Master Erasmus Mundus in

Color in Informatics and Media Technology (CIMET)

Conceiving a Fast and Practical Multispectral-Stereo System

MASTER THESIS REPORT

Presented by

Raju Shrestha

Defended at

Gjøvik University College
4th June, 2010

Jury Committee:

Prof. Alain Trémeau
Prof. Jon Yngve Hardeberg
Prof. Jussi Parkkinen
Ass. Prof. Javier Hernández–Andrés

Supervisors:

Prof. Jon Yngve Hardeberg,
Gjøvik University College, Norway

Ass. Prof. Alamin Mansouri,
University of Burgundy, France

Conceiving a Fast and Practical Multispectral-Stereo System

Raju Shrestha

Master Thesis
Master in **Color in Information and Media Technology (CIMET)**
30 ECTS
Faculty of Computer Science and Media Technology
Gjøvik University College, 2010

Avdeling for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Faculty of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Conceiving a Fast and Practical Multispectral-Stereo System

Raju Shrestha

2010/07/01

Table of Contents

List of Tables

List of Figures

Abstract

Multispectral and 3D imagings are two complimentary imaging technologies with many advantages and great potential for their widespread use in the future, if we can make them faster and more practical. This thesis aims at conceiving such a fast and practical two-in-one stereo-multispectral system.

Multispectral imaging systems remedy the problems of conventional three channel (RGB) color imaging like metamerism and dependency on acquisition conditions, and can at the same time present high spatial and spectral resolution. A multispectral image is composed of several monochannel images of the same object, each image holds data about a specific wavelength, or wavelength interval depending typically on the filter used. The major problems of existing multispectral imaging systems are that they are slow as they require multiple takes and/or they are quite expensive, contributing to the current lack of widespread use in the consumer segment. This thesis has explored creating a fast, practicable and affordable multispectral system with the use of two commercially available digital cameras. Each camera is equipped with an optical filter. These two filters are chosen so that they spread the sensitivities of the cameras in such a way that they become well spaced throughout the visible spectrum covering complementary wave-bands, thus giving rise to a six channel multispectral system.

Like multispectral imaging, 3D imaging systems are also gaining popularity and are of much use in imaging fields. It would be a great advantage if we could have an integrated system capable of both multispectral and at the same time 3D imaging. This thesis, therefore, aims at this in conceiving such a stereo-multispectral system. The two cameras modified with appropriate filters that form a six channel multispectral system are used in a stereoscopic configurations to acquire depth information, making it capable of 3D imaging as well. This leads to a faster and practicable and at the same time affordable multispectral-stereo system.

Both simulation and experimental results have shown that the proposed multispectral system results better estimation of the reflectances both spectrally and colorimetrically and that it is feasible. The 3D images are not much affected by the introduction of the filter(s) as is proved by comparable disparity maps computed out of them.

Such a system could be used for many applications, for example for 3D artwork object acquisition. Knowing the spectral reflectance allows us to simulate the appearance of a 3D object under any virtual illuminant. Moreover, it lets us store this valuable information for future restoration.

Keywords: Multispectral-stereo, multispectral, 3D, stereoscopic, imaging, RGB camera, filter selection

Acknowledgements

First of all, I would like to thank the European Erasmus Mundus Program, The European Commission for giving me this opportunity to take part in this wonderful Master CIMET program. I would also like to thank all the four partner universities: University of Saint Etienne, University of Granada, Gjøvik University College and University of Eastern Finland and corresponding academic coordinators: Professors *Alain Trémeau, Javier Hernández–Andrés, Jon Yngve Hardeberg* and *Jussi Parkkinen* for their efforts in making the program successful and fruitful.

I would like to express my deepest appreciation to my thesis supervisor *Professor Jon Yngve Hardeberg*, whose constant support and guidance in successfully carrying out and completing this thesis work. I am thankful for his all kinds of support and motivation. I feel proud to work with him and learn from his vast knowledge, expertise, experience and intelligence. I am also thankful to my co-supervisor, *Dr. Alamin Mansouri*, University of Burgundy, France for his useful tips, comments and suggestions despite from a distance.

I would also like to thank all the internal and external professors of different courses who taught us during the two years course, and from whom we have learnt not only the taught courses but many other aspects of life. My acknowledgement wouldn't be complete without thanking the CIMET administrative coordinators *Helene Goodsir* and *Thomas Guillobez*, University of Saint Etienne, and *Hilde Bakke*, Gjøvik University College for their constant support and help in all sorts of administrative issues and problems faced during the program. It is my pleasure to thank all CIMET colleagues from all parts of the world for wonderful time spent together during one or more semesters.

I am very much grateful to my family especially my parents for their patience, love and blessings that made it possible for me to come to this point. And, it is my wife *Ranjana Shrestha* who not only gave continuous company, support, love and care; but also a little, lovely son *Raynor Shrestha* who has brought so much joys and happiness in my life. Without her, it would not have been possible to do so well and complete the program successfully. My special thanks and love goes to my lovely wife.

Lastly but not the least, I offer my regards to all of those who directly or indirectly supported me in any respect during the completion of the project.

Raju Shrestha
June, 2010

1 Introduction and Overview

This chapter presents the areas of research of this thesis and the background. We then put forward the research problem, and justification and motivation behind this research work. Research questions upon which this thesis is based on are listed out. Finally, the overview of the thesis organization has been presented.

1.1 Areas of Research

Multispectral imaging is one of the promising imaging technologies with potential to be the next generation imaging technology. It can not only solve many of the problems of conventional three-channel (usually RGB) color imaging like metamerism, illuminant dependency, but also provides more information than just color, and makes highly accurate color reproduction possible. 3D stereo imaging is yet another, getting more popular, technology which allows to capture three dimensional (3D) images, unlike conventional RGB and multispectral imaging systems, which can acquire only two dimensional (2D) images.

Both multispectral and 3D stereo imaging are the areas of research in this thesis work. The term imaging means capture, processing, communication, reproduction and display of images and, all digitally in this digital world. This thesis deals with one of the most important parts: the capture/acquisition. So, the term imaging is used in this thesis basically for the acquisition. We explore the possibility of acquiring both multispectral and 3D images simultaneously with one single system. This thesis covers both the theoretical and practical aspects of both the imaging technologies.

1.2 Background

With the development and advancement of digital cameras, the acquisition and use of images/ pictures have increased tremendously these days. Conventional image acquisition systems which acquire images into three color channels, popularly known as RGB, are by far the most commonly used imaging systems. However, these suffer from several limitations like: they provide only color information, suffers from metamerism and captured images are environment dependent. Different alternative imaging technologies are emerging which try to circumvent those limitations. Among them, multispectral and 3D stereo imaging are the two promising complementary imaging technologies that are of our interest here in this thesis. We briefly discuss these imaging technologies with their characteristics, that would give good background for this thesis.

1.2.1 Conventional Color Imaging

Conventional imaging devices like digital cameras capture still photographs or videos, digitally by recording images via an electronic image sensor (CCD/CMOS). The image data is recorded in three colors, usually red, green and blue (RGB) as illustrated with an image of beautiful rhododendron flower in Figure 1. These devices take advantage of and also suffers from metamerism. Metamerism is the phenomenon by which a pair of spectrally different specimens match in one environment (e.g. illuminant, observer), but not in another environment. They

Figure 1: Three layers of an RGB image

achieve a visual match between two objects with different physical properties through metamerism. But the important problem with this approach is that the data acquired are dependent on the illuminant and the observer (Imai et al. 2000). Furthermore, since the spectral sensitivities of the acquisition device generally differ from the standardized color matching functions, it is not possible to obtain device-independent color. This means that images or reproductions will only match the original under a same set of viewing conditions as the one under which acquisitions were made, which will not be the case in real world. It is quite difficult to accurately estimate the color of objects under arbitrary illumination conditions using current three channel based imaging systems. Since spectral sensitivities of the input device and the human eye are different, it is difficult to estimate color even under the same illumination. And, even if these sensitivities are equivalent, it would still be difficult to estimate the color accurately under different illuminations. As a result, images acquired like this lack the accuracy required for any kind of archiving or analysis. So, what could be the solution to this problem? Multispectral imaging is a solution to the problem. Section 1.2.2 discusses on multispectral imaging.

Besides accuracy, yet another limitation of the conventional three channel imaging systems is that the images acquired with these systems are two dimensional. In the real world, we see the things in three dimensions where we perceive depth as well. 3D image and videos are getting more and more popular these days, thanks in parts to 3D movies like Avatar.

1.2.1.1 Limitations of Trichromatic (RGB) based Camera Systems

Major limitations of the conventional RGB-based systems as pointed out by Yamaguchi et al. (2008) are listed below.

- The RGB values obtained in these systems are device dependent and hence have different meanings, depending on the device characteristics or color processing.
- Since the spectral sensitivity characteristics of color camera are different from the human vision; RGB signal does not have one-to-one correspondence to the tristimulus values perceived by human vision. If the spectral sensitivity is closer to human vision, the color fidelity is improved, but noise behavior is known to become worse.
- The RGB or any other 3-primary color signals that comply with the color space such as sRGB or YCbCr are defined under the white point such as CIE D65 or D50 standard illuminant. To reproduce the color as if the object were placed in the presence of the observer, the color under different illuminant is required, but the standard color space does not support the color under arbitrary illuminant.
- For accurate color reproduction, the spectral reflectance of object and the illuminant spectrum are essentially required for calculating the color under different illuminant. White balance adjustment in RGB space does not give high colorimetric accuracy.
- The color gamut of a normal RGB display is limited, and some of high-saturation colors cannot be reproduced. Even if the display device allows the display of wider color gamut, conventional color signals such as sRGB or ITU-R BT.709 do not support wider gamut color signal.
- The observer metamerism effect, due to the observer dependence of color matching functions, cannot be ignored in high-accuracy color reproduction, for example, in the color proofing of printed materials with color monitors.
- In the image archive, database, or analysis, the utilization of color information is limited, since the RGB signal depends on the devices, illuminants, and preprocesses involved in the imaging systems. For example in the image retrieval using color information, target object cannot be found, if the illumination condition is different.
- RGB systems provide information in two dimensions only while we see the world in three dimensions.

Multispectral imaging is used to overcome most of these problems while 3D stereo imaging is used to acquire 3D images.

1.2.2 Multispectral Imaging

Multispectral imaging captures image data at specific wavelengths across the electromagnetic spectrum. This is unlike conventional imaging which captures three channel values: red, green and blue. A grayscale image is a one channel image and an RGB image is a three channel image (Figure 1). A multispectral image is a multi-channel (more than three) images which can be seen as a stack of images, each representing the intensity image at a given wavelength (Figure 2). Number of channels in multispectral image, is in general, between 4 and 20. Imaging with higher number of channels, may be several hundreds, are known as hyperspectral imaging. There is no clear cut boundary between multispectral and hyperspectral imaging. Hyperspectral image acquisition systems are complex and expensive. We are not considering hyperspectral imaging in this thesis. In multispectral imaging, the wavelengths may be separated by filters or by the use of instruments that are sensitive to particular wavelengths, including light from frequencies beyond the visible light range, such as infrared. From a spectroscopic viewpoint, in the multispectral image, we obtain a spectrum in each pixel (Figure 3). Recent studies of multispectral color reproduction systems revealed that increasing the number of bands in the input device enables the spectrum to be estimated more accurately (Burns and Berns 1996; Day et al. 2003). This makes it possible to improve the accuracy of color estimation under arbitrary illumination spectrum.

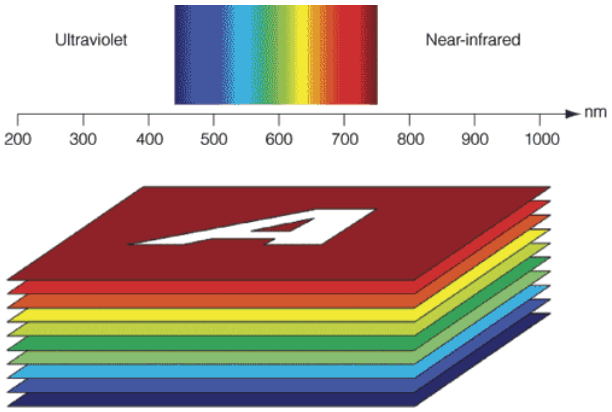


Figure 2: Stack of a multispectral image [Source: Carstensen (2007)]

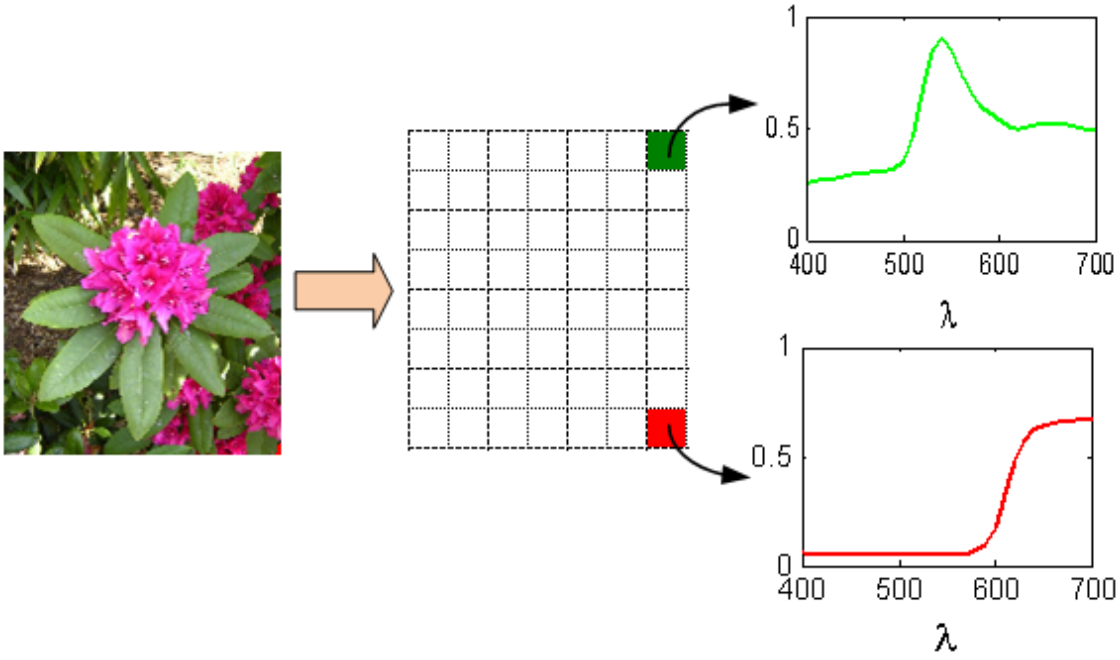


Figure 3: Spectra for each pixel in a multispectral image

In order to realize a multispectral system, several questions need to be addressed: How to do image acquisition? How many channels do we need and correspondingly which filters should we use? What range of spectrum will be covered? How do we evaluate the system?

We will discuss in details the multispectral image acquisition and filter selection algorithms in Chapter 2. The range of spectrum to be covered and number of channels required are application dependent. Unlike conventional color imaging, multispectral imaging can be used to work not only in visible spectrum, but also in infrared and ultraviolet spectrum as well. The number of sensors is restricted by the increased cost and memory requirements as well as manufacturing limitations. Different people suggested different number for the same set of targets. As reported in (Connah et al. 2006), Cohen (1964) came up with 3 channels, Maloney and Wandell (1986) with 5-7, Parkkinen et al. (1989) and Wang et al. (1997) with 8 for Munsell reflectances. Miyake and Yokoyama (1998) recommended 5 channels for paint and artwork reflectances and Maitre et al. (1996) recommended 10-12 channels for the same set of reflectances. Similarly Dannemiller (1992) recommended 3 channels, while Imai (2000) recommended less than 10.

The different results in those studies was due to the fact that they used different decision criteria. Connah et al. (2006) investigated the minimum number of sensors to use based on minimizing reconstruction error for different types of reflectance surfaces. It considered noise in the acquisition to make it more realistic. These studies have shown that not more than 20 channels (usually less) would be required for good multispectral imaging.

Evaluation of a multispectral system is normally done based on the spectral reproduction, or color reproduction or both depending upon the requirement. Section 2.4 discusses commonly used evaluation metrics.

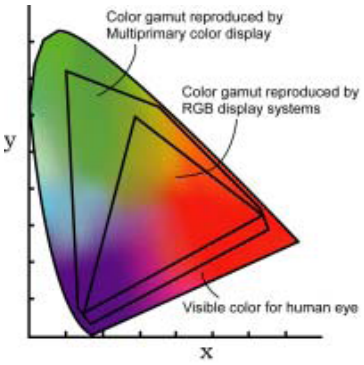


Figure 4: Illustration of gamut expansion by multi-primary color display [Source: Yamaguchi et al. (2008)]

1.2.2.1 Advantages of Multispectral Imaging

Following are the major advantages of multispectral imaging over the conventional color imaging:

- It has been reported that the use of multispectral imaging significantly improves the color accuracy (Pratt and Mancill 1976; Hill and Vorhagen 1994; Tominaga 1996; Burns and Berns 1996; Yamaguchi et al. 1997).

- It makes color reproduction under different illumination environment possible with reasonably good accuracy (Tsumura 2006).

  To display the color under an illuminant different from the image acquisition environment, the image spectral reflectance is estimated from the multispectral image and the spectral power distribution of the illuminant used during the image acquisition. The estimated spectral reflectance is then used to compute color under any arbitrary illuminant.

- An expanded color gamut can be obtained by multi-primary color displays (Ajito et al. 2000), scanner (Cotte and Dupraz 2007) and printers (Berns et al. 2003; Mourad and Kornfeld 2006). Figure 4 shows an example of gamut expansion with multi-primary color display.

- The influence of observer metamerism can be reduced by the spectral color reproduction (Yamaguchi et al. 2002; Murakami et al. 2004)

- The quantitative spectral attributes of an object, useful for its analysis or recognition, can be captured and preserved.

- It is not limited to visual range, rather can also be used in near infrared, infrared and ultraviolet spectrum as well (Horman 1976; Ellrod et al. 2003; Huang 2004; Ononye et al. 2007).

1.2.2.2 Limitations of Existing Multispectral Systems

The current state of the art multispectral systems need multiple shots to capture multispectral images (Tominaga 1999; Hardeberg et al. 2002; Cotte and Dupouy 2003; Huang 2004; Park et al. 2007; Hashimoto and Kishimoto 2008; Everdell et al. 2009). Moreover, the construction and operation of such systems are complex and also costlier. These severly limit the use of the multispectral systems to advanced applications like airborn satellite imaging in remote sensing and researches only. Ohsawa et al. (2004) proposed 6-band HDTV camera systems. It allows one shot image capture. However, it needs custom designed filters and complex optics making it still far away from being practical. This thesis has proposed a fast and practical solution to multispectral imaging.

1.2.3 3D Stereo Imaging

We see the world in three dimensions (3D) through our two eyes and thus perceive the depth. The perception of depth is made possible due to the slight differences in the two retinal images, called horizontal disparity, retinal disparity, or binocular disparity. The differences arise because of the different positions of the eyes in the head. Each eye captures slightly different image of an object, and these two images are combined in the mind by matching up the similarities and adding in the small differences. The small differences between the two images add up to a big difference in the final picture, resulting a three-dimensional stereo picture. This process of perceiving depth in our visual perception is called stereopsis. Stereo imaging tries to emulate this human visual phenomenon to acquire 3D images. Stereoscopic (or 3D) photography/imaging is not new. The stereoscope was invented in 1838 by Charles Wheatstone, and the first 2-lensed stereo camera was built in 1858. Commercially made 3D cameras were popular in the 1920's and the 1950's, but they represented just a tiny fraction of the photography market.

Today, 3D photography that is digitally based is rapidly expanding in popularity, thanks to the advantages that digital photography offers. People have been putting together twin-camera rigs for doing digital 3D photography. Even big manufacturers like Fujifilm have already started producing commercial 3D stereo cameras. 3D imaging (images and videos) are gaining popularity and are of great use in imaging fields. Jaumann et al. (2007) presented the high-resolution stereo camera (HRSC) experiment on Mars Express. There are 3D movies, televisions and video games. Belgian newspaper La Derniere Heure recently issued a special '3D edition', claimed to be the first in Europe.

1.3 Problem Description

We have seen that multispectral and stereo imaging technologies are two complementary technologies that have many advantages compared to conventional color imaging. However, there are still lack of their widespread and general use because of their complexity and cost.

One serious problem in the state of the art multispectral acquisitions is that they need several shots for acquisition, which not only slows down the acquisition process but also makes it unusable for capturing the scenes in motion. The conventional color imaging systems, and even multispectral systems can acquire only 2D information, whereas we see the world in 3D. This thesis targets both speedy and 3D acquisition with the aim at conceiving such a multispectral-stereo system that can act as a multispectral as well as 3D stereo imaging system, and at the same time is fast and practical.

The idea is to use two cameras, whose spectral sensitivities are modified with a pair of appropriate filters from among readily available filters, to form a six channel multispectral system. The two cameras are used in a stereoscopic configuration, thus making it capable of 3D imaging as well. By synchronizing their opteration, the system captures both multispectral and 3D stereo data simultaneously in one shot. This leads to a faster and practicable and at the same time affordable stereo-multispectral system. We investigate feasibility of such systems and also evaluate their performances compared to conventional 3-channel RGB systems.

1.4 Justification, Motivation, and Benefits

Both the multispectral and 3D imaging technologies have potential to be the next generation imaging technologies in the near future if the existing problem of speed, complexity and cost could be solved. This motivates us to do this thesis work in realizing both multispectral and stereo imaging in one system. It will be capable of capturing both multispectral and 3D stereo data simultaneously. And, since it is a one-shot system, this leads to a faster and practicable and at the same time affordable stereo-multispectral systems.

To produce a spectral reflectance that can faithfully represent an object surface, a multispectral camera is coupled with stereo imaging. A reflectance spectrum obtained with multispectral imaging can be associated with each 3D reconstructed point. By linking 3D points to a curve representing the spectral reflectance, the system gives a physical representation of the matter that is ideally independent from the environment: illuminant, observer, and acquisition devices. Knowing the spectral reflectance, thus, lets us simulate the appearance of a 3D object under any virtual illuminant. Moreover, it lets us store this valuable information for future restoration.

This work aims at, if not the general public but at least to enthusiasts, who are interested in multispectral and 3D imaging world, but may not be willing to invest a lot of money and at the same time avoid complexity. Therefore, flexibility, robustness, and low cost are important. With the advancement of digital camera systems, their prices are also decreasing. Adding two appropriate filters will not add much cost. So multispectral-stereo system we envisioned here in this thesis will not only have more mobility, convenience, and attraction, but also affordable.

Thus with the sucess in our goal in this thesis work, of conceiving a fast and practical stereo-multispectral system, it is believed that this work will place a milestone towards the development and use of multispectral-stereo imaging in the near future for wider applications.

1.5 Research Questions

Following research questions have been formulated as the basis of this thesis work.

1. Can we make a fast and practical multispectral system using two digital cameras and a pair of filters?
2. Can we acquire both multispectral and 3D stereo images with one system?
3. Do we need to use same type of cameras or different cameras can also be used?
4. Do such multispectral systems perform better than conventional 3-channel system?
5. How to correspond images from the two cameras?
6. What would be the influence of introduction of filters on the 3D images?

1.6 Research Methodology

This thesis studies both multispectral and 3D stereo imaging aspects of the proposed multispectral-stereo system. The multispectral systems have been investigated with both simulation and experimental approaches. In the simulation based approach, the response of the cameras are computed based on the multispectral camera model and results of the proposed 6-channel multispectral system are compared with conventional 3-channel system. In the experimental approach, the captured responses of the real multispectral camera constructed during the experimental studies are used instead. As much measurements are done as possible, instead of using standard data, for more accurate results.

In the case of stereo system, we have investigated the results of the 3D image captured with the proposed multispectral-stereo system and compared with the results from the pure stereo camera. Individual cameras, and stereo calibration have been carried out to determine the exact intrinsic and extrinsic parameters of the camera system. Then, disparity maps are computed from the pair of stereo images using a stereo matching algorithm. These maps are analyzed and compared with the maps from the pure stereo system to investigate the 3D images acquired with the multispectral-stereo system.

1.7 Paper Publication

With the encouraging results from the simulation studies during the initial phase of the thesis work, we had submitted a paper entitled "**Multispectral image capture using two RGB cameras**", and it has already been accepted in the European Signal Processing Conference (EUSIPCO 2010) to be held in 23-27 August 2010, Aalborg, Denmark.

We have recently submitted yet another paper based on both the simulation as well as experimental results from the perspective of multispectral imaging, in a prestigeous international conference. And, we are also planning to make submission of a paper with special focus on stereo aspect of the work done in this thesis.

1.8 Thesis Organization

This thesis comprises of eight chapters describing in details the work involved and the state of the art technologies and theories.

This chapter introduces the research problem along with the background introduction on conventional imaging technology, its limitations and the two complementary solutions: multispectral and stereo imaging. It points out the justfication and the motivation behind this thesis work and its expected benefits.

The rest of this thesis from chapter two is organized as follows:

Chapter 2 presents theories and state of the art technologies on multispectral imaging, particularly on acquisition. It discusses how multispectral images are acquired, how do we select filters, how to analyze the result and how do we evaluate a multispectral system.

Chapter 3 discusses 3D stereo imaging technologies and theories behind them. Starting with stereo vision, it presents different stereo imaging models, individual camera and stereo camera calibration techniques, estimation of disparity and depth from the stereo images.

Chapter 4 puts forward our proposed multispectral-stereo system, compares with previous related works and presents the methodologies followed in the study of proposed multispectral and stereo imaging systems.

Chapter 5 investigates our proposed system from multispectral image acquisition point of view based on simulation approach. It studies and analyzes the feasibility of such a system based on results of simulations with different types of cameras.

Chapter 6 continues with the proposed multispectral system, this time the presentation is based on real experimental works. Here, it tries to validate the results from the simulation. All the devices, instruments and tools used in the experimental works, their configurations and characterizations and experimental procedures, are thoroughly explained.

Chapter 7 discusses works on stereo aspects of the project work. It put forwards the camera calibration, stereo calibration and stereo matching techniques used and presents the resulting disparity maps on several 3D stereo images acquired with our multispectral-stereo system to validate that the proposed multispectral-stereo system is capable of acquiring 3D stereo images in addition to the multispectral information.

Chapter 8 presents the conclusion and the future perspectives of this thesis work.

The camera sensitivity data obtained from measurement, the list of filters used, camera calibration output files, and the list of OpenCV functions used are given in Appendices A, B, C and D respectively.

Due to the large volume of the program codes in Matlab and Visual C++, they are not included in this report. However, they can be made available through the Norwegian Color Research Laboratory, Gjøvik University College, or by directly contacting me via my email raju_shrestha@hotmail.com.

2 Multispectral Image Acquisition

By terming image acquisition here, we mean digital image acquisition as imaging technologies are, if not all, but mostly digital nowadays. This chapter introduces conventional trichromatic color imaging and the model used. It then presents the state of the art multispectral image acquisition techniques and important issues in the imaging process: selection of filter, acquisition model, estimation of spectral reflectance. At first, we discuss briefly on how we perceive color and model it.

2.1 Color Perception and Modeling

We perceive color from an object when our eye senses the physical stimulus (reflected light) from the object through its three cone receptors: L, M and S. The signals thus sensed are communicated and interpreted by our neural system and the brain to give color perception. The physical stimulus for color is electromagnetic radiation in the visible region of the spectrum (wavelength from $\lambda_{min} = 360$nm to $\lambda_{max} = 830$nm), which is commonly referred to as light. Figure 5 illustrates the phenomenon of perceiving color.

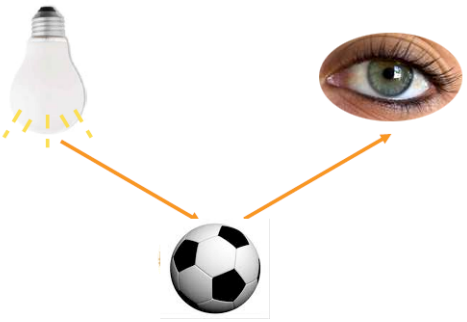


Figure 5: How we see color?

Assuming a fixed set of viewing conditions, the response of the cones can be modeled by a linear system defined by the spectral sensitivities of the cones. Let $f(\lambda)$ be the spectral distribution of light incident on the retina, where λ denotes the wavelength, the responses of the three cones can be modeled as a three component vector $c = [c_L c_M c_S]^T$ given by the equation:

$$c_i = \int_{\lambda_{min}}^{\lambda_{max}} s_i(\lambda) f(\lambda) d\lambda, \; i = L, M, S, \tag{2.1}$$

where s_i denotes the spectral sensitivity of the ith type of cone, and λ_{min}:λ_{max} denotes the interval of wavelengths of the visible spectrum outside of which the spectral sensitivities are assumed to be all zero. The normalized sensitivities of the human eyes are shown in the Figure 6.

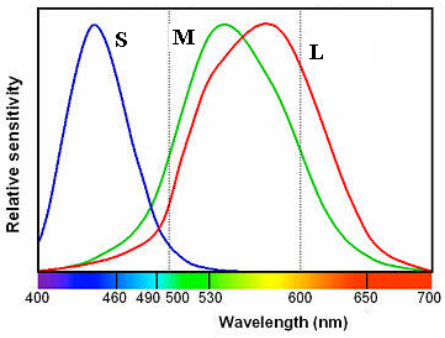


Figure 6: Normalized cone sensitivities of the human eye [Source: `http://www.normankoren.com`]

For non-luminous, reflective objects, the spectral distribution $f(\lambda)$ of the light incident on the retina is the product of the spectral reflectance $r(\lambda)$ of the object surface and the spectral radiance $l(\lambda)$ of the viewing illuminant. We may thus write Equation 2.1 as follows:

$$c_i = \int_{\lambda_{min}}^{\lambda_{max}} s_i(\lambda) l(\lambda) r(\lambda) d\lambda, \; i = L, M, S. \tag{2.2}$$

By uniformly sampling the spectra with a proper wavelength interval $\Delta\lambda$, Equation 2.2 can be numerically approximated by replacing integrals with summations as given below:

$$c_i = \sum_{\lambda_{min}}^{\lambda_{max}} s_i(\lambda) l(\lambda) r(\lambda) \Delta\lambda, \; i = L, M, S. \tag{2.3}$$

It can be expressed in matrix form as follows:

$$C = S^T \text{Diag}(L) R, \tag{2.4}$$

where C is the cone response vector, S is the cone sensitivity matrix, Diag(L) is the diagonal illuminant matrix with entries from the samples of $l(\lambda)$ along the diagonal and R is sampled spectral reflectance of the object.

Since, the cone response functions are quite difficult to measure directly, non-singular linear transformations of the cone responses are readily determined through color matching experiments. A standardized set of color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are defined by the CIE, and is widely used in colorimetric definitions. The CIE 1931 XYZ color matching functions are shown in Figure 7. Defining $A = [\bar{x}, \bar{y}, \bar{z}]$ as the matrix of sampled color matching functions, we can represent a color stimulus using its CIE XYZ tristimulus values X as follows:

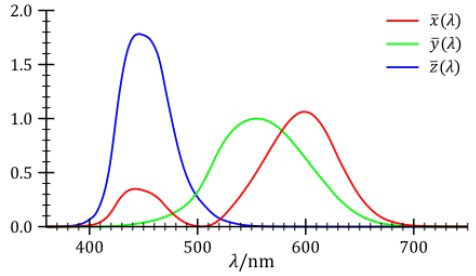


Figure 7: CIE XYZ color matching functions [Source: `http://en.wikipedia.org/wiki/CIE_1931_color_space`]

$$X = A^T \text{Diag}(L) R. \tag{2.5}$$

RGB color values can be obtained from X by means of transformation matrix.

2.2 Color Image Acquisition

With the arrival of digital image acquisition devices like digital cameras and scanners, the image acquisition process becomes quite simpler compared to quite complex image acquisition process in conventional photographic techniques and at the same time high resolution cameras allow a perfect spatial resolution and an excellent color fidelity. Here we are interested in digital camera, so our discussion here will be on color imaging with digital camera. As digital imaging technology has been improved much and high resoultion cameras are available in reasonably cheaper price in last few years, digital cameras have rapidly become more popular.

2.2.1 Digital Camera Model

The structure and operation of the digital camera is very similar to human eye, as both of them are based on two major components: a lens assembly, and an imaging sensor. The lens assembly captures a portion of the light emanating from an object, and focuses it onto the imaging sensor. The imaging sensor then transforms the pattern of light into a video signal, either electronic or neural. Retina is the imaging sensor in our eyes, whereas CCD or CMOS sensors are used in digital cameras.

A digital image is a long string of 1s and 0s that represent all the tiny colored dots or pixels that collectively make up the image. These devices sample and quantize the continuous-space, analog, real-world images to acquire digital images. In order to get a full color image, most sensors use filtering to look at the light in its three primary colors. The camera combines all three colors to create the full spectrum. There are several ways of recording the three colors in a digital camera, for example: single sensor with demosaicing, three sensors. In any case, there will be three values, usually, Red, Green and Blue (RGB) corresponding to each pixel. Thus, like three cone receptors and cone sensitivities in color perception, there are three channels: R,

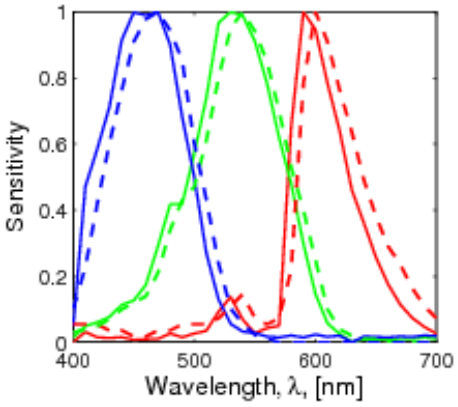


Figure 8: Camera sensitivities of Nikon D70 (solid) and Canon 20D (dashed)

G, and B; and correspondingly three channel sensitivities. Sensitivities of the three channels, also known as camera sensitivities, in general, differ for different cameras. Figure 8 illustrates camera sensitivities of Nikon D70 and Canon 20D cameras.

Like cone response as described in Section 2.1, camera response can be modeled by

$$c_i = \int_\lambda s_i(\lambda) l(\lambda) r(\lambda) d\lambda, \; i = R, G, B. \tag{2.6}$$

All the symbols are the same as defined in Section 2.1 except s_i which denotes camera sensitivities instead of cone sensitivities. In discrete sum form, Equation 2.6 can be expressed as

$$c_i = \sum_\lambda s_i(\lambda) l(\lambda) r(\lambda) d\lambda, \; i = R, G, B. \tag{2.7}$$

And, in matrix form the model is given by

$$C = S^T \text{Diag}(L) R. \tag{2.8}$$

This model represents the camera response in the ideal case when there is no acquisition noise. But in practice, some degree of noise is always present in any electronic device that transmits or receives a signal, and digital camera is no exception. So, the noise model needs to be incorporated into the camera model to make it more realistic.

A noisy camera can be modeled mathematically by adding acquisition noise n to the noise free camera model given by Equation 2.8 as (Connah et al. 2006)

$$C = S^T \text{Diag}(L) R + n. \tag{2.9}$$

In order to model the acquisition noise n, we need to know the possible sources of noise that may give rise to different types of noises in the camera. Section 2.2.2 discuses different sources and types of noise.

2.2.2 Noise Sources

Noise can be defined as any unwanted signal added to a device output which contains no information. The first major source of noise comes from the camera sensor (CCD/CMOS) itself. A digital camera has several sources of inherent noise that are all dependent on factors such as time, signal, and temperature.

The main sources of temporal noise in a digital camera include dark current (DC), random shot noise, reset transistor noise, CCD clocking noise, quantization noise and noise from the output amplifier. Dark current noise is dependent on the operating temperature and the integration time. The dark current noise also varies across the pixels of the imager, leaving a fixed pattern noise. The dark noise can be dealt with by taking several dark images at the appropriate integration time, averaging them, and using the average pixel referenced data as the zero level for each pixel. The reset transistor noise and output amplifier noise is generally dealt with by using different methods of sampling the signal which is beyond the scope of this research. The random shot noise is caused by the random arrival of photons at the sensor and it cannot be eliminated, but can be reduced by taking several images and averaging them.

Another source of noise comes from the scene illumination. In practice, there is always certain amount of variation in the light due to the unevenness in lighting itself and also due to the optics of the system. This noise is dealt with by taking a reference image of a uniform target, such as an even gray or white surface of known reflectance and using it to flat field the image.

Digital cameras contain an imaging sensor which delivers an analog signal. This signal is converted by an Analog-to-Digital Converter (ADC) into a stream of digital numbers with certain bits resolution. If the noise on the analog signal is in the same order of magnitude or smaller than the quantization step of the ADC, the quantization procedure creates errors which can be described by quantization noise added to the signal. Thus, quantization noise arises from the discrete nature of the sampling/conversion process and hence it is separate and distinct from other sources of error.

2.3 Multispectral Image Acquisition

It is well known that with a panchromatic imaging sensor and three well-chosen optical filters, it is possible to obtain a fairly good reconstruction of the color tristimulus values of the reference human observer as defined in colorimetry. As discussed in Section 1.2.1, the limitation of conventional three color imaging can be circumvented by increasing the number of channels in the image acquisition and increase the color quality significantly by using multispectral imaging. A multispectral image, thus, is an image where each pixel contains information about the spectral reflectance of the imaged scene. Multispectral image acquisition aims to estimate the spectral reflectances of the scene by using more than three filters in sequence. Here we will discuss on state of the art multispectral image acquisition techniques.

The main function of a multispectral acquisition system is to allow acquisition of a scene in different bands of the light spectrum. Multispectral systems can acquire images beyond visible spectrum, in near infrared, infrared and ultraviolet range as well. However, we are interested in visible band only here in this thesis work.

There have been many reports on devices for the acquisition of mulispectral images. Maloney and Wandell (1986); Hill and Vorhagen (1994); Tominaga (1996); Burns and Berns (1996); Maitre et al. (1996); Yamaguchi et al. (1997); Farina et al. (2000); Levenson et al. (2003); Tsuchida et al. (2005) are some of the example works. A multispectral image acquisition system, in general, comprises of object observed, light source (illumination), optical path, color filter, and imaging sensor/camera as shown in Figure 9.

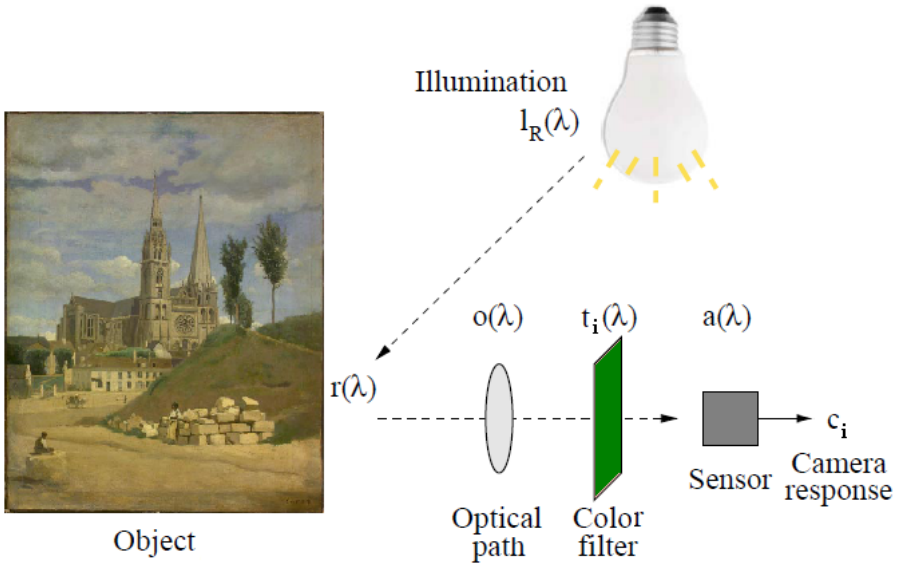


Figure 9: Components of multispectral image acquisition process [Source: Hardeberg (1999)]

2.3.1 Types of Multispectral Image Acquisitions

Multispectral image acquisition systems can be categorized into different types based on different aspects.

Based on ***band size***, multispectral image acquisition can be categorized into following two types:

1. *Narrow band:* This approach needs narrow band sampling of the spectrum. For this, monochromator devices, liquid crystal tunable filters (LCTF), narrow-band interference filters or tunable light lasers are used. This approach is generally quite accurate spectrally, but acquisition is often quite slow, expensive and those special equipments are required.

2. *Wide band:* This approach typically uses a smaller set of wide band optical filters to capture digital image data and then uses spectral reconstruction algorithms to estimate the spectral reflectances of an object. This approach produces reasonable results because both manmade and natural materials generally have smooth spectral reflectance shapes, thus the number of channels can be reduced to between eight and ten channels with broader band pass filters while still achieving accurate results. Subsection 2.3.4 discusses on how to estimate spectral responses from camera response.

Based on the ***techniques used to separate bands***, we can categorize three types of multispectral image acquisition:

1. *Use of Filters:* In a typical filter based imaging systems, either a rather large set of traditional optical filters in a filter wheel, or a tunable filter (Miller and Hoyt 1995; Hardeberg 1999; Hardeberg et al. 2002) capable of many different configurations are employed. With filter wheel, mechanical rotor (manual or computer controlled (Brauers et al. 2009)) is used to rotate the wheel and select a filter at a time. Tunable filters allow to select the band electronically thus allowing faster acquisition compared to the system with filter wheel. In both cases, minimizing the actual number of filters used while keeping the error sufficiently low is important to reduce operational costs, acquisition time, and data storage space. How to select filters from among a large set of filters is an important problem in multispectral acquisition. Section 2.3.2 discusses on filter selection algorithms.

2. *Extension of Color Filter Array:* In this type of acquisition systems, instead of using filters in front of sensors, extended mosaicked color filter array (CFA) is used to obtain more than three color bands (Garcia et al. 2000; Miao and Qi 2004). This type of multispectral system is still under research for its practical use.

3. *Use of Tunable Light Source:* Here the light source itself is tuned into different bands. One example of doing this is the use of tunable laser. Another way of having tunable light source is the use of multiplex multi-color LED illumination (Park et al. 2007; Everdell et al. 2009; Bouchard et al. 2009). It is gaining popularity for allowing faster acquisition and with availability of LED in more and more colors.

Each channel data is captured with a sensor (CCD or CMOS). Based on ***types of sensors*** used, the acquisition system can be categorized as follows:

1. *Linear Array Sensor:* Linear arrays sensors are used in devices like scanners and specially designed cameras like CRISATEL (Cotte and Dupouy 2003). With linear array sensors, the system scans the scene line-by-line. For example, the CRISATEL camera has a 12000 pixel linear array vertically mounted and scans horizontally by a step motor up to 20000 vertical lines, thus enabling it to capture 12000×20000 pixels image.

2. *Matrix Sensor:* This is the case when using digital cameras whose sensor (CCD or CMOS) is normally designed in matrix form. With a monochrome digital camera, a whole image scene (one channel at a time) can be captured at once without need of scanning (Hardeberg 1999; Hardeberg et al. 2002).

   With a high quality tri-chromatic digital camera in conjunction with a set of appropriate optical filters allows to acquire unique spectral information (Imai 1998; Imai and Berns 1999; Tominaga 1999; Imai 2000; Huang 2004; Valero et al. 2007; Yamaguchi et al. 2008). This method enables three channels of data to be captured per exposure as opposed to one. With a total of n colored filters, there are $3n + 3$ camera responses for each pixel (including responses with no colored filters) correspondingly giving rise to a $3n + 3$ channel multispectral image. This greatly increases the speed of capture and allows the use of technology that is readily and cheaply available and does not need to be specialized. Such cameras are generally optimized to achieve good perceptual quality of the image.

3. *Two matrix sensors:* In all the above types of multispectral image acquisition techniques including one with a trichromatic digital camera, we have to take several shots with each chosen filter; this severely limits the usability for non-stationary scenes. This is among the important problems we are trying to answer in this thesis work. So how about if we use two matrix sensors instead of one?. Here, instead of one, we use two trichromatic digital cameras. This is the approach proposed in this thesis work. The idea is to select a pair of appropriate color filters, one for each camera, so that they will modify the sensitivities to give six reasonably well spaced channels in the visible spectrum. Chapter 4 will cover related works and our proposed multispectral-stereo system in details.

2.3.2 Selection of Filters

Which filters are to be used in a multispectral image acquisition is one of the important questions to be addressed in designing a filter-based multispectral acquisition system. Several filter selection algorithms have been proposed (Hardeberg 1999, 2004; Day 2003*a*). They are:

1. *Equi-spacing of filter dominant wavelength:* This is the first and most intuitive method where we simply try to find a combination of filters which spread the dominant wavelengths equally spaced across the spectrum of interest. However, it is not practicable where there is a large number of filters (Hardeberg 1999, 2004; Day 2003*a*).

2. *Progressive optimal filters:* In this method, the kth filter is selected from the set of available filters as the one that shows the highest correlation to the progressive optimal filter of the first k characteristic reflectances (PCA analysis), having only positive values (Hardeberg 1999, 2004).

3. *Maximizing filter orthogonality:* It chooses a set of filters whose non-normalized vectors have a maximized orthogonality (Hardeberg 1999, 2004).

4. *Maximizing orthogonality in characteristic reflectance vector space:* Here, the idea is to select filters that have a high degree of orthogonality after projection into the vector space spanned by certain most significant characteristic reflectances (Hardeberg 1999, 2004).

5. *Exhaustive Search:* All possible combinations of the filters are evaluated and the one that gives the best result is chosen (Hardeberg 1999, 2004; Day 2003*a*). So it clearly gives the optimal solution. However, the complexity of this method could be prohibitive, since it requires the evaluation of large number of permutations or combinations; especially when there is a large number of filters. For selecting k filters from given set of n filters, it requires $P(n, k) = \frac{n!}{(n-k)!}$ permutations, or $C(n, k) = \frac{n!}{k!(n-k)!}$ combinations, depending upon the multispectral system designed.

There are some other filter selection approaches proposed in the literatures. For example, Novati G. and R. (2004) proposed filter selection method based on filter vector analysis. Chatzis et al. (2006)'s approach is based on feature vector analysis.

2.3.3 Multispectral Image Acquisition Model

The multispectral image acquisition systems are usually thought as inherently device independent as they are capable of recording the spectral reflectance of the surface imaged in each pixel of the scene by discounting the spectral characteristics of the acquisition system and of the illuminant. It is assumed that the spectral radiance of the illuminant used for the acquisition is known, either by direct measurement, or indirectly by estimation of the spectral characteristics of the acquisition system. To obtain device independent image data of high quality, it is important to know the spectral characteristics of the system components (like camera, illuminant etc.) involved in the image acquisition process.

Denoting the spectral radiance of the illuminant by $l_R(\lambda)$, the spectral reflectance of the object surface imaged in a pixel by $r(\lambda)$, the spectral transmittance of the optical systems in front of the detector array by $o(\lambda)$, the spectral transmittance of the ith optical color filter by $t_i(\lambda)$ and the spectral sensitivity of the sensor array by $a(\lambda)$. In a multispectral system, a set of filters are used in place of optical color filter represented in the Figure 9. Let K be the number of filters used. Image is captured by changing the filter at a time giving K channel image data. Supposing a linear optoelectronic transfer function of the acquisition system, the camera response c_i for an image pixel is then given by

$$c_i = \int_{\lambda_{min}}^{\lambda_{max}} a(\lambda)t_i(\lambda)o(\lambda)l(\lambda)r(\lambda)d\lambda + n_i = \int_{\lambda_{min}}^{\lambda_{max}} s_i(\lambda)r(\lambda)d\lambda + n_i,\ i = 1, 2, ..., K, \tag{2.10}$$

where $s_i(\lambda) = a(\lambda)t_i(\lambda)o(\lambda)l(\lambda)$ denotes the spectral sensitivity of the ith channel and n is the noise of the acquisition system.

Like in Equation 2.9; by uniformly sampling the spectra into N equal wavelength intervals, we can express the multispectral camera model with noise in matrix form as

$$C_i = S_i^T \mathrm{Diag}(L)R + n_i, \tag{2.11}$$

where $S_i = [s_i(\lambda_1)s_i(\lambda_2)...s_i(\lambda_N)]^T$ and $R = [r(\lambda_1)r(\lambda_2)...r(\lambda_N)]^T$ are the vectors containing the spectral sensitivity of the ith channel of the acquisition system, and the sampled spectral reflectance, respectively.

System unknowns in S_i vectors, known as spectral sensitivity functions are estimated with one of the two methods:

- The first method is based on direct spectral measurements, requiring quite expensive equipment, in particular a wavelength-tunable source of monochromatic light. The camera characteristics are determined by individually evaluating the camera responses to monochromatic light from each sample wavelength of the visible spectrum (Park et al. 1995; Burns 1997; Verdu et al. 1998; Sugiura et al. 1998).

- The second method is based on the acquisition of a number of samples with known reflectance or transmittance spectra. By observing the camera output to known input, we may estimate the camera sensitivity. Several authors have reported the use of such methods, e.g. (Pratt and Mancill 1976; Sharma and Trussell 1993; Farrell and Wandell 1993; Sherman and Farrell 1994; Hubel et al. 1994; Burger and Sherman 1994; Sharma and Trussell 1996; Maitre et al. 1996; Jon et al. 1998).

Section 2.3.4 will discuss on how to estimate spectral reflectance, i.e. the device independent image data from the camera response.

2.3.4 Spectral reflectance estimation from camera responses

This is the problem of estimating the reflectance of the object R from the camera responses $C = [C_1 C_2 ... C_n]^T$ in Equation 2.11. That means to estimate the $N \times 1$ vector R given the $K \times 1$ vector of camera responses C and the matrix S. This is a system of K linear equations in N unknowns. For an exact solution it is sufficient to set the number of knowns equal to the number of unknowns (i.e. K=N). However, in practice, normally N>K. This causes infinite possible solutions to the equations.

Several methods of spectral reflectance estimation from camera responses have been proposed (Maloney and Wandell 1986; Imai 2000; Haneishi et al. 2000; Shi and Healey 2002; Day 2003*b*; Imai and Berns 2002; Imai et al. 2003; Cheung et al. 2005; Connah and Hardeberg 2005; Mansouri et al. 2005; Zhao and Berns 2007; Zhang and Xu 2008). The four most commonly used methods that are used here in this thesis are described below.

- **Least Square-Wiener (LSQW):** It is based on building a mapping from camera responses to reflectance that minimizes the least squares error (Finlayson 1997; Hardeberg 1999). Given k reflectances in the training data set, we seek a matrix M that transforms camera responses to reflectance such that

$$\min_{M} \sum_{i=1}^{k} \|r_i - Mc_i\|^2, \tag{2.12}$$

  where r_i is reflectance of the i^{th} surface of the training set, c_i is the corresponding camera response. With usual notation, R denoting the reflectance matrix and C denoting the camera response matrix, the solution to the Equation 2.12, known as the least-square solution is given by

$$M = RC^+ = RC^T(CC^T)^{-1}, \tag{2.13}$$

  where $C^+ = C^T(CCT)^{-1}$ is the pseudo-inverse of the matrix C. Assuming zero noise in the acquisition, C in Equation 2.13 can be replaced by $E^T R$, giving the reflectance estimate $\tilde{R}$ as

$$\tilde{R}_{(test)} = R_{(train)} R^T_{(train)} E(E^T R_{(train)} R^T_{(train)} E)^{-1} C_{(test)}, \tag{2.14}$$

  where $E = Diag(L)S$. The simulated camera output C is computed for the test target reflectance $R_{(test)}$ using Equation 2.14. With noise n, this equation is modified as:

$$\tilde{R}_{(test)} = R_{(train)} R^T_{(train)} E(E^T R_{(train)} R^T_{(train)} E + nn^T)^{-1} C_{(test)}. \tag{2.15}$$

  This estimate is dependent upon the correlation RR^T of the training set, and hence the solution is strongly tied to the choice of characterization set.

- **Linear Model**: A linear-model approach formulates the problem of the estimation of a spectral reflectance $\tilde{R}$ from the camera responses C as finding a transformation matrix (or reconstruction matrix) Q that reconstructs the spectrum from the K measurements as follows

$$\tilde{R} = QC. \tag{2.16}$$

  The goal now is to determine the matrix Q that minimizes a given distance metric $d(r, \tilde{r})$. *Linear regression* method determines Q from the training data set using

$$Q = R_{(train)} C^+. \tag{2.17}$$

  The pseudo-inverse C^+ may be difficult to compute and when the the problem is ill-posed, it may not even give any inverse, so it may need to be regularized. Subsection 2.3.5 discusses on regularization.

  Maloney (1992), and Imai and Berns (1999) estimate the spectral reflection by approximating R by linear combination of a small number of basis functions:

$$R \approx Bw, \tag{2.18}$$

where B is a matrix containing the basis functions obtained from training data set, and w is a weight matrix. Different approaches have been proposed for computing w.

Maloney (1992) computes w from the training data set by using $C = E^T R$ and substituting the value or R from Equation 2.18 and then inversing:

$$w = (E^T B)^+ C. \tag{2.19}$$

The reflectance of the test target is then estimated using

$$\tilde{R}_{(test)} = Bw = B(E^T B)^+ C_{(test)}. \tag{2.20}$$

Imai and Berns (1999) avoids the requirement of spectral properties of the channels and illumination to be known but rather assumes a linear relationship between camera responses and the weights that represent reflectance in a linear model:

$$w = MC, \tag{2.21}$$

where M is the transformation matrix which can be determined empirically via a least squares fit as

$$M = wC^+. \tag{2.22}$$

w is computed from Equation 2.18 as

$$w = B^{-1} R_{(train)} = B^T R_{(train)}. \tag{2.23}$$

The reflectance of the test target is then estimated using

$$\tilde{R} = Bw = BMC_{(test)} = BwC^+_{(train)} C_{(test)} = BB^T R_{(train)} C^+_{(train)} C_{(test)}. \tag{2.24}$$

- **Polynomial Model:** With this model (Connah and Hardeberg 2005; Bianco et al. 2008), the reflectance R of characterization dataset is directly mapped from the camera responses C through a linear relationship with the n degree polynomials of the camera responses:

$$\begin{aligned} R(\lambda_1) &= m_{11}C_1 + m_{12}C_2 + m_{13}C_3 + m_{14}C_1C_2... \\ R(\lambda_2) &= m_{21}C_1 + m_{22}C_2 + m_{23}C_3 + m_{24}C_1C_2... \\ \vdots &\quad \vdots \\ R(\lambda_N) &= m_{N1}C_1 + m_{N2}C_2 + m_{N3}C_3 + m_{N4}C_1C_2... \end{aligned} \tag{2.25}$$

It can be written in matrix form as

$$R = MC_p, \tag{2.26}$$

where M is the matrix formed from the coefficients and C_p is the polynomial vector/matrix from n degree polynomials of the camera responses as $(C_1, C_2, C_3, C_1^2, C_1C_2, C_1C_3, C_2C3...)^T$. The polynomial degree n is determined through optimization such that the reconstruction error is minimized. Complete or selected polynomial terms (for example, polynomial without crossed terms) could be used depending on the application. Transformation matrix M is determined from the training dataset using

$$M = RC^+_{p(train)}. \tag{2.27}$$

Substituting the computed matrix M in Equation 2.26 , the reflectance of the test target is estimated as

$$\tilde{R}_{(test)} = RC^+_{p(train)} C_p(test). \tag{2.28}$$

Since non-linear method of mapping camera responses onto reflectance values may cause over-fitting the characterization surface, regularization can be done as described in Subsection 2.3.5 to solve this problem.

- **Neural Network Model:** Artificial neural networks simulate the behavior of many simple processing elements present in the human brain, called neurons (Mansouri et al. 2005). Neurons are linked to each other by connections called synapses. Each synapse has a coefficient that represents the strength or weight of the connection. Learning or training is accomplished by adjusting these weights to cause the neural network to appropriate output results. The learning rules specify how to calculate the modification of the weights based on the objective. Figure 10 illustrates a basic one output perceptron (a type of neural network).

  Here, a_j is the activation of the jth output cell, $x_i$is the answer of the ith input cell and w_{ij} is the strength of connection between ith input cell and jth output cell. The answer o_j of the jth output cell, which is function of the neuron activation a_j, is only 0 or 1 in the case of deterministic perceptron. Behavior of the perceptron is modified by modifying the function f that translates the activation into a response. Advantage of neural networks model are, they are robust to noise. Mansouri et al. (2005) proposed robust spectral reconstruction algorithm based on hetero associative memories linear neural networks. Heteroassociative memories associate a stimulus in the input to a response in the output even if the two vectors have not the same size.

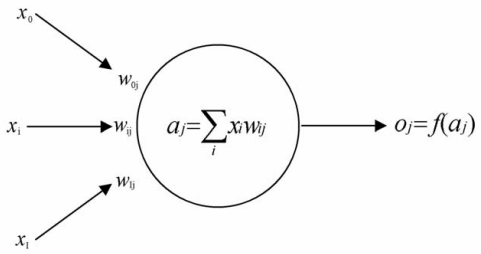


Figure 10: Basic one output perceptron

The neural network is trained with the training dataset using Delta rule also known as Widrow-Hoff rule. It continuously modifies w to reduce the difference (the delta) between the expected output value e and the actual output o of a neuron. This rule changes the connection weights in the way that minimizes the mean squared error of the neuron between an observed response o and a desired theoretical one like:

$$w_{ij}^{t+1} = w_{ij}^{t} + \eta(e_j - o_j)x_i = w_{ij}^{t} + \Delta w_{ij}, \tag{2.29}$$

where e is the expected response, t is the number of iteration,and η is a learning rate.

The weights w thus computed is finally used to estimate the reflectance of the test target using

$$\tilde{R} = wC_{(test)}. \tag{2.30}$$

The spectral reflectance, thus estimated from the camera responses, gives the device independent data of the imaged scene. An appropriate distance metric is used to evaluate accuracy of the reconstructed spectrum.

2.3.5 Regularization

Regularization introduces additional information in an inverse problem in order to solve an ill-posed problem or to prevent over-fitting. Non-linear method of mapping camera responses onto reflectance values is the potential for over-fitting the characterization surfaces. Over-fitting is caused when the number of parameters in the model is greater than the number of dimensions of variation in the data. Among many regularization methods, Tikhonov regularization (Dyas 2000; Gulliksson and Wedin 2000) is the most commonly used method of regularization. How Tikhonov regularization is done in pseudo-inverse computation using SVD (Singular Value Decomposition) is discussed here.

SVD states that any matrix M can be decomposed into three separate matrices as

$$M = USV^T, \tag{2.31}$$

where U and V are orthogonal matrices ($U^TU = VV^T = I$, where I is the identity matrix) and S is a diagonal matrix whose diagonal entries s_{ii} are known as singular values. The singular values decrease in value as they increase in index, i.e. $s_{11} > s_{22} > s_{33}$... etc. The pseudo inverse of M is computed directly using the singular value decomposition, so M^+ is given by

$$.M^+ = VS^{-1}UT \tag{2.32}$$

Here, the inverse S^{-1} has diagonal entries $1/s_{ii}$. If the value of s_{ii} becomes very small, then the reciprocal can become very large. As a result the solution for the coefficients can become very unstable, and especially sensitive to small amounts of noise. In turn, this instability can cause severe over-fitting of the data. Regularization is one way to solve this problem. Tikhonov regularization adds small positive values α to s_{ii} values prior to computing the inverse, such that $1/(s_{ii} + \alpha)$ does not become too large. The Equation 2.31 can then be written as:

$$M = U(S + \alpha I)V^T. \tag{2.33}$$

Since there is no pre-determined optimal value for the regularization parameter α, it must be determined experimentally. Regularization can also be done by truncating a number of last singular values in computing the pseud-inverse.

2.4 Evaluation of Multispectral Imaging System

Given a multispectral system, it is important to know how good the system's performance is. In general, performance of a multispectral system is measured by comparing the estimated reflectances or color values with the original or measured values using some distance metrics. Basically, there are two broad categories of evaluation metrics (distance metrics), colorimetric and spectral used in measuring the performance of multispectral systems.

2.4.1 Colorimetric Evaluation Metrics

These metrics are based on the measure of color difference between the estimated color and the original (or measured) color. There are several color difference formulas; for example, CIELUV, CIELAB, CIE94, CIEDE2000. The most commonly used is the CIELAB color difference (ΔE^*_{ab}), measured by the Euclidean distance in $L^*a^*b^*$ color space:

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}. \tag{2.34}$$

Major limitation of colorimetric metrics is that they produce bad correlation to spectral matches, particularly for metameric pairs (Imai et al. 2002).

2.4.2 Spectral Evaluation Metrics

This type of metrics is based on *Spectral Accuracy*, which is a measure of the similarity between the estimated spectra against that of the original (or measured) spectrum. The metrics used for spectral matches can impact everything from the selection of the filters used for multi-channel capture to the evaluation of the spectral estimation. Spectral metrics can distinguish between metamers but do not take into account human vision. There are several spectral evaluation metrics proposed in the literature (Imai et al. 2002). Root Mean Square Error (RMSE), Goodness of Fit Coefficient (GFC), Integrated Radiance Error (IRE) are some of the examples. Among these metrics, RMSE is the most commonly used spectral metric in multispectral imaging and hence used in this thesis as well.

Root Mean Square Error (RMSE) is the most commonly used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or measured. RMSE is a good measure of precision.

Denoting $R(\lambda)$ as the original/actual spectral value, $\tilde{R}(\lambda)$ as the reconstructed/estimated spectral value at wavelength λ and letting n be the number of wavelengths, the RMSE is computed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{R}(\lambda_i) - R(\lambda_i)\right]^2}. \tag{2.35}$$

The lesser the RMSE value, the better the system is, with zero value as the best. RMSE is normally expressed in percentage, meaning simply multiplying the value by 100 (not relative error).

2.5 Applications of Multispectral Imaging

Multispectral imaging is applicable to several domains (Hardeberg 1999), such as remote sensing (Swain and Davis 1978), astronomy (Rosselet et al. 1995), medical imaging (Taxt and Lundervold 1994), analysis of museological objects (Maitre et al. 1996), cosmetics (Doi et al. 2005), medicine (Farkas et al. 1996), high-accuracy color printing (Berns 1998; Berns et al. 1998), computer graphics (Peercy 1993), and multimedia (Jon et al. 1998). And many research groups worldwide are working on multispectral imaging, for example, at Rochester Institute of Technology, USA (Burns and Berns 1996; Burns 1997; Berns 1998; Berns et al. 1998), at RWTH Aachen, Germany (Keusen 1996; Koenig and Praefcke 1998; Hill 1998), at the university of Chiba, Japan (Haneishi et al. 1997; Yokohama 1997; Miyake and Yokoyama 1998), University of Joensuu, Finland (Hauta-Kasari et al. 1999; Kaarna and Parkkinen 2000; Toivanen et al. 2003) and Norwegian Color Research Laboratory, Gjøvik University College (Hardeberg 2004; Connah and Hardeberg 2005; Connah et al. 2006).

In addition to accurate imaging colorimetry, multispectral imaging has been employed for computational color constancy (Maloney and Wandell 1986; Tominaga 1996) and color image analysis (Maitre et al. 1996; Farina et al. 2000; Martinez et al. 2002; Levenson et al. 2003). Multispectral imaging can be applied to many different situations, including characterization of materials according to surface chemistry and characterization of component distribution in composite materials. It can also be used to remove irrelevant material from the analysis and find defects and foreign matter (Mehl et al. 2004; Kleynen et al. 2005; Blasco et al. 2007).

3 Stereo Image Acquisition

This chapter covers theories and state of the art technologies behind stereo imaging. Starting with stereovision and perception of depth, it presents stereo image acquisition methods, stereo camera model. It also presents the stereo camera calibration, stereo matching problem and estimating the disparity and in turn the depth from stereo images.

3.1 Stereo Vision and 3D Perception

We can see the world in three dimensions (3D) with our eyes. The obvious question arises, how can we see 3D view? Because of different positions of the two eyes in our head, they perceive the world in slightly different perspectives leading to different retinal images, with relative displacements of objects known as disparities in the two monocular views of a scene (Figure 11). These separate images are sent to the brain for processing. When the two images arrive simultaneously in the back of the brain, they are united into one picture. The mind combines the two images by matching up the similarities and adding in the small differences. The small differences between the two images add up to a big difference in the final picture! The combined image is more than the sum of its parts. It is a three-dimensional stereo picture. This process of perceiving depth in our visual perception is called *stereopsis* and such a vision is called *stereovision* or *binocular vision*. Thus, stereo vision is the cause by the fusion of the two views from our two eyes.

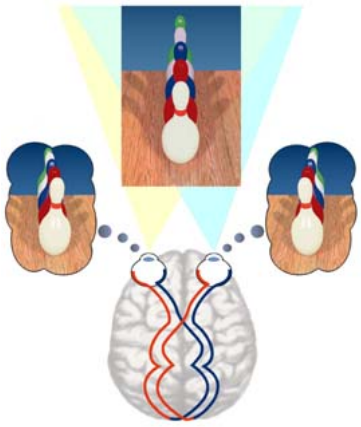


Figure 11: Depth perception from two eyes [Source: `http://www.vision3d.com/stereo.html`]

The word “stereo" comes from the Greek word “stereos" which means firm or solid. With stereo vision, we see an object as solid in three spatial dimensions: width, height and depth (or x, y and z). It is the added perception of the depth dimension that makes stereo vision so rich and special. The size and direction of the disparities of an object gives its relative depth. If we know the geometry of the imaging system, we can obtain the absolute depth-information. Depth is an important parts of image analysis in the study of computer vision.

Nowadays, stereo vision is not only used in mixing two images together but also in many other areas. For example, we use stereo vision on a robot, to simulate what human can see by their eyes. With this vision, robot can do some tasks for human.

3.2 Stereo Imaging

One of the limitations of the conventional color imaging with trichromatic camera is that the captured image is in 2D. Now our obvious curiosity and desire would be how if we can acquire images in 3D. And simple answer would be to try to emulate human stereo vision with a stereo camera. A Stereo camera is a camera (or two cameras rigged together) that has two lenses about the same distance apart as your eyes and takes two pictures at the same time. This emulates the way we actually see and therefore creates the 3D effect when viewed. Stereo photography or stereography, sets out to capture the two images as seen by the eyes, by using camera lenses in place of the eyes’ lenses and film or sensor in place of the retinas.

People have been putting together twin-camera rigs for doing 3D photography. Even big manufacturers like Fujifilm have already started producing commercial 3D stereo cameras. 3D imaging (images and videos) are gaining popularity and are of great use in imaging fields. There are International Stereoscopic Union (`http://www.isu3d.org`) and clubs for example, Stereo Club of Southern California (`http://www.la3dclub.com`), Sydney Stereo Camera Club (`http://oz3d.info`), Stereo Club Tokyo(`http://www.stereoclub.jp`), Stereo division in Photography Society of America (`http://www.psa-photo.org`), enthusiasts' personal web pages, educational, research and commercial sites for providing information and promoting stereo imaging.

An alternative approach to the traditional stereo using two cameras is active stereo. In an active stereo vision system, one of the cameras is replaced with a projector or a laser unit, which projects onto the object of interest a sheet of light at a time (or multiple sheets of light simultaneously). The term "active" here thus signifies that energy is projected into the environment. The perspective projection matrix of the camera and the equations of the planes containing the sheets of light relative to a global coordinate frame are computed from calibration and then the 3D coordinates of object points are computed by triangulation which simply involves finding the intersection of a ray from the camera and a plane from the sheet of light of the projector or laser. From the geometrical calibration of the system, the object points' positions and depth are calculated (Alexander and Ng 1987). Batlle et al. (1998); Scharstein and Szeliski (2003); Griesser et al. (2004); Hirschmuller (2006) are some of the example works in this area. This is not within the scope of this thesis work.

In Section 3.3, we present simplified stereo imaging model and explain how to estimate depth from two stereo images.

3.3 Stereo Imaging Model and Depth Estimation

The underlying principle of binocular vision is that of triangulation. Given a single image, the three-dimensional location of any visible object point must lie on the straight line that passes through the center of projection and the image of the object point (Figure 12). The determination of the intersection of two such lines generated from two independent images is called triangulation.

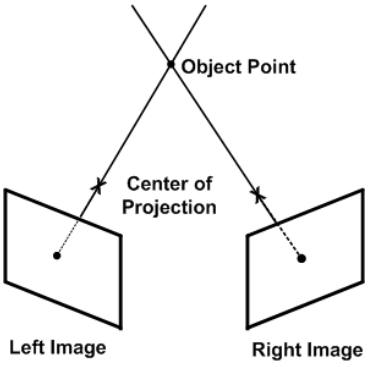


Figure 12: Principle of triangulation in stereo imaging

Figure 13 shows the stereo imaging model with two cameras, where $Image_l$ is taken by the left camera and $Image_r$ is taken by the right camera. $x_l y_l z_l$ is the coordinate system of the left camera ($camera_l$) and $x_r y_r z_r$ is the coordinate system of the right camera ($camera_r$) and XYZ is the 3D world coordinate system. Assume that the world coordinate system is aligned with the left camera coordinate system and the two image planes are coplanar along the x-axis. Two cameras are separated by distance b. Let the image point in the $camera_l$ corresponding to any world coordinate point P(X,Y,Z) is $P_l(x_l, y_l)$ and the corresponding image point in the $camera_r$ is $P_r(x_r, y_r)$. Both the cameras are assumed to be identical with the same resolution and same focal length of f. If we let the aperture of a thin lens decrease to zero (also known as Pinhole Camera Model), all rays are forced to go through the optical center, and therefore they remain undeflected. From this model, we can estimate depth of stereo imaging. This imaging setup can be seen as a section in XZ plane as shown in the Figure 14. With this setup, any world point P will have same Y and Z values corresponding to the two camera coordinate systems, but different X values because of the camera separation. Let these two X values be denoted as X_l and X_r respectively. Now from the geometry (similar triangles), the following relations can be obtained.

$$\frac{-x_l}{X_l} = \frac{f}{Z-f} \quad => \quad X_l = \frac{x_l}{f}(f - Z) \tag{3.1}$$

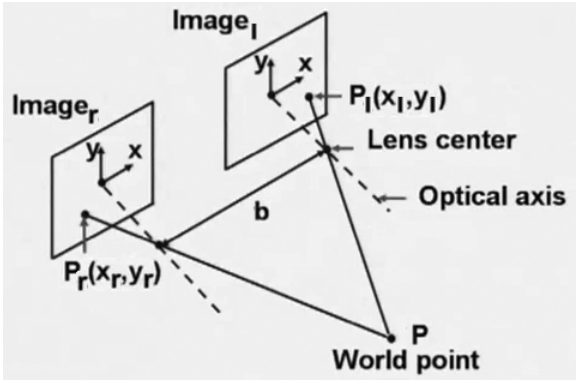


Figure 13: Stereo imaging model

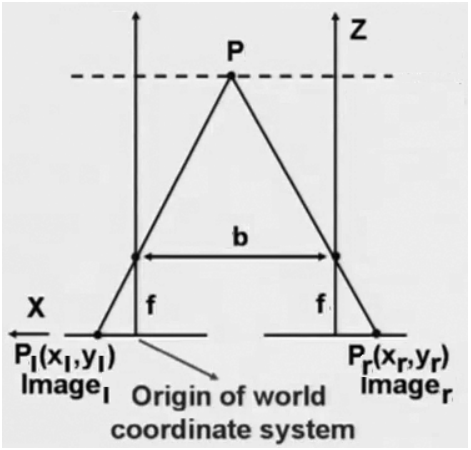


Figure 14: Sectional view of the model

$$\frac{x_r}{-X_r} = \frac{f}{Z-f} \quad => \quad X_r = \frac{x_r}{f}(f-Z) \tag{3.2}$$

Since two cameras are separated by distance b,

$$X_r = X_l + b. \tag{3.3}$$

Substituting values of X_l and X_r from Equations 3.1 and 3.2 respectively, we get

$$Z = f - \frac{fb}{x_r - x_l}. \tag{3.4}$$

This gives the Z-coordinate of the point P, which is the same for camera$_l$ coordinate system, camera$_r$ coordinate system as well as for the world coordinate system. Here, $x_r - x_l$ is the disparity of the two cameras. Thus, knowing the camera separation, disparity and the focal length of the camera, we can compute the depth of the image point. Knowing Z, we can compute X and Y coordinates from the following equations:

$$X = \frac{x}{f}(f-Z), \tag{3.5}$$

and,

$$Y = \frac{y}{f}(f-Z). \tag{3.6}$$

Here, in order to compute the depth, the important parameter to be measured is the disparity. The correct and fast estimation of disparities is a difficult problem. The problem is that given a point in the left image, what is the corresponding point in the right image?. This problem is called *stereo correspondence* problem or *stereo matching* problem. Section 3.6 discusses on this.

Thus, reconstruction of the world seen through stereo cameras can be divided in two steps:

- *Correspondence/Matching problem:* For every point in one image, find out the corresponding point on the other and compute the disparity of these points.
- *Triangulation:* Given the disparity map, the focal distance of the two cameras and the geometry of the stereo setting (relative position and orientation of the cameras), compute the (X,Y,Z) coordinates of all points in the images.

The simple stereo imaging model discussed here gives a general idea on how we can estimate depth from two stereo images. Some of the assumptions made in the model may not be valid in real case, for example the two image planes of the two cameras may not be in the same plane and the focal lengths of the two cameras also may not be the same. Moreover, there will be radial and tangential lens distortion and skewness problem. Figure 15 shows a generalized stereo imaging model where two cameras are not aligned. In such case, we need to determine the relative position and orientations of the two cameras and then align them through geometric transformations (translation, rotation and scaling) for depth estimation. These parameters are determined by means of camera calibration. Calibration of a stereo camera is the task of relating the ideal pinhole model of the camera with an actual device (internal calibration) and retrieving the relative position and orientation of the cameras (external calibration). The Section 3.4 discusses on camera calibration.

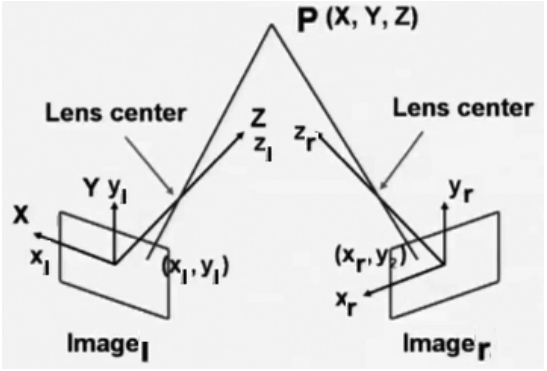


Figure 15: Generalized stereo imaging model

There are, in general, two ways of extracting three-dimensional structure from a pair of images: calibrated route and uncalibrated route (Fisher 2007).

1. *Calibrated route:* In this method, both the cameras are first calibrated with respect to some world coordinate, then so called epipolar geometry is calculated by extracting the essential matrix of the system, and from this, the three dimensional Euclidean structure of the imaged scene is computed. The Section 3.4 discusses on camera calibration.

2. *Uncalibrated or Self-calibrated route:* Self-calibration refers to the process of calculating all the intrinsic parameters of the camera using only the information available in the images taken by that camera. This method corresponds to the way in which biological systems determine 3D structure from vision. No calibration frame or known object is needed: the only requirement is that there is a static object in the scene, and the camera moves around taking images. All that is assumed is that we have a stereo camera system that is capable, by comparing the two images, of establishing some correspondence between them. Each such correspondence, written $(\mathrm{m}, \mathrm{m}')$, indicates that the two image points m and m' are very likely to be the images of the same world point M. Thus, the system does not know its intrinsic and extrinsic parameters. This is known as the uncalibrated system. Here, a quantity known as the fundamental matrix, which is a 3×3 matrix that relates corresponding points in stereo images, is calculated from image correspondences and this is then used to determine the projective 3D structure of the imaged scene. This is ideal for a mobile camera, such as a camera mounted on a mobile robot. The actual camera movement itself does not need to be known. Faugeras (1993); Deriche et al. (1994); Pilu (1997) are some of the example works in this direction. Kuffar and Takaya (2007) demonstrates the use of the epipolar geometry to estimate the depth (distance) to a point on the object structure in a pair of stereo images.

In both approaches the underlying principle of binocular vision is that of triangulation. Given a single image, the three-dimensional location of any visible object point must lie on the straight line that passes through the center of projection and the image of the object point. The determination of the intersection of two such lines generated from two independent images is called triangulation.

3.4 Camera Calibration

Camera calibration, often referred to as camera resectioning, is a way of examining an image, or a video, and deducing what the camera situation was at the time the image was captured. The objective of camera calibration is to find the *external/extrinsic parameters* (position and orientation relatively to a world co-ordinate system), and the *internal/intrinsic parameters* (principal point, focal length, skew coefficients and distortion coefficients (Brown 1966; Fryer and Brown 1986)) of the camera. The extrinsic parameters also describe the relationship between the cameras.

Several camera calibration algorithms have been proposed (Tsai 1986; Faugeras 1993; Kumar and Hanson 1994; Heikkila and Silven 1997; Fitzgibbon and Zisserman 1998; Zhang 2000; Kurazume et al. 2002). One of the most used camera calibration techniques is the one proposed by Tsai (Tsai 1986, 1987) as it is suitable for a wide area of application since it can deal with coplanar and non-coplanar points. It also offers the possibility to calibrate internal and external parameters separately. This option is particularly useful since it gives the possibility to fix the internal parameters of the camera, when known, and carry out only pose estimation.

If we know the internal camera parameters (provided by the manufacturer or computed using a known target, usually a checkerboard plane), the situation called pose estimation, we just need to recover the six parameters relative to the position and orientation of the camera. Methods proposed in Heikkila and Silven (1997) and Zhang (2000) can be used in this situation. Some methods for pose estimation as well as a sensitivity analysis can be found in Kumar and Hanson (1994).

We are discussing here the model from Bouguet (2006), as it has been designed on top of Zhang (2000) and Heikkila and Silven (1997). The Subsections 3.4.1 and 3.4.2 respectively present how intrinsic parameters and extrinsic parameters are computed with this model.

3.4.1 Intrinsic Parameters

There are five intrinsic parameters in a general camera model. They are:

- *Focal length:* It is the focal length of the camera. It is denoted by the 2×1 vector, $fc = [f_x, f_y]^T$. The ratio f_y/f_x, often called *aspect ratio*, is different from 1 if the pixel in the sensor array are not square. The values are expressed in units of horizontal and vertical pixels.

- *Principal point:* Point of intersection of the principal axis with the image plane is called the principal point. It is denoted by the 2×1 vector, $cc = [c_x, c_y]^T$.

- *Skew coefficient:* The skew coefficient defining the angle between the x and y pixel axes is denoted by the scalar, α_c. Thus pixels are even allowed to be non-rectangular. Some authors refer to that type of model as *affine distortion* model. For new cameras where pixels are assumed to be rectangular, the value of $\alpha_c = 0$.

- *Distortions:* In general, two types of distortions are considered: *radial* and *tangential*. Radial distortion is a failure of a lens to be rectilinear: a failure to image lines into lines. The tangential distortion is due to decentering, or imperfect centering of the lens components and other manufacturing defects in a compound lens. Brown (1966) and Fryer and Brown (1986) introduced these distortion models which are used here. The image distortion coefficients (both radial and tangential) are denoted by the 5×1 vector d. The coefficient of 6^{th} order radial distortion term is the fifth entry of the vector d. For the lenses which do not have imperfection in centering, the tangential component of distortion can be discarded. Moreover, it is often not necessary to push the radial component of distortion model beyond the 4th order (as used by Zhang (2000)); in such case $d(5) = 0$.

Let P be a point given by the vector $XX_c = [X_c, Y_c, Z_c]^T$ in camera reference coordinate. Projecting this point on the image plane according to the intrinsic parameters (f, c, α, d), the (pinhole) image projection x is given by

$$x = \frac{f}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix}. \tag{3.7}$$

The image projection in normalized form x_n is then given by

$$x_n = \begin{bmatrix} \frac{X_c}{Z_c} \\ \frac{Y_c}{Z_c} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}. \tag{3.8}$$

After including lens distortion, the new normalized point coordinate x_d is defined as follows:

$$x_d = \begin{bmatrix} X_d(1) \\ X_d(2) \end{bmatrix} = \left(1 + d(1)r^2 + d(2)r^4 + d(5)r^6\right) x_n + dx, \tag{3.9}$$

where $r^2 = x^2 + y^2$, and dx is the tangential distortion vector:

$$dx = \begin{bmatrix} 2d(3)xy + d(4)(r^2 + 2x^2) \\ d(3)(r^2 + 2y^2) + 2d(4)xy \end{bmatrix}. \tag{3.10}$$

The final pixel coordinate (x_p, y_p) of the projection of P on the image plane, obtained after applying the distortion is given by

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} f_x \left(x_d(1) + \alpha_c x_d(2) + c_x\right) \\ f_y x_d(3) + c_y \end{bmatrix}. \tag{3.11}$$

It can be expressed in augmented homogeneous matrix form as

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \alpha_c f_x & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d(1) \\ x_d(2) \\ 1 \end{bmatrix}. \tag{3.12}$$

In order to compute intrinsic parameters from this relation, a set of checkerboard calibration images (normally 20) in different orientations are used and the parameters are computed by optimization process using iterative gradient descent with an explicit (closed-form) computation of the Jacobian matrix. Subsection 3.4.3 presents the calibration steps to be followed.

3.4.2 Extrinsic Parameters

The two extrinsic parameters to be estimated in camera calibration are:

- *Rotation:* It is defined by a 3×3 matrix, R.
- *Translation:* It is given by 3×1 translation vector, T.

If the point P is given by the vector $XX = [X, Y, Z]^T$ in the object reference coordinate, then XX and XX_c are related to each other through the following rigid motion equation:

$$XX_c = R.XX + T. \tag{3.13}$$

It can be expressed in augmented homogeneous coordinates form as

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.14}$$

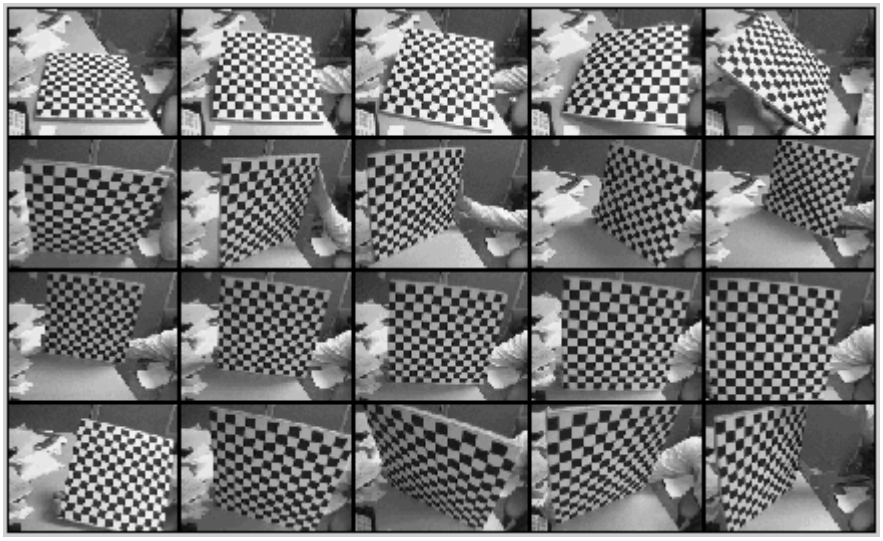


Figure 16: Calibration images [Source: Bouguet (2006)]

Once the coordinates of a point is expressed in the camera reference frame, it may be projected on the image plane using the intrinsic camera parameters. Rotation vector Om is also used to express rotation parameter, which is related to the rotation matrix R through the Rodrigues rotation formula: $R = \text{rodrigues}(Om)$. Rodrigues' rotation formula (see Wikipedia (2010*b*) for details) gives a convenient way to write the general rotation matrix. There are many other camera calibration techniques and tools available in the academic world. The links to many of the well known tools are given in Bouguet (2006).

3.4.3 Calibration Steps

In computer vision, we usually perform calibration by analyzing images taken from camera. It involves checkerboard (or some other planar calibration pattern) which should have distinct width and height (measured in checkerboard squares) and be even. Normally 20 images of checkerboard are taken in distinct orientations as shown in Figure 16. Calibration software then finds inner corners of the checkerboard on every image and then uses them together with information about real-world size of checkerboard to determine intrinsic camera parameters. The corners are extracted to an accuracy of about 0.1 pixel. This is done in two steps: initialization and then nonlinear optimization.

First, it computes the initial intrinsic parameters. The distortion coefficients are all set to zeros initially. Then the initial camera pose is estimated as if the intrinsic parameters have been already known. The initialization step computes a closed-form solution for the calibration parameters, not including any lens distortion. The non-linear optimization step then minimizes the total reprojection error (in the least squares sense) over all the calibration parameters. The optimization is done by iterative gradient descent (see Wikipedia (2010*a*) for details) with an explicit computation of the Jacobian matrix.

Knowing the intrinsic parameters, extrinsic parameters can also be computed from the Equation 3.14.

3.5 Stereo Camera Calibration

Stereo calibration is important for accuracy in 3D reconstruction. Calibrating stereo cameras is usually dealt with by first calibrating each camera independently, and then applying geometric transformation of the external parameters to find out the geometry of the stereo setting. The two cameras are calibrated separately as discussed in Section 3.4 using left and right camera images from the same checkerboard calibration images. We thus obtain the intrinsic parameters: intrinsic parameters fc_l, cc_l, αc_l, d_l and fc_r, cc_r, αc_r, d_r corresponding to left and right cameras respectively. From the extrinsic parameters R_l, T_l and R_r,T_r, we can compute rotation and translation parameters R and T that defines the position of the right camera with respect to the left camera.

With all these information of stereo configuration, we can recompute the internal and external parameters for either left or right or both together (as desired) so as to minimize the reprojection errors on both camera for all calibration grid locations. This will give smaller uncertainties on the intrinsic parameters (especially that of the focal values) due to the fact that the global stereo optimization is performed over a minimal set of unknown parameters.

Then if we consider a point P in 3D space, its two coordinate vectors $Xl = [X_l, Y_l, Z_l]^T$ and $Xr = [X_r, Y_r, Z_r]^T$ in the left and right camera reference frames respectively, they are related to each other through the rigid motion transformation:

$$Xr = R * Xl + T. \tag{3.15}$$

Let $xl = Xl/Zl = [x_l, y_l, 1]^T$ and $xr = Xr/Zr = [x_r, y_r, 1]^T$ be the coordinate vectors of perspective projections Pl and Pr of P on the two camera planes. Now, the problem, known as triangulation, consists of obtaining Xl and Xr from xl and xr. Equation 3.15 may be written as

$$xrZr = R * xlZl + T. \tag{3.16}$$

This leads to the relation:

$$\begin{bmatrix} -Rxl & xr \end{bmatrix} \begin{bmatrix} Zl \\ Zr \end{bmatrix} = T. \tag{3.17}$$

Let $A = \begin{bmatrix} -Rxl & xr \end{bmatrix}$ (a 3×2 matrix). The least square solution for this equation is then given by

$$\begin{bmatrix} Zl \\ Zr \end{bmatrix} = \left(A^T A\right)^{-1} A^T T. \tag{3.18}$$

Denoting $\alpha_l = -Rxl$, an explicit expression for Zl and Zr may be expanded as

$$Zl = \frac{\|xr\|^2 \langle \alpha_l, T \rangle - \langle \alpha_l, xr \rangle \langle xr, T \rangle}{\|xl\|^2 \|xr\|^2 - \langle \alpha_l, xr \rangle^2}, \tag{3.19}$$

and,

$$Zr = \frac{-\|xl\|^2 \langle xr, T \rangle + \langle \alpha_l, xr \rangle \langle \alpha_l, T \rangle}{\|xl\|^2 \|xr\|^2 - \langle \alpha_l, xr \rangle^2}. \tag{3.20}$$

where $\langle ... \rangle$ is the standard scalar/dot product operation. Equations 3.19 and 3.20 give the depth values of the image with reference to the two cameras. Thus, in order to determine the depth, we need to first determine corresponding points in the left and right images. This is done through stereo matching algorithms.

3.6 Stereo Matching

Stereo matching or correspondence problem is one of the important problems in stereo imaging. It tries to find corresponding points in the left and right images in order to measure disparity. However, it is not just the small displacements of image parts. Besides, numerous other image variations occur between the two stereo views, making the reliable computation of disparities more difficult. The varying view geometry leads to a variety of systematic image variations, including occlusion effects. Sensor noise and different transfer functions of the left and right cameras may introduce stochastic signal variations. In addition, since most object surfaces in real-world scenes display specular reflection, the intensities observed by the imaging systems are not directly correlated with the object surfaces, but nearly always have a viewpoint dependent component, which moves independently of the surface in question.

This problem has been studied over several decades in computer vision and many researchers have proposed several methods to tackle this whole set of distorting signal variations.

3.6.1 Stereo Matching Methods

There are mainly three types of stereo matching methods: area-based or intensity-based, feature-based and phase-based methods. These methods are discussed briefly in the subsequent subsections.

- *Area-based/Intensity-based Stereo Matching*

  Area-based approaches compare directly intensity values within small image patches of the left and right view, and try to maximize the correlation between these patches.

  Classical area-based approaches minimize the deviation or maximize the correlation between patches of the left and right view. A large enough patch size assures a stable performance (via the central-limit theorem). The technique for scoring region matching varies, but usually is based on the number of pixels that are the same on an exact or near-exact point basis. Zitnick and Kanade (1999) successfully implemented stereo image matching and technique for point-matching within a region.

  The computation of the correlation measure, and then maximization of this value in area-based approach are computationally expensive processes, since extensive search is required in configuration space. This problem is usually solved with hierarchical approaches, where coarse-to-fine multiresolution matching scheme is used for disparity estimation. The matches obtained at the coarser level are used to guide the matching process gradually up to the finest level. For example, Hannah (1985); Marapane and Trivedi (1994); Hung et al. (1998) used this scheme. However, for generic image data there is no guarantee that the disparity information obtained at the coarse scales is valid. The disparity estimate might be wrong, might have a different value than at finer scales, or might not be present at all. Thus hierarchical approaches will fail under various circumstances.

  Luo et al. (1995); Koo and Jeong (2001); Sun (2002); Stefano et al. (2004); Wang and Zheng (2008) are some more example works in the area-based stereo matching.

- *Feature-based Stereo Matching*

  In feature-based stereo algorithms, the intensity data is first converted to a set of features that are assumed to be a more stable image property than the raw intensities. The matching stage operates only on these extracted image features.

  There are many different types of features and thus the obvious question here would be what type of features that one should use? Edge elements, corners, line segments, and curve segments are features that are robust against the change of perspective, and they have been widely used in many stereo vision work. Edge elements and corners are easy to detect, but may suffer from occlusion; line and curve segments require extra computation time, but are more robust against occlusion (they are longer and so are less likely to be completely occluded). Higher level image features such as circles, ellipses, and polygonal regions have also been used as features for stereo matching, these features are, however, restricted to images of indoor scenes.

  Most feature-based stereo matching systems are not restricted to using only a specific type of features, instead, a collection of feature types is incorporated. For instance, the system proposed by Weng et al. (1988) combines intensity, edges, and corners to form multiple attributes for matching; Lim and Binford (1987), on the other hand, used a hierarchy of features varying from edges, curves, to surfaces and bodies (2D regions) for high-level attribute matching. Recent sparse stereo approaches generally make use of affine transform invariant features such as Maximally-Stable Extrema Regions (MSER) (Matas et al. 2004), Scale Invariant Feature Transform (SIFT) operator (Lowe 2004), or Speeded Up Robust Features (SURF) (Bay et al. 2006).

  Some other example works in feature-based stereo matching are Horaud (1989); Goulermas and Liatsis (2000); Tang et al. (2006).

The problem with this method is only a few specific feature-classes are generally utilized in feature-based algorithms. Therefore, most areas of the stereo images end up in the "no feature present"-class, which is not considered further in the matching process. This leads to a tremendous data reduction which speeds up processing, but makes it impossible to calculate disparity estimates for most of the image regions. In order to obtain dense disparity maps, one is forced to interpolate these missing values.

To complicate matters further, every feature detected in the left image can potentially be matched with every feature of the same class in the right image. This is the classical "false matches" problem, which is basic to all feature-based stereo algorithms. The problem can only be solved by the introduction of one or more additional constraints to the final solution of the matching problem. The stereo matching constraints is discussed in Subsection 3.6.2.

- *Phase-based Stereo Matching*

  A third way for calculating disparities is known as phase-based methods (Sanger 1988; Jenkin and Jepson 1994). These approaches derive Fourier-phase images from the raw intensity data. Extraction of the Fourier phase can be considered as a local contrast equalization reducing the effects of many intensity variations between the two stereo views. Algorithmically, these approaches are in fact gradient-based optical flow methods, with the time-derivative approximated by the difference between the left and right Fourier-phase images (Barron et al. 1994). The Fourier phase exhibits wrap-around, making it again necessary to employ hierarchical methods, which has drawbacks already discussed. Furthermore, additional steps have to be taken to ensure the exclusion of regions with ill-defined Fourier phase (Fleet and Jepson 1993).

- *Other Methods*

  There are many other methods proposed in the literature which tried to improve the existing classical methods. For example, Henkel (1997) proposed an approach based on the detection of coherence between simple disparity units modeling complex cells in human visual cortex, and also proposed a neural network approach to stereovision in Henkel (1998, 2000). Li et al. (2007) proposed multi-view edge-based stereo by incorporating spatial Coherence. Frisby (1998) and Goulermas and Liatsis (2000) are some other example works.

3.6.2 Stereo Matching Constraints

The stereo matching process is a very difficult search procedure. In order to minimize false matches, some matching constraints must be imposed. These constraints are usually derived from reasonable assumptions about the physical properties of object surfaces, and rule out certain combination of matches. Classical constraints include the uniqueness of a match, figural continuity and the preserved ordering of matches along horizontal scan lines (Frisby 1998). In conjunction with the features extracted from the images, constraints define a complicated error measure which can be minimized by direct search techniques or through cooperative processes. A list of the commonly used constraints as pointed out by Fisher (2007) is given below.

- *Similarity or compatibility*: For the intensity-based approach, the matching pixels must have similar intensity values (i.e. difference lower than a specified threshold) or the matching windows must be highly correlated. For the feature-based approach, the matching features must have similar attribute values (Grimson 1981).

- *Uniqueness*: Almost always, a given pixel or feature from one image can match no more than one pixel or feature from the other image (Marr and Poggio 1979). The uniqueness constraint may not be applicable to the line segment-based approach. This constraint can also fail if transparent objects are present in the scene. Furthermore, given a pixel (or feature) in one image, its corresponding pixel (or feature) may be occluded in the other image. In this case, no match should be assigned to that pixel.

- *Continuity*: The cohesiveness of matters suggests that the disparity of the matches should vary smoothly almost everywhere over the image (Marr and Poggio 1979). This constraint fails at discontinuities of depth, for depth discontinuities cause an abrupt change in disparity.

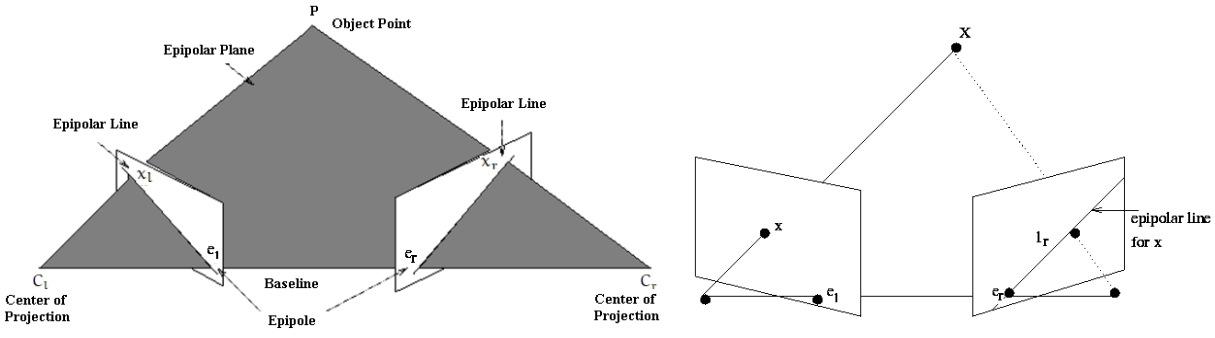
- *Ordering*: The ordering of features is preserved across images (Baker 1981).

- *Relaxation*: It is a global matching constraint to eliminate false matches (Barnard and Thompson 1980). An example of applying relaxation to stereo matching: assign to each candidate match a probability value based on some criteria on the "goodness of match"; iteratively update this probability value for each candidate match; finally, delete those matches whose probability value is below a certain threshold. The update process in each iteration is as follows: increase probability of a given candidate match by a value that is proportional to the number of neighboring matches that have consistent disparity values.

- *Epipolar*: The epipolar constraint states that given a feature point in the left image, the corresponding feature point must lie on the corresponding epipolar line (Deriche et al. 1994). This constraint reduces the search space from two-dimensions to one-dimension. This can be explained with Figure 17(a).

  The *epipole* is the point of intersection of the line joining the optical centers, that is the baseline, with the image plane. Thus the epipole is the image, in one camera, of the optical center of the other camera. The *epipolar plane* is the plane defined by a 3D point P, and the optical centers C_l and C_r. The *epipolar line* is the straight line of intersection of the epipolar plane with the image plane. It is the image in one camera of a ray through the optical center and image point in the other camera. All epipolar lines intersect at the epipole.

  Thus, a point x in one image generates a line in the other on which its corresponding point must lie. We see that the search for correspondences is thus reduced from a region to a line. This is illustrated in Figure 17(b).



(a) Epipolar constraint

(b) The epipolar line along which the corresponding point for x must lie

Figure 17: Epipolar geometry [Source: Fisher (2007)]

Unlike all other constraints, the epipolar constraint would never fail and could be applied reliably once the epipolar geometry is known.

3.6.3 Open problems in Stereo Matching

The complexity of the correspondence problem depends on the complexity of the scene. As described in the previous sections, there are constraints and schemes that can help reduce the number of false matches, but many unsolved problems still exist in stereo matching. Some of these problems as pointed out in Fisher (2007) are:

- *Occlusion:* Given a feature in the first image projected by a scene point P, the “corresponding” feature of P may not exist in the second image because from the second camera’s viewpoint, P may be occluded.

- *Discontinuity:* This is related to occlusion. Most matching algorithms use the continuity and/or relaxation as a global constraint to reduce false matches. This constraint works everywhere in the images except at depth discontinuities.

- *Regularity and Repetitivity:* Repetitive patterns in the scene cause ambiguity in matching. Most existing matching algorithms are still not robust enough to handle this.

3.7 Stereo Image Acquisition

Based on the number and types of cameras used, stereo image acquisition methods can be categorized into four types (ISU 2010):

1. *Using a single digital camera:* The simplest method for taking a 3D photo requires just a single camera, and a stationary subject. A picture is taken first, by pacing feet firmy on the ground, with the weight on the left foot. Then,another picture is taken by shifting the weight to the right foot. We now have a stereo pair of images, one for the left eye and one for the right eye, which can be viewed in 3D. More reliable results can he achieved if the camera is mounted on a tripod topped with a slide-bar which facilitates accurate movement between exposures (Figure 18). Obviously, these techniques, called “sidestep" or “cha-cha" 3D, only works for subjects that are not in motion. And, quantitive studies like depth estimation would not be possible as there is no precise geometry involved.

2. *Using a beam splitter:* The beam-splitter (more correctly called an image-splitter) is a device which attaches to the front of the lens of an ordinary digital camera and by means of mirrors or prisms produces two images of the same scene side-by-side on a single frame of film. The advantages of this technique are: simplicity, economy and reliability.

   There are, however, substantial disadvantages as well. The device introduces distortions in opposite directions in the two images because of one eye-view effectively passing through one side of the lens, and the other eye-view passing through the other side of the lens. There is a risk of ghosting or flare due to the scattering of light in the mirror system.

3. *Using a single stereo camera:* Stereographic cameras have been manufactured for almost as long as photography has existed as a technique. However, there were no commercial digital stereo cameras until recently. With gaining popularity and use, manufacturers have started producing 3D cameras. We have found two commercial stereo cameras in the market: Fujifilm FinePix REAL 3D W1 (Figure 19(a), `http://www.fujifilm.com/products/3d/camera/finepix_real3dw1/`) and 3D VuCam (Figure 19(b), `http://www.3dstereo.com/viewmaster/cam-3dvu.html`). We have used the first one, which is claimed as the world’s first 3D digital camera, in this thesis work as the quality of images it captured is better and also less prone to noise.

Figure 18: Stereo photography with single camera

(a) Fujifilm FinePix REAL 3D W1

(b) 3D VuCam

Figure 19: Two commercial stereo cameras

4. *Using two digital cameras:* Stereo enthusiasts design their own stereo camera by putting two cameras together on a support bar or by joining them with twin-camera rigs. The distance between the cameras is important for the impression of depth. The distance between adult human eyes (interpupillary separation) varies between possible extremes of 55 mm and 75 mm, with an average of around 63.5 mm. Since the viewing of stereographs requires the brain to carry out the same processes as in normal vision, it is essential that the information supplied to it matches so far as is possible as in normal binocular vision. Any excessive separation or misalignment of the two images during viewing causes the brain great problems, which manifest themselves as double vision (diplopia) eyestrain and headaches. So, it is recommended to separate two cameras by 63.5mm.

   Usually this method requires technical expertise and a lot of work as the two shutters need to be released simultaneously. Specialized synchronization software like Stereo Data Maker is available for this; list of cameras supported by the software can be found from its website `http://www.stereomaker.net`. Using such software, a controller can be designed by the user by wiring together the two cameras. Such a controller can control camera actions including external flash electrically through lanc ports, and thus allows us to synchronize image capture from two cameras lesser than one millisecond. Do it yourself tutorial entitled “How to Build Your Own 3D Camera Rig for Under $20" by Kurland (2009) explains the construction of the controller and the stereo camera itself. Commercial controllers are also available in the market. One example is Pokescope Digital 3D Camera Controller (`http://www.pokescope.com`).

   The two cameras may be joined in different orientations: horizontal (side-by-side) or vertical (base-to-base) as shown in Figures 20(a) and 20(b) respectively. The second orientation allows us to reduce the lens separation. But it has a number of disadvantages, chief among them are the vertical format, which is not well-suited to many stereo subjects, and the limited scope for reducing the width of the images to form a suitable stereo window.

   Even with two cameras of the same type, differing focal lengths and other variations may occur unless the lenses are specially selected as matching pairs, resulting in discrepancies between the two images which, although not too troublesome in a hand-viewer, can make for difficult viewing by projection. Under no circumstances should separate flash units be used on each camera in order to overcome lack of synchronization, as the resulting variations in shadow positions render it almost impossible to view the stereo pair satisfactorily.

Quite a work has been done in stereo image acquisition. Ross (1993) proposed a custom built high-speed, physically robust stereo ranging system to be used on trinocular stereo rig and three specially modified CCD cameras. Stereo matching was performed using the sum-of-sum-of-squared differences (SSSD) technique. Like 3D images, 3D movies can also be taken using two movie cameras similarly joined in a stereoscopic configuration. In this thesis, our focus is on still image acquisition only.

(a) Horizontal orientation (b) Vertical orientation

Figure 20: Two different orientations of cameras

3.8 Golden Rules of Stereo Photography

Gerhard P. Herbig phrased *Golden Rule* to present three 3 basic rules which must be considered for highest stereo enjoyment. The three rules concern with shooting, mounting and viewing of the stereo image respectively. The rules and the reasons behind them are given below as it is (ISU 2010).

1. *The depth information in a stereo image must not exceed certain limits*. Since our brain can process only a limited amount of depth information, the depth contained in a stereo image must be limited so that the entire stereo picture can be observed as a uniform spatial image. A violation of the shooting rule leads to a decay of the spatial image into individual image sections. If we want a stereo photo that is pleasant to view, adherence to the shooting rule is a condition which should be observed under all circumstances.

2. *No part of the spatial image may be cut by the stereo window*. With stereo projection (and when looking into a hand-held stereo viewer) the spatial image is usually seen behind an apparent window. This window is called the stereo window. This stereo window is actually the spatial image of the stereo frame. Whether the image scenery can be seen in front of or behind the stereo window depends exclusively on the position of the single film chips within the stereo mounts.

   The stereo window is usually seen like a firm wall. When the mounting rule is violated, the spatial image is often too far in front of the stereo window. Parts of the picture seem to penetrate or “impinge on" the wall. This “erroneous image" clearly causes reduced enjoyment when looking at spatial images. However, the mounting rule is a can-rule, i.e. it should be kept, but it doesn’t have to be kept under all circumstances.

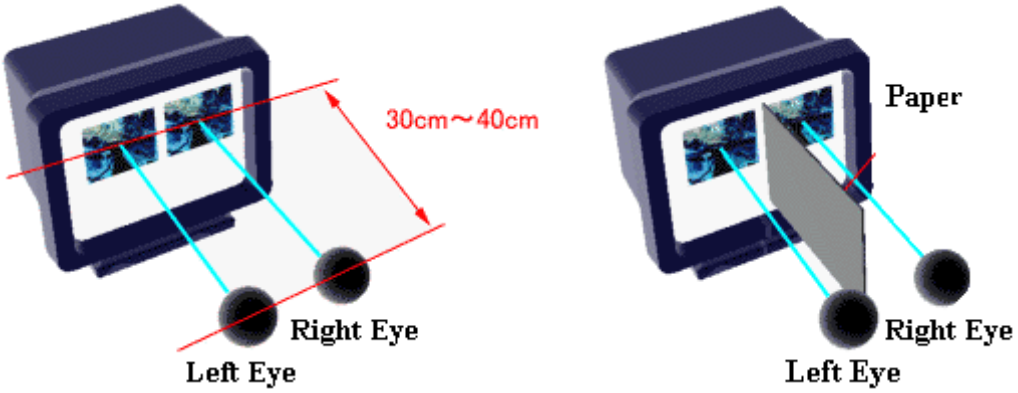
3. *No person viewing a projected stereo image should be forced to diverge his eyes*. When looking in the distance the viewing axes of the eyes are parallel, whereas looking at near objects the eyes converge inwards a bit. With natural vision, diverging viewing axes never occur. If forced to look at stereo images using diverging viewing axes for a longer time, pounding headaches are the inevitable result. Even if some people are capable of fairly amazing eye gymnastics, looking at stereo image should be enjoyable for every viewer. Therefore the projection rule has to be enforced under all circumstances in any stereo projection.

More questions and answers concerning stereo photography, mounting and stereo projection are given in (Herbig 2005).

3.9 Viewing Stereo Images

The two images taken with a stereo camera are mounted so that when presented to the eyes, either directly in a viewer which focuses the images and ensures that each eye sees only its appropriate image, or indirectly by projecting on a screen using polarized light and polarizing spectacles to ensure the two images on the screen retain their separate identities, the brain then integrates into a single stereo image in the normal way. There are several different ways of viewing stereo images (Stereoeye.jp 2010). We will discusse major techniques below.

1. *Parallel Viewing:* In this viewing, two images are displayed on the screen as in Figure 21(a) and then the left image is viewed with the left eye and the right image is viewed with the right eye. At first, the viewer looks at the wall behind the screen and then looks down to the stereo photograph slowly without focusing soon. Then, two picture will be one over another and 3D image will be seen. It may need bit of practice as it is not easy to view in this way. The view can be made bit easier by placing a piece of paper between the screen and the face separating two eyes and the two images as shown in the Figure 21(b). It is to be noted that eyes should be kept at a distance of more than 15 inch from the screen and head is not tilted while viewing.



(a) (b)

Figure 21: Parallel viewing [Source: Stereoeye.jp (2010)]

2. *Cross Eyed Viewing:* In this viewing, the left image is viewed with the right eye and the right image with the left eye, hence the name cross-eyed. In order to facilitate, we can put one finger at the center of the screen and then look at the top of the finger as shown in the Figure 22. We will see 3 photographs behind the finger. Then, keeping it, and focusing on the center photograph slowly, we will see it in 3D.

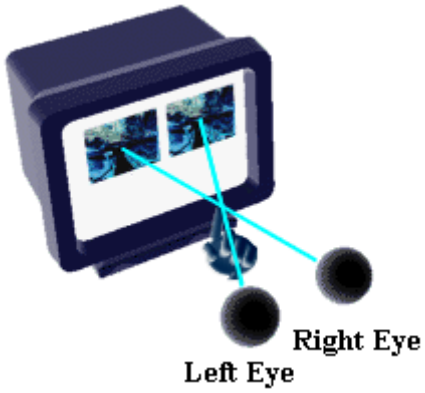
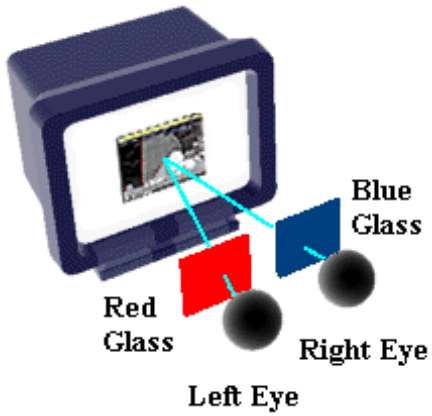


Figure 22: Cross-eye viewing [Source: Stereoeye.jp (2010)]

3. *Anaglyph Viewing:* Anaglyph images are images created from the two images (left and right) filtered with two different colors and superimposed (Figure 23(a) illustrates one such image). When viewed with two color glasses (each lens a chromatically opposite color, usually red and cyan), called color coded or anaglyph glasses as shown in the Figure 23(b), they reveal an integrated stereoscopic image. The visual cortex of the brain fuses this into perception of a three dimensional scene or composition.

(a) An anaglyph image



(b) Anaglyph vewing

Figure 23: Anaglyph image and viewing [Source: Stereoeye.jp (2010)]

4. *Projection Viewing:* In this method of viewing, the left image and the right image are projected on to a screen by two projectors with polarizing filters as shown in Figure 24. And we view the screen by wearing 3D glasses that have polarizing filters. So, the image projected from left projector goes into our left eye and the image projected from right projector goes into our right eye, and we see the 3D image. This is the most common viewing method used for watching 3D movies in the theater.

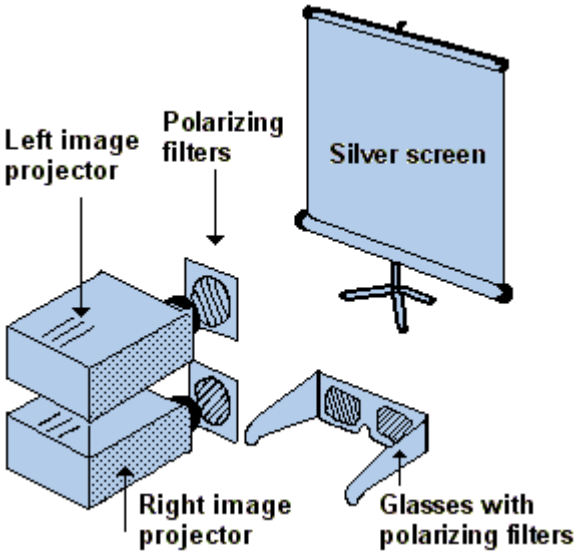
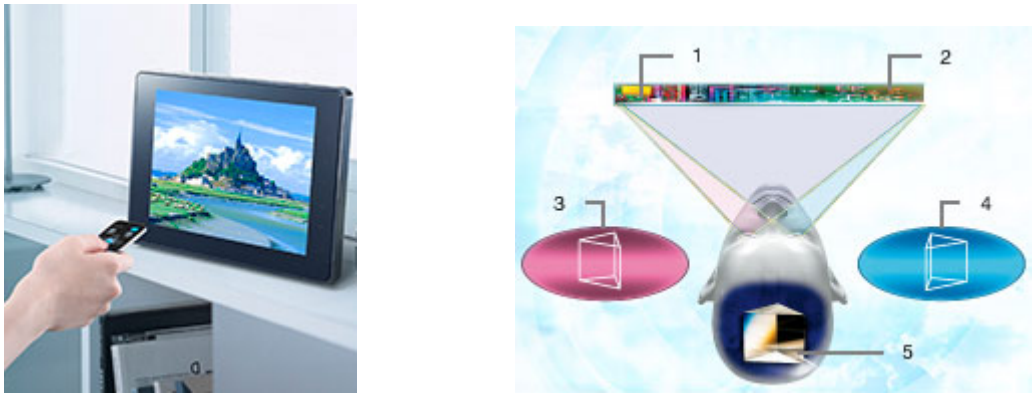


Figure 24: Projection viewing

Different types of commercial 3D viewers are available in the market. One innovative technology is the 3D digital viewing technology from Fujifilm, introduced in the first modern 3D camera, *Fujifilm FinePix REAL 3D W1* and the complimentary device, *Fujifilm FinePix REAL 3D V1*. Both of them are based on modern digital LCD display system. The viewer supports remote and touch key technolgy (Figure 25). The 3D viewing technology is based on the phenomenon known as parallax, which is the human's ability to perceive depth that come from viewing an object along two different lines of sight. The monitor on the 3D digital camera uses an originally developed "Light Direction Control System" and, the 3D digital viewer adopts a "Parallax Barrier System" to precisely direct light to the right and left eyes in a way that simulates parallax. Both enable easy and natural viewing of 3D images with just the naked eye. They allow automatic as well as manual parallax adjustment for fine tuning for a more natural 3D effect. The viewer technology is illustrated in Figure 25(b).



(a) The viewer (b) Viewer technology

Figure 25: Fujifilm 3D viewer technology [Source: Fujifilm (2010*a*)]

The numbered labels in the figure are given below:

1. Light Direction Control System/ Parallax Barrier System
2. 3D LCD monitor
3. Image received by the left eye
4. Image received by the right eye
5. 3D image

4 Proposed Multispectral-Stereo System

This chapter presents the proposed multispectral-stereo system using two digital cameras or a stereo camera. We already know that it consists of two parts: multispectral and stereo. We discuss both of these aspects in our proposed system. At first, we discuss previous related works on two camera and/or six channel based multispectral image acquisition. Chapter 2 already discussed state of the art multispectral imaging technologies. Our main focus here is in the works on multispectral image acquisition using two cameras. Model of the proposed multispectral system has also been presented. We then present the stereo imaging aspects of the system. Finally, we put forward the study methodology of the multispectral and stereo systems.

4.1 Previous Related Works

Here, we present previous works done in multispectral imaging with 6-channel system with one or two trichromatic cameras.

The architecture of a six band high-definition television (HDTV) camera system (Figure 26) for color reproduction or motion pictures based on spectral information has been proposed by Ohsawa et al. (2004). In their system, the light is divided into two optical paths by a half mirror, and incident on two conventional three-charge coupled device (CCD) cameras after transmission through the specially designed interference filters inserted in each path. The two HDTV cameras capture three band images in sync to compose each frame of the six band image. The total spectral sensitivities of the six band camera are the convoluted spectral characteristics of the optical components: the objective lens, the half mirror, the IR cutoff filter, the interference filters, the CCD sensors, etc. Six-channel uncompressed HDTV video signal is captured and recorded on a magnetic hard disk. The spectral transmittances of the interference filters used for the six band separation were specifically designed so that two different interference filters divide each of the spectral sensitivities of the R, G, and B sensors into two different parts in the spectrum region. They evaluated the system based on CIELAB color differences between the measured and reproduced color; Wiener estimation technique has been used. They obtained average and maximum CIELAB color difference for Gretag Macbeth color checker as $\Delta E^*_{ab} = 1.4$ and 4.2 respectively; where $\Delta E^*_{ab} = 4.1$ and 8.2 in the case of the conventional HDTV camera.

Figure 26: 6-band HDTV camera [Source: Ohsawa et al. (2004)]

Hashimoto and Kishimoto (2008) proposed a two-shot type 6-band still image capturing system using commercial digital camera and custom color filter. It uses a digital camera and a custom designed color filter and needs to take two shots, one without and one with the filter, thus resulting 6-channel output. The filter is custom designed in such a way that it cuts off the left side (short wavelength domain) of the peak of original spectral sensitivity of blue and red, and also cuts off the right side (long wavelength domain) of the green. It shows the 6-channels system results more accurate color and wider color range. The problem with this is that it still needs two shots and is therefore incapable of capturing objects in motion.

4.2 Proposed Multispectral System with two RGB Cameras

We have seen that state of the art multispectral systems devised with digital camera and filters acquire multispectral images in multiple shots, placing one filter at a time. This not only makes the system very slow but also unable to capture objects in motion. Moreover, the construction of such a system is very complex and costly. A solution to this problem is to use two cameras-based system as discussed in the above section. They used it with HDTV cameras along with custom designed filters. Our proposed system extends this further making it more practicable and useful.

We use two normal digital (RGB) cameras and a pair of appropriate optical filters in front of each camera. The idea is to select one or two appropriate optical filters from among a set of available filters, so that they will modify the sensitivities of one or two cameras to give six, if not quite, but reasonably well spaced channels in the visible spectrum. Selection of the filters can be done using a filter selection method presented in Section 2.3.2. Depending upon the sensitivities of the two cameras, it is also possible to select a single filter for one camera while the other camera left without any filter. The subsequent combination of the images from the two cameras provides a multispectral image of the objects in the acquired scene. It is faster, as the image can be acquired in one shot, thus making it capable of taking multispectral images of objects in motion as well. The multispectral system thus constructed will be of six channel, 3 each contributed from the two cameras. For natural and manmade surfaces whose reflectances are more or less smooth, it is recommended to use as few channels as possible (Connah et al. 2006) and we have found that six channel would be good enough in such cases.

Moreover, by joining the two cameras in a stereoscopic configuration, it allows us to capture 3D stereo images also. This makes the system capable of acquiring both the multispectral and stereo 3D images simultaneously. Figure 27 illustrates a multispectral-stereo system constructed from two digital cameras and two filters in front of two cameras. Stereo imaging needs extra procedure and precision issues to be taken into account as discussed in Chapter 3. One shot acquisition can be made possible by synchronizing the two cameras as discussed in Section 3.7. This makes the system fast as we aim for. Instead of using two cameras and join them in a stereoscopic configuration, we can use a readymade stereo camera. In both the cases, the system acts as a two-in-one multispectral stereo system. The two cameras need not be of same type, instead, any two cameras can be used, provided the two are operated in same resolution.

Figure 27: Illustration of a multispectral-stereo system with two cameras and two filters

From multispectral point of view, our proposed system looks similar to the one presented in Ohsawa et al. (2004) and discussed above in Section 4.1. However, there are significant differences in the approach and the aim in the two. The following list points out how the proposed approach differs from the previous work.

- The previously proposed system was targeted for video acquisition with two high-end television cameras. Our proposed system is targeted with two commercially available normal digital cameras, which makes it more practical.

- Beam splitter has been used in the front to divide the incident light from the objective lens into reflected and transmitted beams, which are then focused on the RGB CCD sensors of each camera head. Our approach is more simple and practical in that it uses one or two filters simply attached to the lenses of the two cameras; and no need to use beam splitter.

- Specifically designed interference filters were used in the previous work. This may need precision and specialized filter design work which introduces higher cost, and in turn might limit its usability to researchers only. Here, we will select filters from among readily available filters in the market.

- Unlike the previous work, to make it more realistic, acquisition noise has been considered in the proposed system.

- The two sets of RGB sensors in the cameras used in Ohsawa et al. (2004) were designed to have the same spectral sensitivities. Here, we are making it more flexible with the provision of using any two cameras having the same resolution.

- Ohsawa et al. (2004) evaluates the system solely based on color reproduction. However, we analyze the system based on both spectral and colorimetric reproduction.

- Moreover, we are taking advantage of the stereoscopic configuration allowing it to acquire 3D images simultaneously along with the multispectral information. We aim to have 2-in-1 *Multispectral-Stereo* system.

Thus, our proposed multispectral-system is faster, cheaper and practical solution, as it is the one shot acquisition which can be constructed from even commercial digital cameras and readily available filters; and additionally, capable of capturing 3D images at the same time.

4.3 Multispectral Camera Model

We present here the model for the proposed 6-channel multispectral system.

Let s_1 and s_2 be the spectral sensitivities of the two cameras, t_1 and t_2 are the spectral transmittances of the selected filter pair, L is the spectral power distribution of the light source and R is the spectral reflectance of the surface captured by the camera. As there is always acquisition noise introduced into the camera outputs, let n denotes the acquisition noise. The response of the camera C is then, given by the multispectral camera model (Equation 2.11) as

$$C_i = S_i^T \mathrm{Diag}(L) R + n. \tag{4.1}$$

The spectral sensitivities of the six-channel multispectral system (S) are obtained by combining sensitivity matrices of the two cameras as

$$S = [S_1, S_2]. \tag{4.2}$$

where $S_1 = \mathrm{Diag}(t_1)s_1$ and $S_2 = \mathrm{Diag}(t_2)s_2$.

4.3.1 Noise Model

Among several sources of noise as discussed in Section 2.2.2, DC noise, shot noise and quantization noises are considered as major sources. Here, we model random shot noise and quantization noise and model the noisy multispectral camera model with these noise models. DC noise is assumed to be corrected while the other noises are assumed to be negligible or corrected for.

Shot noise n_{shot} arises from the inherent uncertainty in the generation, reflection, and capture of light. This is a Poisson process (Healey and Kondepudy 1994), thus the variance of the shot noise component increases with increasing input intensity. This is modeled using multiplicative Gaussian noise as

$$n_{shot} = [\zeta_1 c_1, \zeta_2 c_2, ..., \zeta_p c_p]^T, \tag{4.3}$$

where each of the ζ_i is a pseudo-random variable taken from a Gaussian distribution with zero mean and variable standard deviation, and c_i represents the ith sensor response.

Quantization noise n_{quant} is incorporated by directly quantizing the simulated responses after the application of shot noise. This is done by rounding the sensor outputs to simulate their analogue-to-digital conversion into discrete 8-, 10- and 12-bit representations. Thus the camera outputs of our model were initially represented by any value in the range 0-1. As an example, in order to model 12-bit digital resolution, these continuous values were scaled to the range 0-4095 and then rounded to the nearest integer.

Thus, the ultimate noisy multispectral camera model is given by

$$C = S^T \mathrm{Diag}(L) R + n_{shot} + n_{quant}. \tag{4.4}$$

4.4 Methodology

The proposed integrated acquisition system can be considered to consist of two different systems: multispectral and 3D stereo. These two systems act in parallel and we study both of them separately. Different methodologies are followed while studying the two acquisition systems. We first investigate the multispectral system and then study the stereo system based on the results we obtained from these. The methodologies followed in these two types of acquisition systems are discussed briefly in the following subsections.

4.4.1 Multispectral Acquisition System

The proposed multispectral system acquires images from two cameras giving six channels, three channels from each camera. Since the two cameras are in a stereoscopic configuration, the question on how to find corresponding points in the two images can be handled with stereo matching algorithms discussed in Section 3.6. From these six channel data we can estimate the spectral reflectances of the surface whose image has been captured. The performance of a six-channel camera is evaluated in terms of its accuracy to capture surface spectral reflectances.

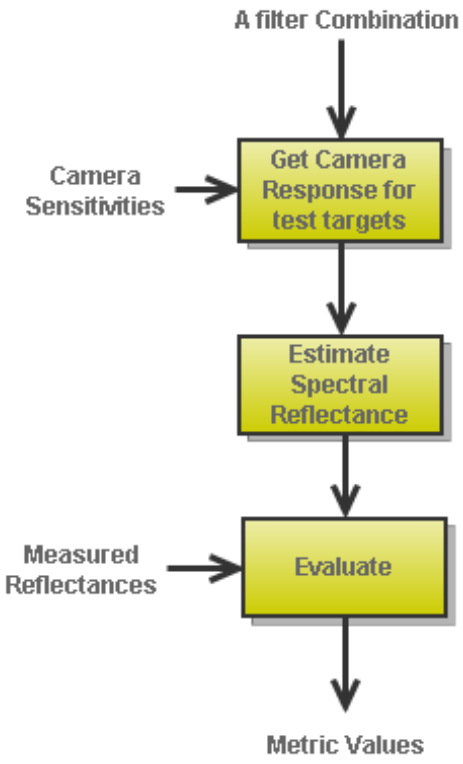


Figure 28: Process flow for study of multispectral acquisition system

Figure 28 shows the process flow, showing the study methodology of the multispectral acquisition system. For each pair of filter combination (a single filter in front of one of the cameras and the other left without any

filter is also considered a valid combination), and knowing the camera sensitivities of the two cameras, we obtain the simulated camera responses for the training and test targets. Spectral reflectances of the test target patches are then estimated and result is compared with the measured spectral reflectances using evaluation metrics. Camera sensitivities may or may not be required depending on the estimation method used. Section 2.3.4 presented different estimation methods used in this thesis. RMSE and ΔE^*_{ab} are used as the evaluation metrics and results are investigated based on optimizing these metrics (see Section 2.4 for details on these metrics). A filter pair that gives the minimum metric value is picked as the best filter combination for the cameras used.

Studies of the multispectral system has been carried out with both *simulation* as well as *experimental* approaches. Simulation based approach simulates the camera response using the camera model given by Equation 4.1 while the experimental approach uses the real camera responses captured with the multispectral-stereo camera system. Chapter 5 presents in details the simulation based approach and Chapter 6 discusses thoroughly the acquisition system based on experiments.

4.4.2 3D Stereo Acquisition System

The pair of images captured with the two cameras (left and right) corresponds to the pair of stereo images as the two cameras are setup in stereoscopic configurations. With these pair of stereo images, 3D stereo image can be viewed with any of the stereo viewing methods and/or tools as discussed in Section 3.9.

The process flow of the methodology for the study of the 3D stereo acquisition system is illustrated in Figure 29. By calibrating the stereo cameras as discussed in Sections 3.4 and 3.5, we determine the intrinsic and extrinsic parameters. Then for a given pair of stereo images captured with the system, the images are stereo rectified and then compute the disparity maps. Since one or two filters are introduced in front of one or two lenses of the two cameras, we study how it affects in the quality of the 3D images. The disparity maps obtained with the proposed system are compared with the disparity maps obtained from the pure stereo system without the filter(s). The comparisons are made based on precision and recall metrics. Precision measures the exactness, whereas the recall measures the completeness of the disparity maps computed with the proposed stereo system compared to the reference ones. Stereo image acquisition along with all these issues has been studied and analyzed in details with experimental works in Chapter 7.

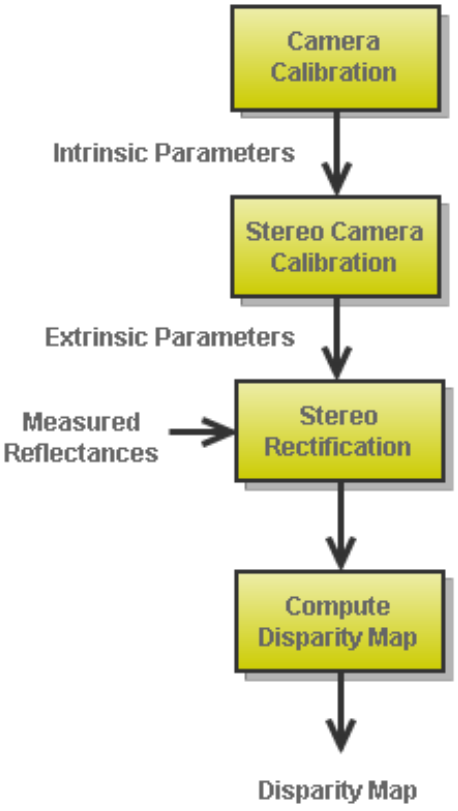


Figure 29: Process flow for study of 3D stereo acquisition system

5 Multispectral Acquisition with Mulstispectral-Stereo System: A Simulation Approach

In this chapter, we study the proposed approach of multispectral image acquisition using two digital cameras, with simulation approach. Simulation has been carried out with different pairs of cameras including the first 3D stereo camera from Fujifilm. Starting with the simulation setup used, we discuss the simulation model and process, and then present the results obtained with the simulation on systems constructed with different camera pairs. Simulation programs have been developed in Matlab version 2009b.

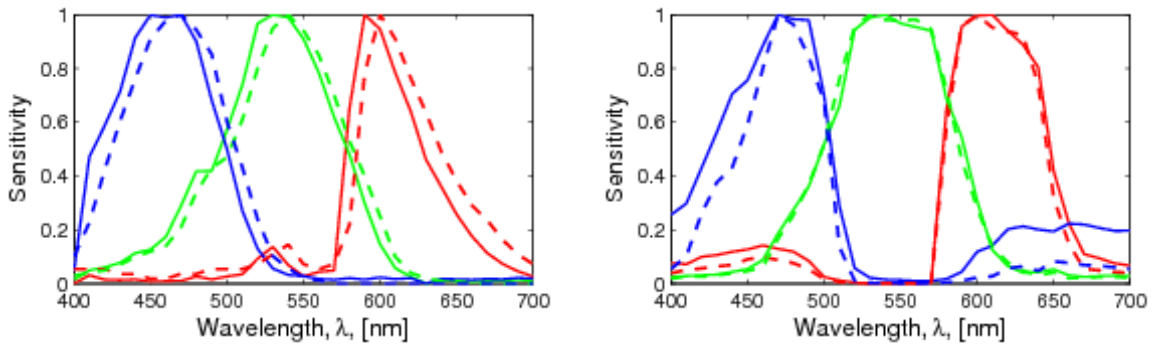
5.1 Simulation Setup

Simulation basically obtains the camera responses for the target color charts using the multispectral camera model given in Section 4.3. Section 5.2 discusses the simulation process. Simulation requires the camera sensitivities, set of filters of known transmittances, and Macbeth color checkers as targets. Here, we discuss the setup used in the simulation.

1. *Cameras :* Different cameras were used for the simulation. Simulation has been done with the sensitivities of real and imaginary cameras. Sensitivity data are obtained from available dataset and some are extracted from the plots given in papers. Here, the results with the two commercial SLR cameras: Nikon D70, Canon 20D, and with the first 3D stereo camera, the Fujifilm FinePix REAL 3D W1 (we call it simply Fujifilm3D) are presented. The sensitivities of Nikon D70, Canon 20D cameras are obtained from ISET software. The sensitivities of the Fujifilm3D are obtained from the camera characterization (see Section 6.2 for details). Figure 30 shows the pictures of the three cameras and their sensitivities are shown in Figure 31. *Note that in all the plots from here after, plots for left camera are drawn with solid lines and that for the right camera are drawn with dashed lines*.

(a) Nikon D70 (b) Canon 20D (c) Fujifilm FinePix REAL 3D W1

Figure 30: Three different cameras used in the simulation



(a) Nikon D70 (solid) and Canon 20D (dashed) (b) Fujifilm 3D (Left - solid, Right - dashed)

Figure 31: Camera sensitivities

2. *Filters:* Two hundred and sixty five optical filters of three different types: exciter, dichroic and emitter from Omega (Omega 2010) are used in the simulation. Rather than mixing filters from different vendors, one vendor has been chosen as a one stop solution for the filter, and Omega has been picked as it has a large selection of filters and data is available online (Omega 2010). The list of all the filters used has been given in Appendix B.

   Moreover, we have got a multiband Schott BG36 filter for the experimental work and so it has also been used with Fujifilm3D in simulation. Figure 32(a) shows some example filters and 32(b) shows the Schott BG-36 filter.

(a) Example filters [Source: `http://www.midwestopticalsystems.com/filters.html`]

(b) Schott BG-36 filter

Figure 32: Optical filters

3. *Test Target:* The Gretag Macbeth color checker DC-MCC240 (Figure 33) has been used as the training target, and the classic Macbeth color checker-MCC24 (Figure 34) as the test target in the simulation. The reflectances might have changed due to few years of time of purchase of the charts available in the laboratory. So for better accuracy, reflectances of the color patches of these color charts have been measured. Section 6.4 describes the measurement procedure. The measured reflectances of the charts are shown in Figures 35 and 36. The outer surrounding gray patches, the glossy patches in the S-column of the DC chart, and inner white patches except the central four have been omitted from the training dataset. The average of the four central white patches is used as one white patch.

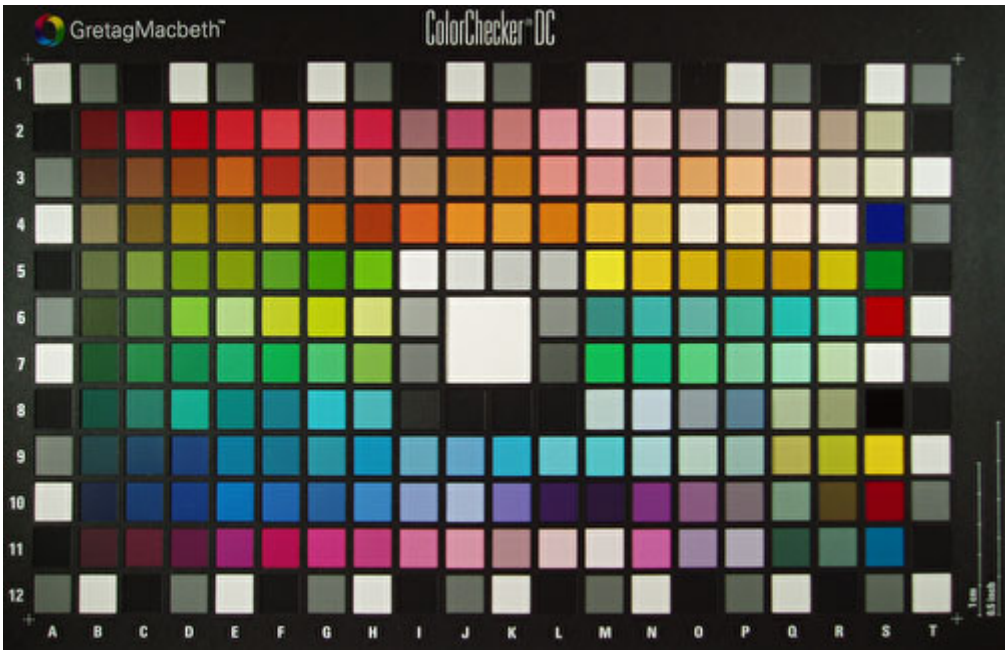


Figure 33: Macbeth color checker DC

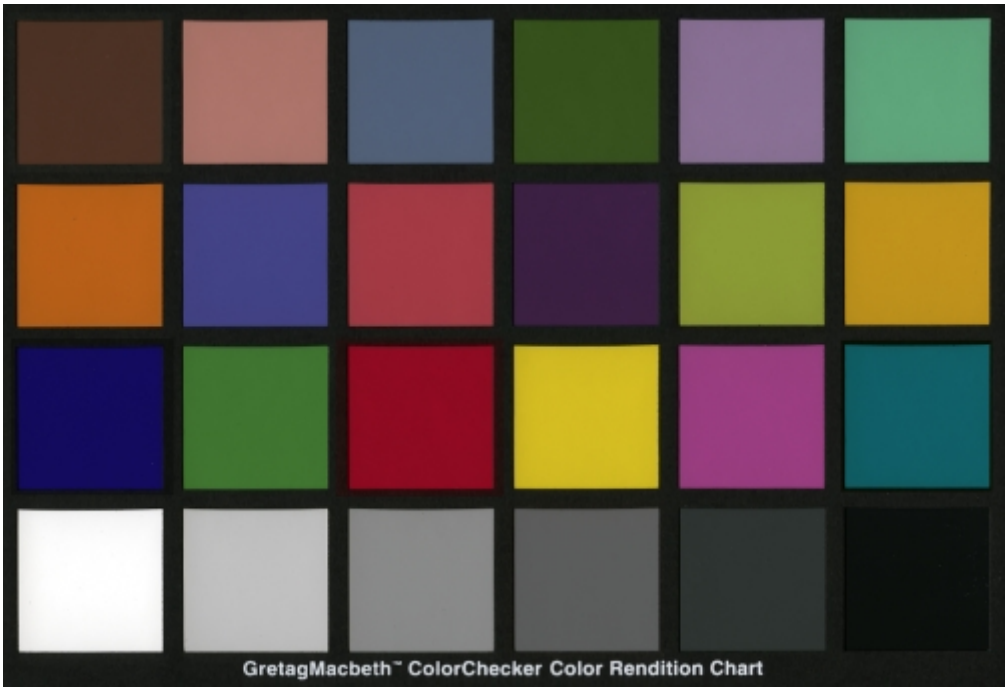


Figure 34: Macbeth color checker classic

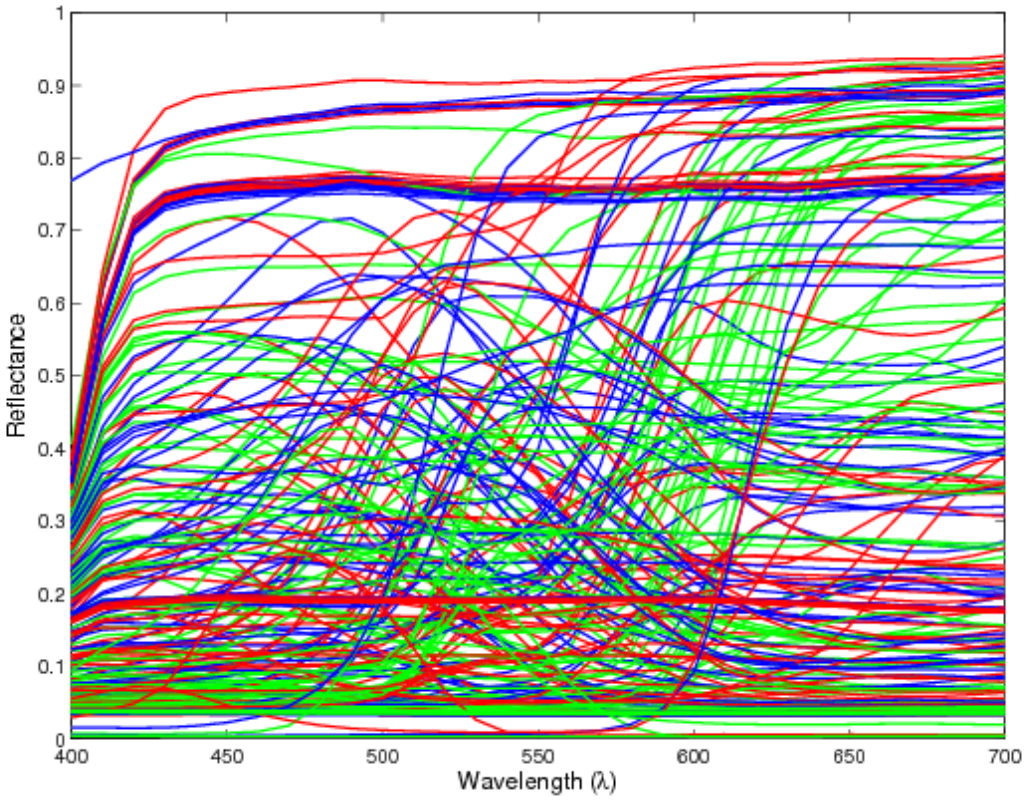


Figure 35: Reflectances of Macbeth color checker DC patches

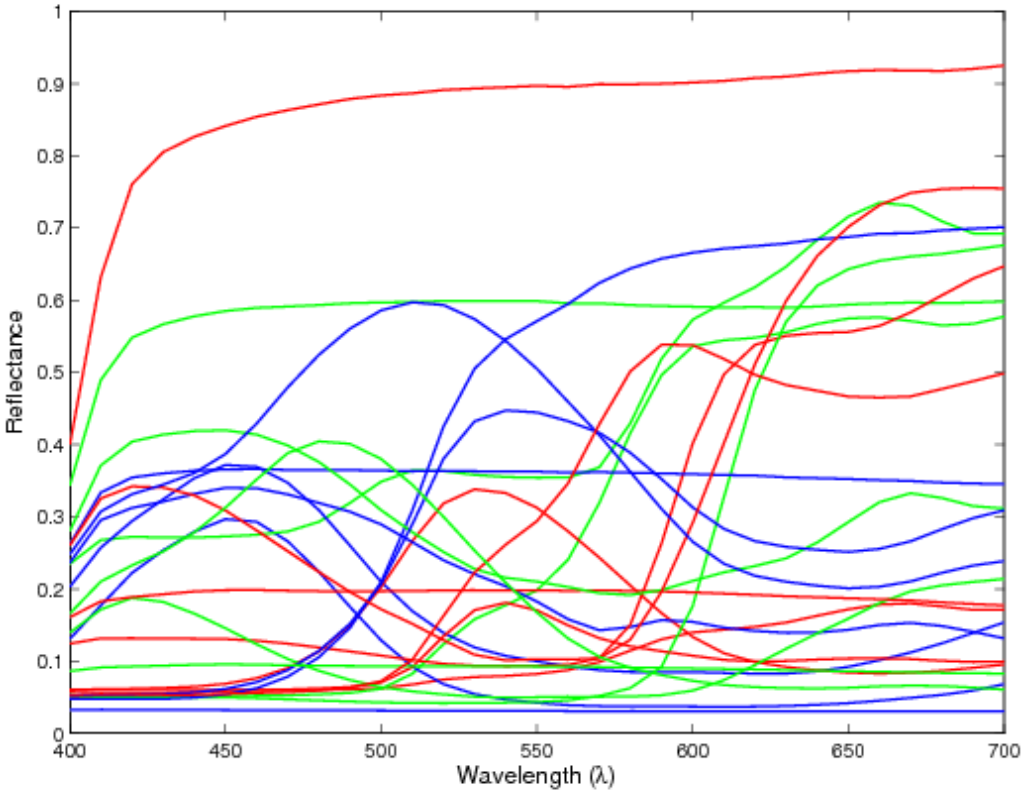


Figure 36: Reflectances of Macbeth color checker classic patches

5.2 Simulation Process

Knowing the camera spectral sensitivities (s_1 and s_2), selected filters' spectral transmittances (t_1 and t_2), and the spectral power distribution of the light source (L), the simulated camera output of the resulting six-channel multispectral acquisition system (C) is obtained using the multispectral camera model as presented in Section 4.3. Simulated random shot noise and quantization noise are introduced as described in the Subsection 5.2.2.

The reconstructed/estimated reflectance ($\tilde{R}$) is obtained for the test target from the training target reflectance (R_{train}). Different estimation methods as described in Section 2.3.4 have been investigated. All the inverse calculations are regularized with Tikhonov regularization technique (see Section 2.3.5 for detals), and the regularization parameter (α) is obtained through optimization such that the estimation error is minimized.

The estimated reflectances are evaluated using the two evaluation metrics: RMSE and ΔE^*_{ab} (see Section 2.4 for details). CIE D50 illuminant and CIE XYZ 1964 color matching functions are used for color computation.

The best pair of filters is exhaustively searched, according to each of the evaluation metrics, from among all available filters with which the multispectral system can reconstruct the measured reflectances of the twenty four Macbeth color checker patches. The results corresponding to minimum mean of the three evaluation metrics are obtained. To speed up the process, the combinations not fulfilling the constraints described in Section 5.2.1.1 are skipped. The 265 filters leads to more than 70,000 possible combinations. The constraints introduced reduce the processing down to less than 20,000 combinations.

Simulation program has been developed in Matlab 2009b.

5.2.1 Selection of Filters

Section 2.3.2 presented different algorithms for the selection of filters. Here, we have to choose at most two filters from the set of available filters. Therefore, the exhaustive search method is feasible and a logical choice because of its guaranteed optimal results. Since the method takes quite long time when the number of available filters becomes large, we introduce additional constraints to reduce the number of evaluations. The Section 5.2.1.1 presents the selection constraints introduced in our simulation.

5.2.1.1 Additional Constraints

We can ignore filter combinations which result small transmittances in one or more channels. Infeasible filter pairs are excluded based on criteria: "Filter pairs that result maximum transmission factor of less than 40%, and less than 10% of the maximum transmission factor in one or more channels are excluded".

5.2.2 Noise Simulation

Random shot noise n_{shot} is simulated as certain percentage of the camera output multiplied by normally distributed random numbers whose value is in the range [0,1]. Recent measurements of noise levels in a trichromatic camera suggest that realistic levels of shot noise are between 1% and 2% (Barnard and Funt 2002). So 2% normally distributed random noise is introduced as a random shot noise in the simulation. Given that a pseudo-random variable is used to simulate noise, the results of these simulations will vary from trial to trial. In order to discount this effect, the simulations are repeated 10 times and the average is recorded. Moreover, for the stability of the results in multiple runs of the simulation, the noise is saved and the same noise is used in different runs. In order to check the stability of the system with different levels of noise, the system has been tested with different levels of noise in 1% and 10% range. Section 5.4 presents and discusses this in details.

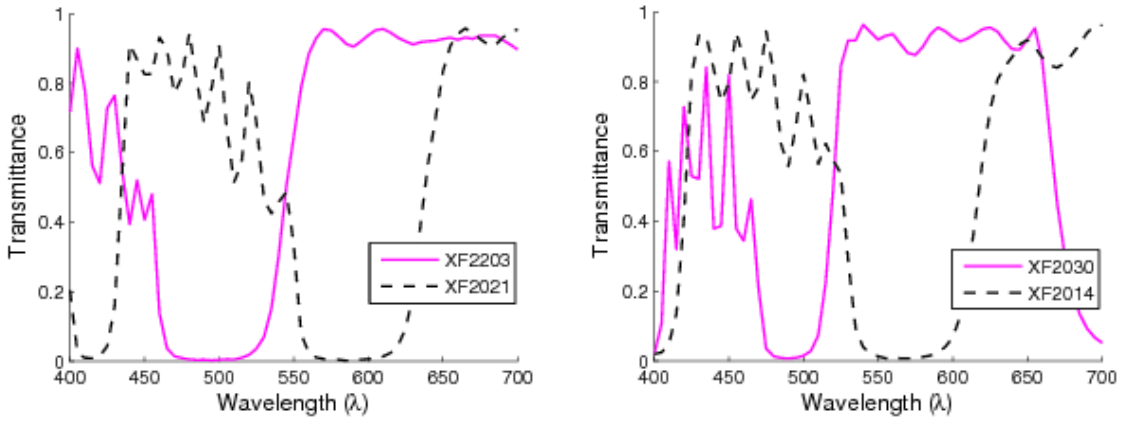
Quantization noise n_{quant} is incorporated by directly quantizing the simulated responses after the application of shot noise. 12-bit quantization noise is used. Other sources of noise, such as dark noise, are assumed to be negligible or corrected for.

5.3 Simulation Results

Simulation has been carried out on several pairs of real and imaginary cameras. The results of two representative cases: NikonD70–NikonD70 pair (Same camera pair), NikonD70–Canon20D pair (different camera pair) along with the Fujifilm3D camera used in the experiment have been presented here. As the experiment has been carried out with Fujifilm3D camera and a Schott BG-36 filter (Chapter 6), simulation has also been carried out with similar setup, with one lens covered by the filter, while the other lens left open.

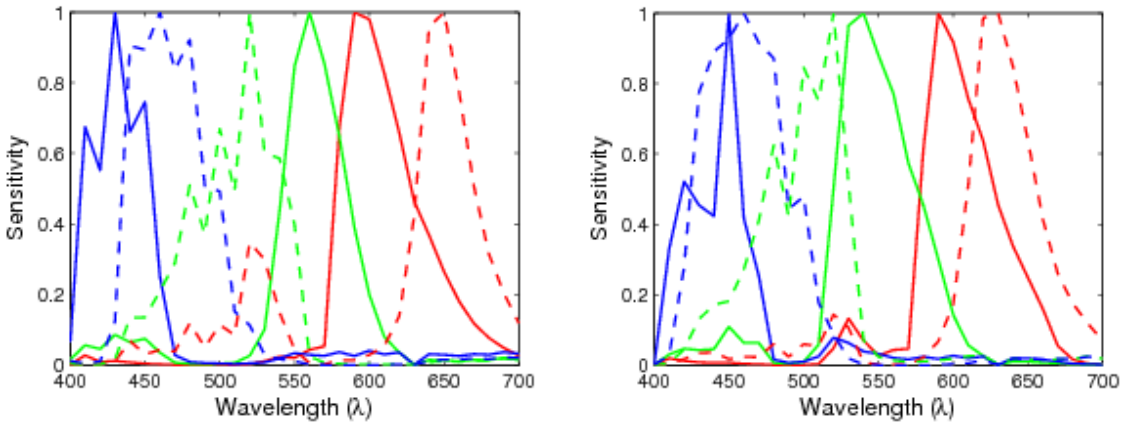
5.3.1 With two Nikon D70 Cameras

The simulation picks different filter pairs for minimum mean RMSE and for the minimum mean ΔE^*_{ab} depending on the estimation method. Three estimation methods: Maloney and Wandell, Imai and Berns, and Least Square-Wiener are investigated in this simulation. The polynomial and neural network methods are very slow for the simulation of a large number of combinations. So for comparative studies, these methods are investigated in the simulation with the Fujifilm3D camera below. For minimum mean RMSE, all the three methods select the same filter pair (XF2203–XF2021) as the best combination. Similarly, the same pair of filters (XF2030–XF2014) has been picked by all the three methods for minimum mean ΔE^*_{ab}. Figures 37(a) and 37(b) show the transmittances of the filters selected in these two cases respectively. The normalized effective channel sensitivities of the multispectral system corresponding to the optimal results of the two metrics are shown in Figure 38. In these plots, thin curves correspond to the modified sensitivities of the left camera and the thick curves correspond to that of the right camera.



(a) For Minimum RMSE (b) For Minimum ΔE^*_{ab}

Figure 37: Transmittances of the filters selected (NikonD70–NikonD70)



(a) For Minimum RMSE (b) For Minimum ΔE^*_{ab}

Figure 38: Multispectral 6-channel normalized sensitivities (NikonD70–NikonD70)

Tables 1 and 2 show statistics of estimation errors of the different methods for 3-channel and 6-channel systems respectively. Figure 42 shows the statistics of the RMSE and ΔE^*_{ab} error values respectively.

Table 1: Estimation errors with 3-channel system (Nikon D70)

Method	**RMSE %**			ΔE^*_{ab}		
	Max	**Mean**	**Std**	**Max**	**Mean**	**Std**
Maloney & Wandell	10.60	3.49	2.21	6.82	2.04	1.60
Imai & Berns	11.37	3.42	2.33	10.23	2.75	2.45
Wiener	11.83	3.33	2.44	6.26	1.98	1.53

Table 2: Estimation errors with 6-channel system (NikonD70–NikonD70)

Method	**For minimum RMSE**						**For minimum ΔE^*_{ab}**					
	RMSE %			ΔE^*_{ab}			**RMSE %**			ΔE^*_{ab}		
	Max	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**
Maloney & Wandell	2.51	1.02	0.56	1.04	0.43	0.28	2.57	1.11	0.57	1.23	0.37	0.28
Imai & Berns	2.54	1.02	0.55	1.04	0.43	0.28	2.59	1.10	0.55	1.22	0.37	0.29
Wiener	2.55	1.02	0.55	1.05	0.42	0.27	2.55	1.08	0.53	1.22	0.37	0.28

We now see the results of the reflectance estimation. Since the results are similar with all the three methods, the results obtained with the Wiener method has been given as illustration. Figures 39 and 40 show the estimated spectral reflectances along with the measured reflectances of the 24 color patches of the classic Macbeth color checker, for the two cases. The corresponding estimation result obtained with a 3-channel camera is shown in Figure 41. The filters selected by the simulation for the two cameras (left-right) and the two metric values are also shown in these plots.

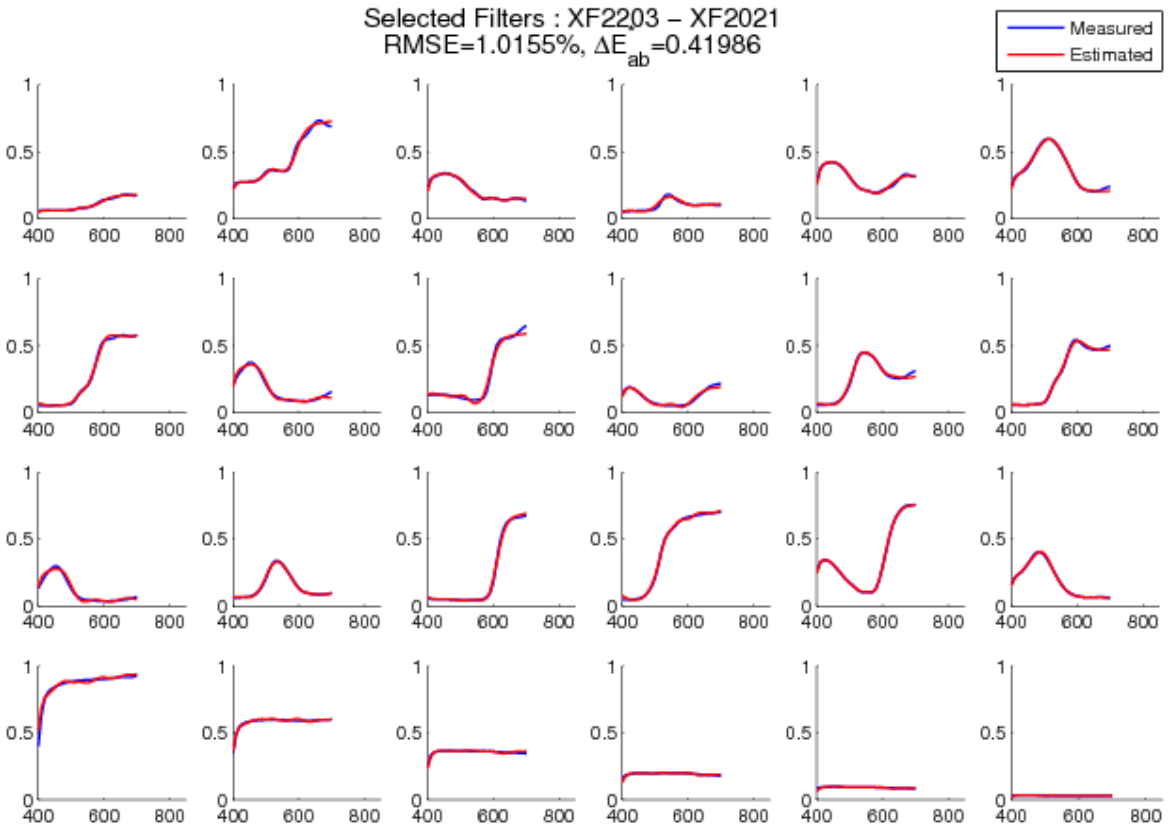


Figure 39: Reflectances with 6-channel for minimum RMSE (NikonD70–NikonD70)

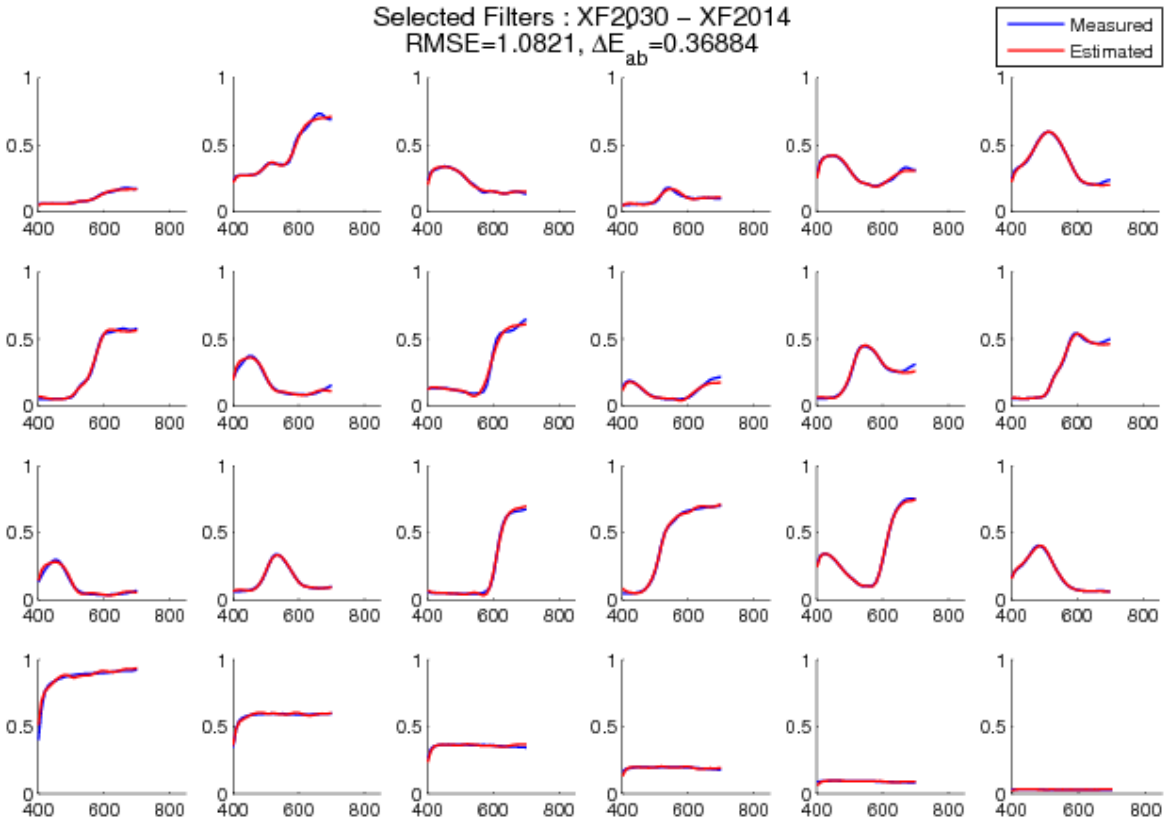


Figure 40: Reflectances with 6-channel for minimum ΔE^*_{ab} (NikonD70–NikonD70)

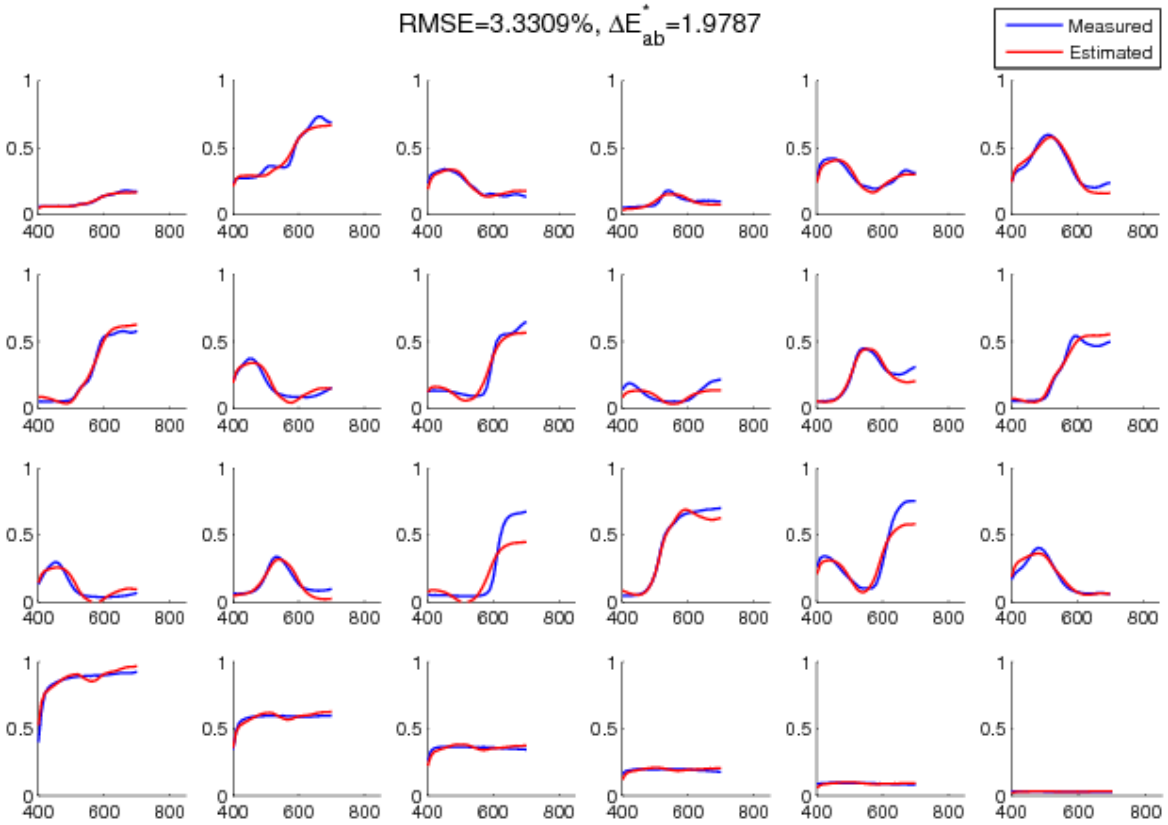
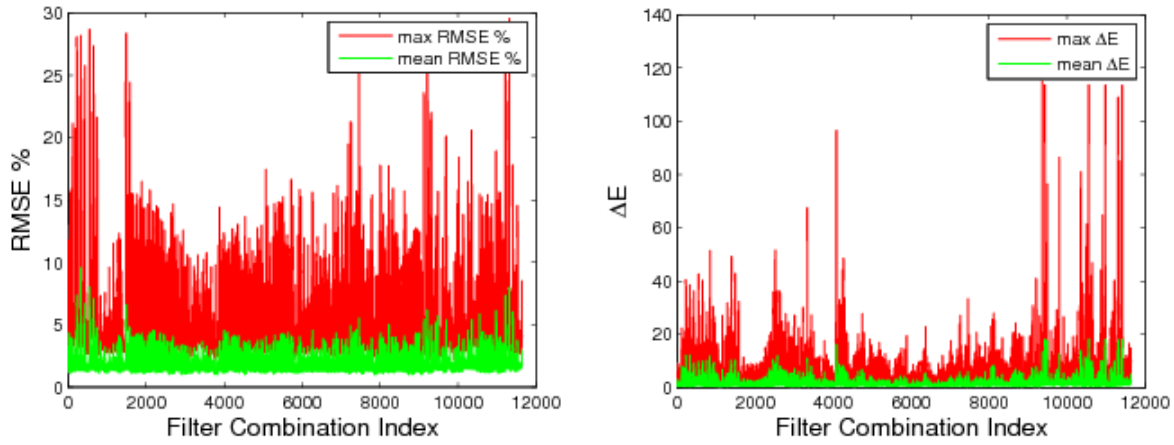


Figure 41: Reflectances with 3-channel (Nikon D70)



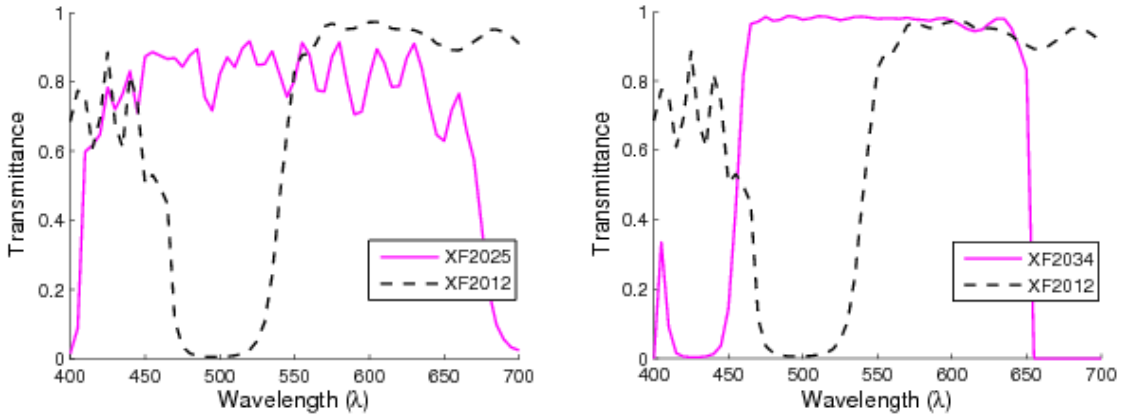
(a) RMSE

(b) ΔE^*_{ab}

Figure 42: Error statistics (NikonD70–NikonD70)

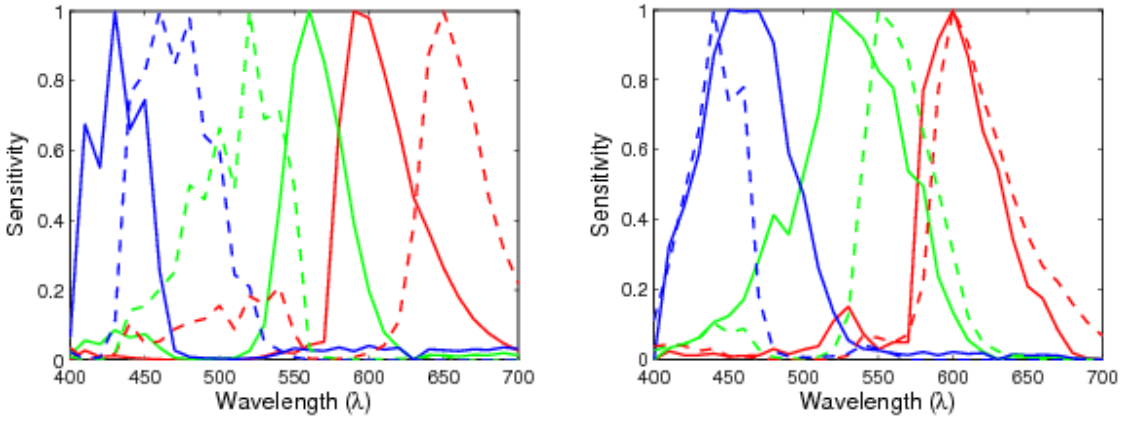
5.3.2 With Nikon D70 and Canon 20D Cameras

In this case, all the three estimation methods select the same filter pair (XF2203–XF2021), just like with two Nikon D70 cameras, for the minimum mean RMSE. For the minimum ΔE^*_{ab}, Maloney and Wandell, and Wiener methods pick the filter pair (XF2025–XF2012), while Imai and Bern method chooses the filter pair (XF2034–XF2012). The filter pair selected for the minimum RMSE is already given above with the two Nikon D70 case, in Figure 37(a). The two pairs of filters chosen by three different methods for the minimum ΔE^*_{ab} are shown in Figure 43. Normalized 6-channel sensitivities corresponding to the optimal results of the two metrics using Wiener method are shown in Figure 44.



(a) (b)

Figure 43: Transmittance of the filter pairs selected for minimum ΔE^*_{ab} (NikonD70–Canon20D)



(a) For Minimum RMSE (b) For Minimum ΔE^*_{ab}

Figure 44: Multispectral 6-channel normalized sensitivities (NikonD70–Canon20D)

Tables 3 and 4 show the statistics of estimation errors of the different methods for 3-channel and 6-channel systems respectively. Figures 45 and 46 show the result of the estimations obtained with the Wiener method for the minimum RMSE and the minimum ΔE^*_{ab} respectively. The corresponding estimation result obtained with a 3-channel camera is shown in Figure 47.

Table 3: Estimation errors with 3-channel system (Canon20D)

Method	**RMSE %**			ΔE^*_{ab}		
	Max	**Mean**	**Std**	**Max**	**Mean**	**Std**
Maloney & Wandell	10.60	3.49	2.21	6.82	2.04	1.60
Imai & Berns	11.37	3.42	2.33	10.23	2.75	2.45
Wiener	11.83	3.33	2.44	6.26	1.98	1.53

Table 4: Estimation errors with 6-channel system (NikonD70–Canon20D)

Method	**For minimum RMSE**						**For minimum ΔE^*_{ab}**					
	RMSE %			ΔE^*_{ab}			**RMSE %**			ΔE^*_{ab}		
	Max	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**
Maloney & Wandell	2.50	1.00	0.54	1.58	0.54	0.38	2.89	1.31	0.69	1.06	0.36	0.26
Imai & Berns	2.54	1.00	0.55	1.62	0.54	0.38	2.77	1.19	0.60	1.03	0.37	0.25
Wiener	2.57	1.01	0.54	1.57	0.53	0.37	2.94	1.28	0.68	1.06	0.35	0.26

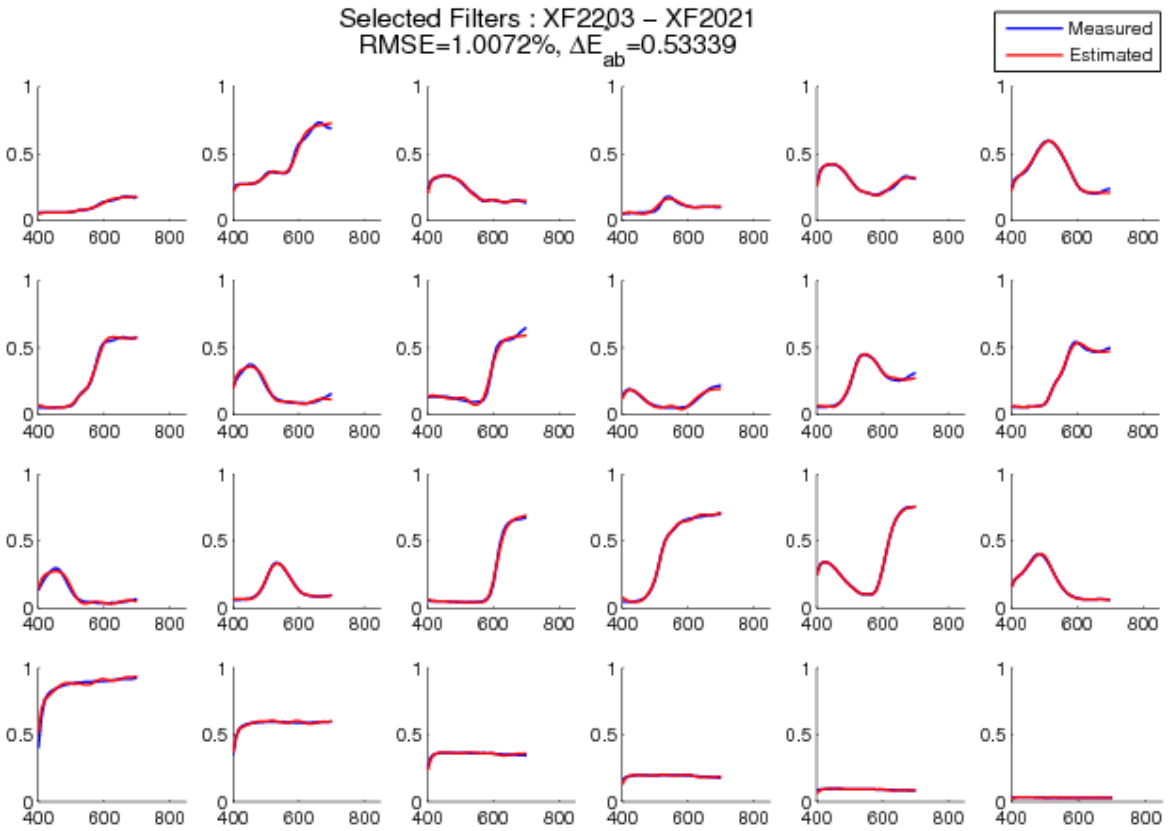


Figure 45: Reflectances with 6-channel for minimum RMSE (NikonD70–Canon20D)

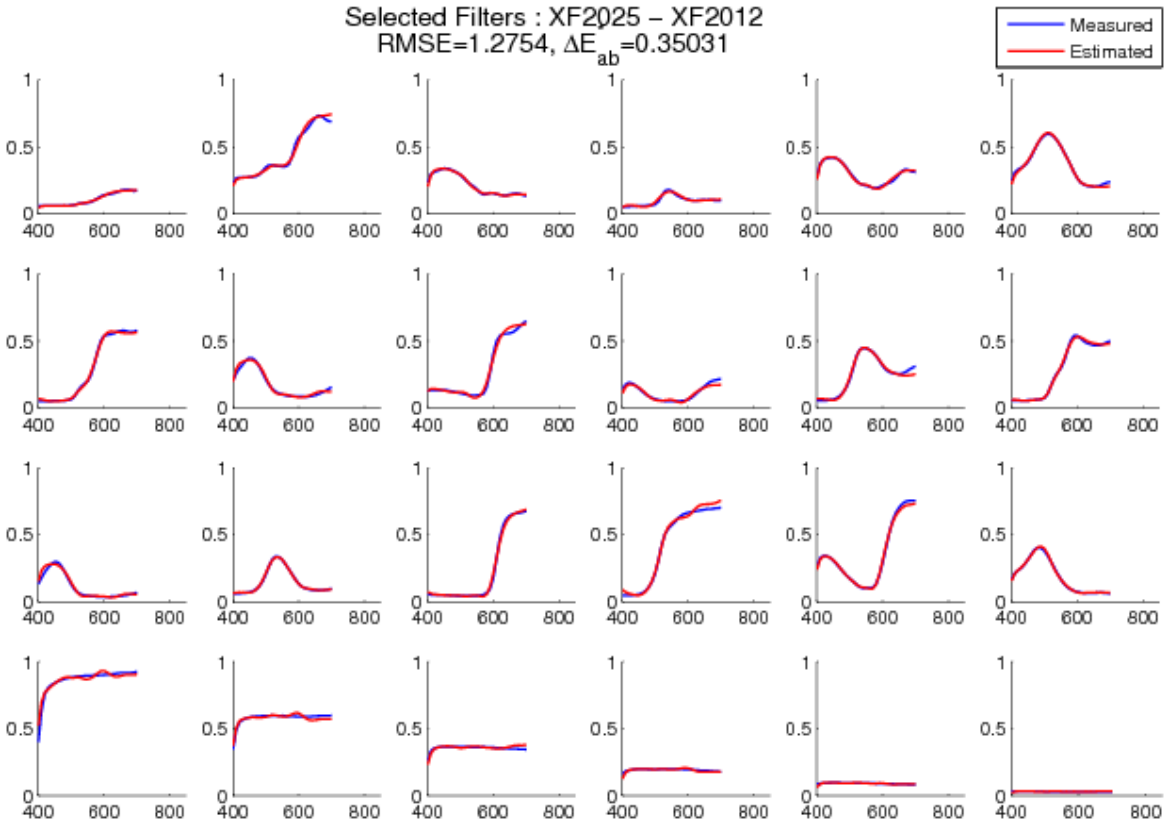


Figure 46: Reflectances with 6-channel for minimum ΔE^*_{ab} (NikonD70–Canon20D)

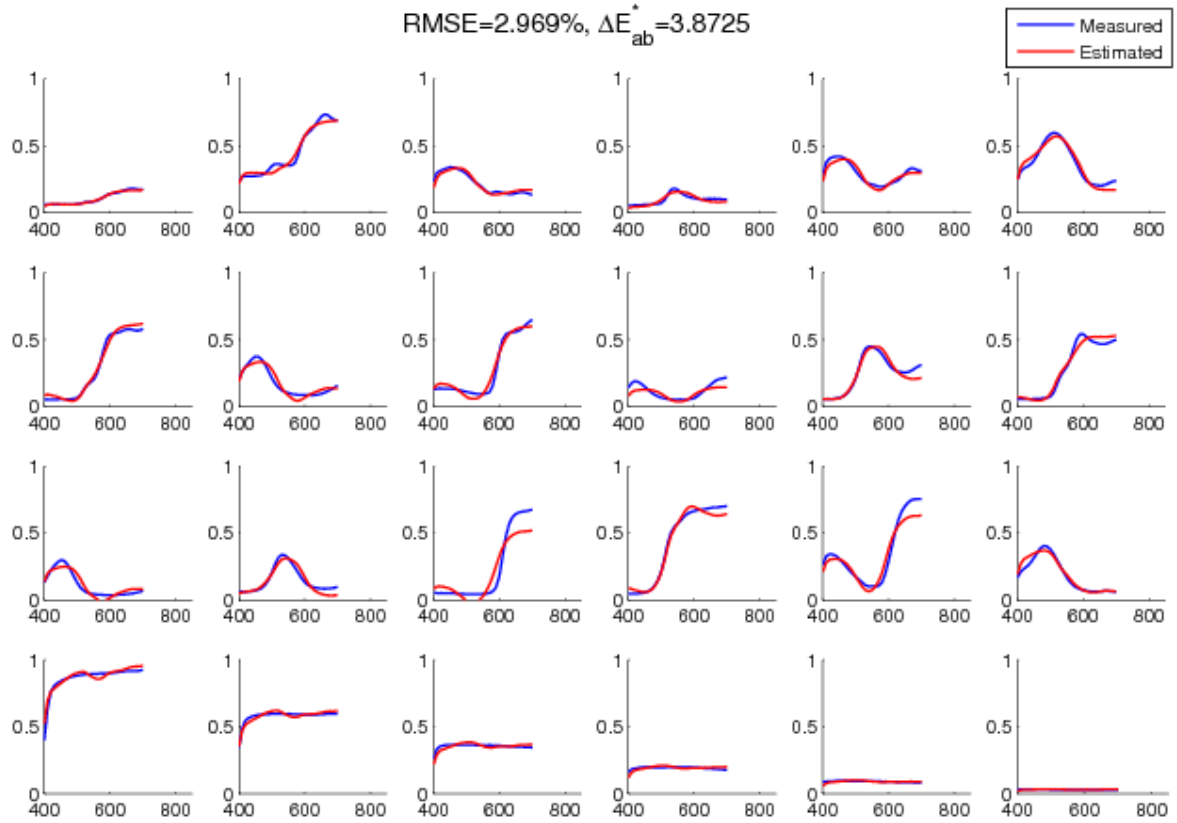
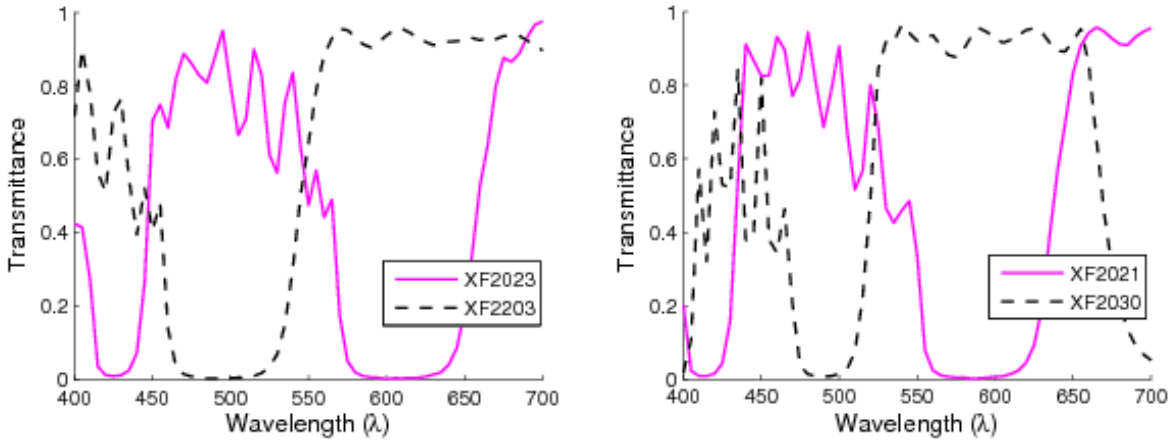


Figure 47: Reflectances with 3-channel (Canon20D)

5.3.3 With Fujifilm 3D Stereo Camera

Simulation picks the filter pair (XF2023–XF2203) by all the three estimation methods for the minimum mean RMSE. Similarly, for the minimum mean ΔE^*_{ab}, the filter pair (XF2021–XF2030) has been selected by all the three methods. These selected filters in the two cases are given in Figure 48. The corresponding normalized effective six channel sensitivities of the resulting multispectral system are given in Figure 49).

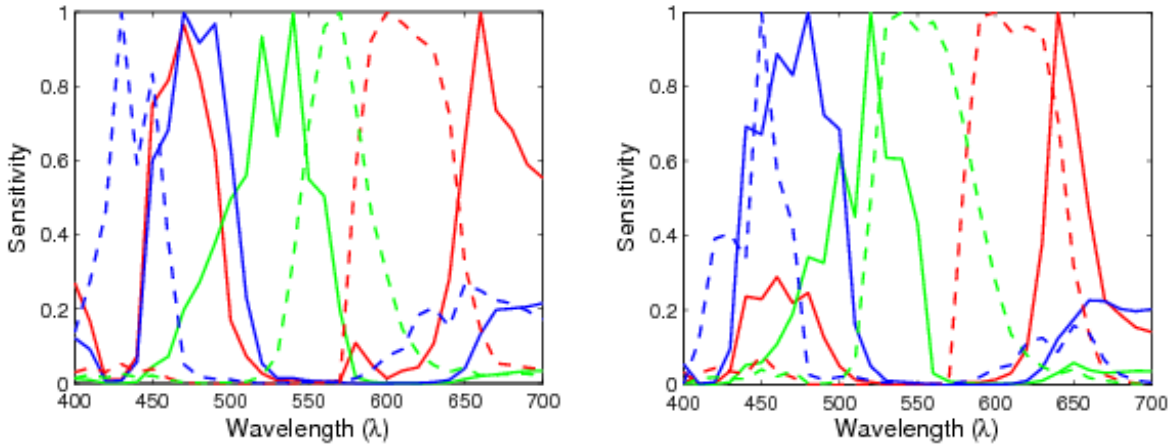


(a) For Minimum RMSE

(b) For Minimum ΔE^*_{ab}

Figure 48: Transmittances of the filters selected (Fujifilm3D)

Tables 5 and 6 show the statistics of estimation errors of the different methods for 3-channel and 6-channel systems respectively. The results of the estimation with Wiener method with the two metrics are given in Figures 50 and 51. The corresponding estimation result obtained with a 3-channel camera is shown in the Figure 52.



(a) For Minimum RMSE (b) For Minimum ΔE^*_{ab}

Figure 49: Multispectral 6-channel normalized sensitivities (Fujifilm3D)

Table 5: Estimation errors with 3-channel system (Fujifilm3D-Left)

Method	RMSE %			ΔE^*_{ab}		
	Max	Mean	Std	Max	Mean	Std
Maloney & Wandell	8.73	3.10	1.78	19.61	4.59	4.27
Imai & Berns	9.11	3.11	1.84	19.26	4.65	4.29
Wiener	9.56	3.11	1.91	20.70	4.62	4.38

Table 6: Estimation errors with 6-channel system (Fujifilm3D)

Method	For minimum RMSE						For minimum ΔE^*_{ab}					
	RMSE %			ΔE^*_{ab}			RMSE %			ΔE^*_{ab}		
	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std
Maloney & Wandell	2.55	1.03	0.52	2.29	0.72	0.59	2.70	1.13	0.59	0.95	0.46	0.24
Imai & Berns	2.58	1.04	0.53	2.30	0.73	0.61	2.75	1.12	0.58	1.28	0.48	0.32
Wiener	2.68	1.04	0.54	2.12	0.73	0.57	2.86	1.13	0.59	1.01	0.46	0.25

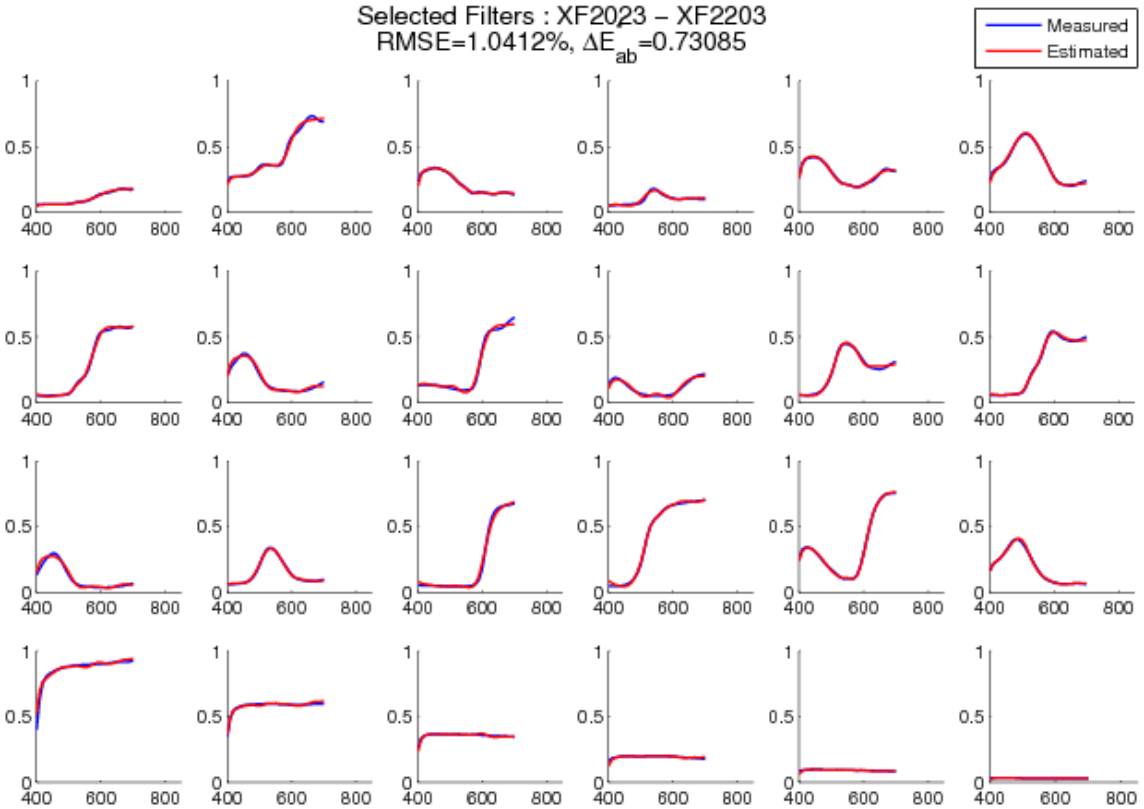


Figure 50: Reflectances with 6-channel for minimum RMSE (Fujifilm3D)

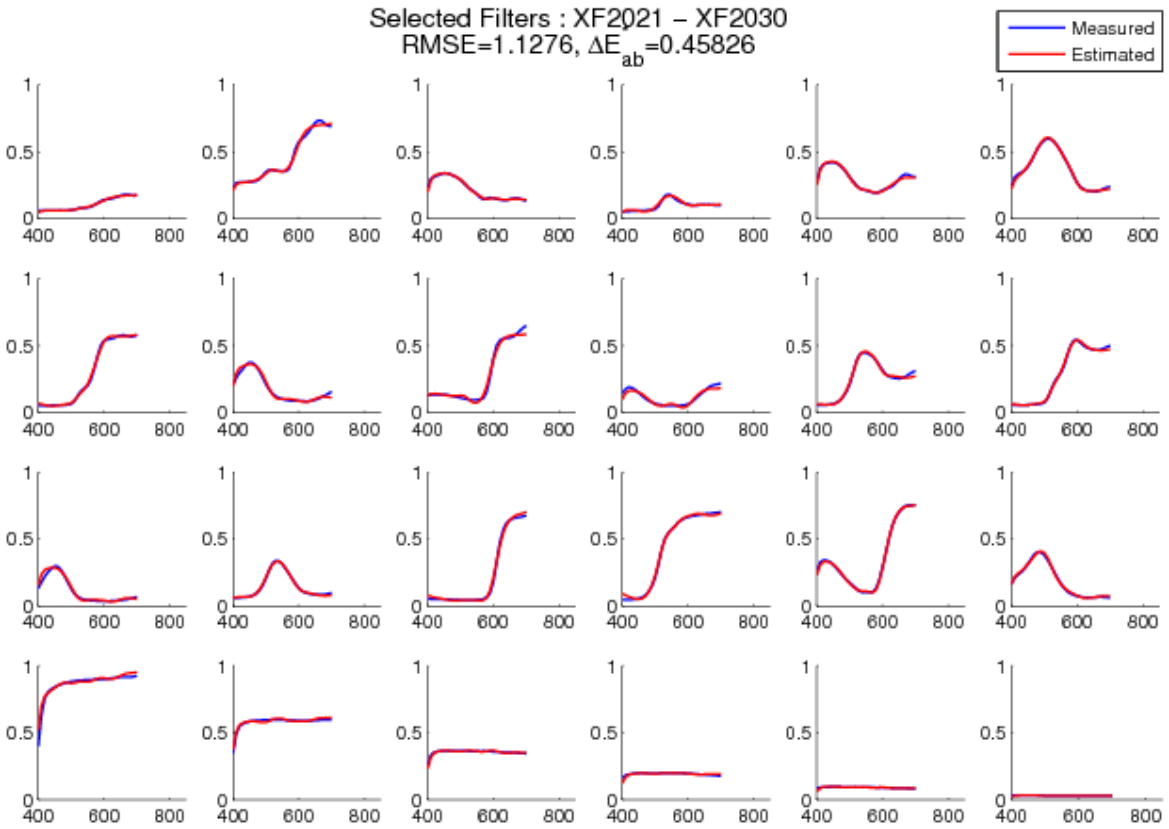


Figure 51: Reflectances with 6-channel for minimum ΔE^*_{ab} (Fujifilm3D)

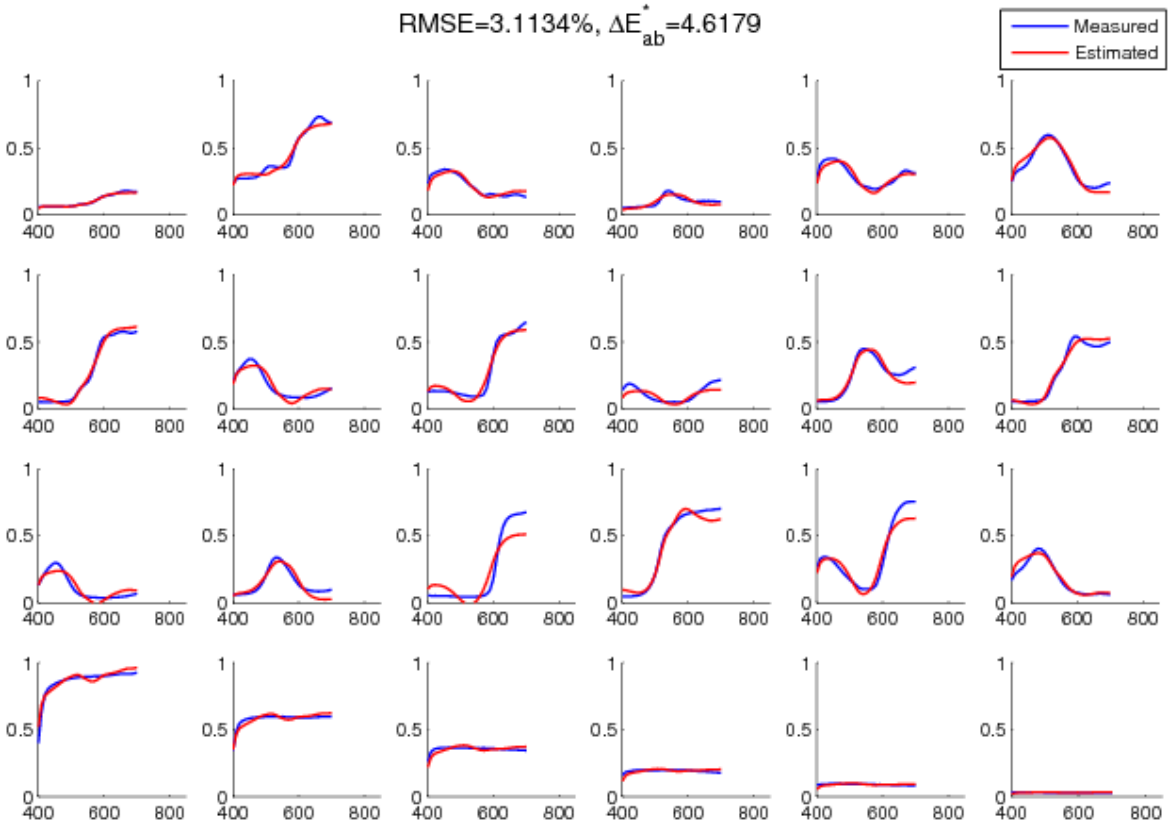


Figure 52: Reflectances with 3-channel (Fujifilm3D-Left)

5.3.4 With Fujifilm 3D Stereo Camera and a Multiband Schott BG-36 Filter

The experimental study (Chapter 6) has been carried out with a multiband Schott BG-36 filter in front of one lens of the camera and the other lens without any filter. So, in order to have comparative analysis of simulation and experimental results, simulation has been done with this setup as well.

The transmittance of the Schott BG-36 filter (obtained from the catalogue provided by the manufacturer of the filter) and the normalized effective channel sensitivities of the resulting six channel multispectral system are given in the Figures 53 and 54 respectively. Figure 55 gives the result of the estimated reflectances of the MCC24 along with their measured reflectances. Figure 52 shows the result of the estimation with a single camera (since both left and right cameras are similar, left one has been selected for illustration). The statistics of estimation errors of the different methods for 3-channel and 6-channel systems are given in Table 7. The results of polynomial and neural network methods are also given. The degree of polynomial has been determined by optimizing the estimation error and the optimization picked the second order polynomial. Similarly, the learning rate η of the neural network method has also been determined through error optimization and number of iterations has been set to a larg number 50000.

Table 7: Estimation errors for 3 and 6-channel system (Fujifilm3D+SchottBG36)

Method	**3-Channel System**						**6-Channel System**					
	RMSE %			ΔE^*_{ab}			**RMSE %**			ΔE^*_{ab}		
	Max	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**
Maloney & Wandell	8.73	3.10	1.78	19.61	4.59	4.27	3.38	1.60	0.77	6.96	1.79	1.55
Imai & Berns	9.11	3.11	1.84	19.26	4.65	4.29	2.92	1.57	0.63	4.84	1.57	1.04
Least Squares	9.56	3.11	1.91	20.70	4.62	4.38	2.78	1.61	0.65	4.38	1.51	1.03
Neural Network	9.56	3.11	1.91	20.63	4.58	4.37	2.96	1.58	0.73	4.97	1.67	1.10
Polynomial (degree 2)	8.72	3.16	1.67	18.70	5.62	4.37	2.48	1.54	0.57	3.87	1.57	0.95

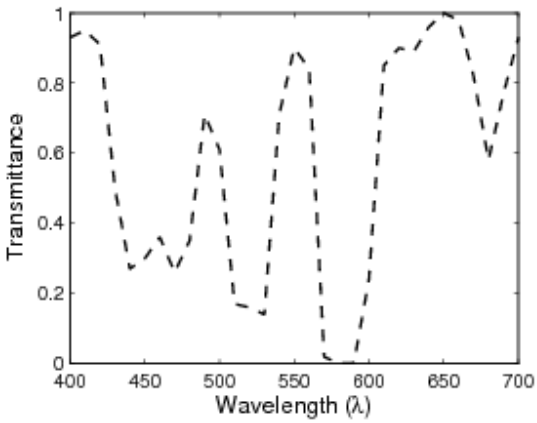


Figure 53: Transmittance of Schott BG-36 filter

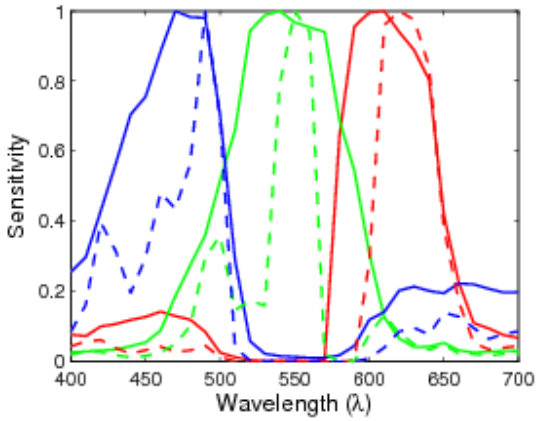


Figure 54: Normalized 6-channel normalized sensitivities (Fujifilm3D+SchottBG36)

5.4 Influence of Noise Level

It is well understood that acquisition noise plays a vital role in the performance of a digital imaging system. In order to see how the noise level influences the performance of the 3-channel RGB camera and 6-channel multi-spectral systems, we haved investigated the changes in the estimation errors through simulation with Fujifilm 3D and Schott BG-36 filter as above with different levels of noise. Table 8 shows the statistics of the results of the estimation errors for both 3-channel and 6-channel systems and Figure 56 shows the trend of error changes (RMSE and ΔE^*_{ab}) with the level of noise. The results are obtained with least square-Wiener estimation method. The trend is similar with other methods too.

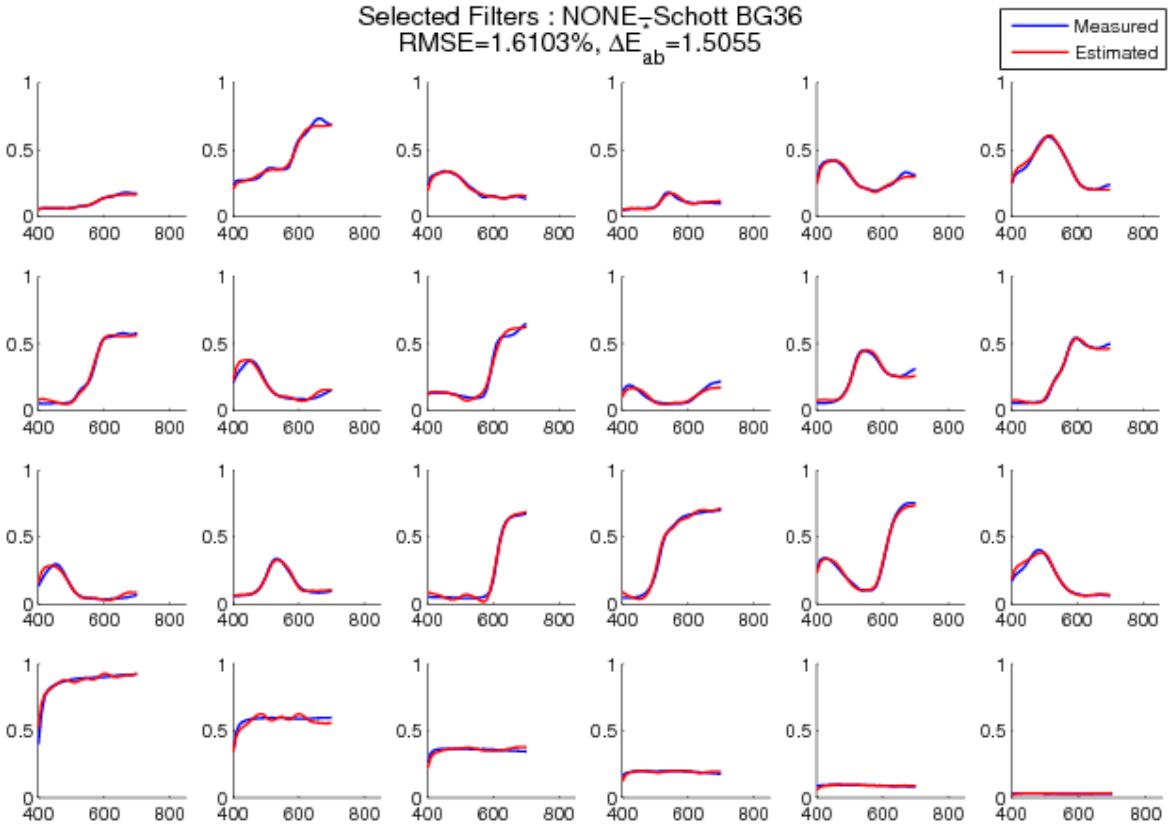


Figure 55: Reflectances with 6-channel multispectral system (Fujifilm3D+SchottBG36)

Table 8: Changes in estimation errors with the level of noise

Noise Level %	3-Channel System						6-Channel System					
	RMSE %			ΔE^*_{ab}			RMSE %			ΔE^*_{ab}		
	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std
0	10.58	3.77	2.33	20.58	4.75	4.56	3.27	1.93	0.86	3.49	1.34	0.94
2	10.53	3.77	2.33	20.17	4.78	4.51	3.74	1.98	0.95	5.19	1.44	1.10
4	10.66	3.80	2.32	21.02	4.77	4.52	3.32	2.12	0.86	4.84	1.74	1.11
6	10.52	3.80	2.33	20.22	4.89	4.47	4.79	2.24	1.16	7.68	2.28	1.77
8	10.56	3.77	2.30	19.93	4.91	4.42	4.59	2.47	1.29	6.85	2.25	2.04
10	10.46	3.88	2.34	20.12	4.77	4.44	7.14	2.61	1.57	11.75	2.99	2.68

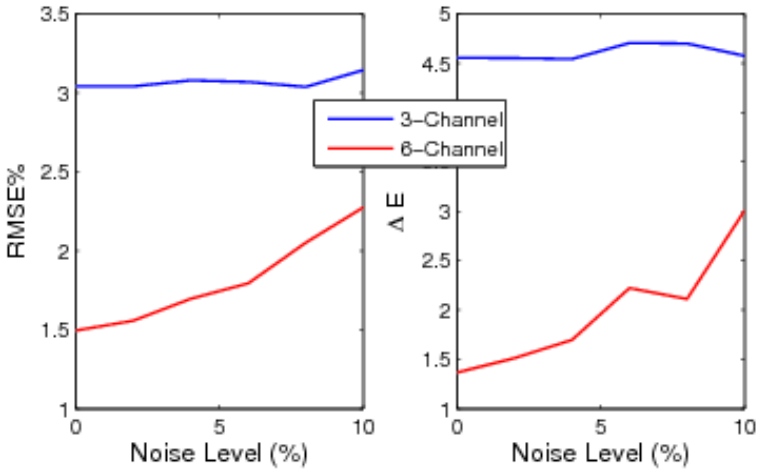


Figure 56: Influence of noise on estimation errors with 3 and 6-channel systems

The result shows that both RMSE and ΔE^*_{ab} estimation errors are significantly lesser with the 6-channel system than with the 3-channel system. However, the errors increase somewhat exponentially with the increase in the noise level in the case of 6-channel system. The increase in the noise level causes relatively small variation (increase) in the estimation errors in the case of 3-channel system. This infers that noise will have greater influence in the performance of the multi-channel multispectral systems.

5.5 Discussion

In this chapter, we have studied our proposed multispectral system with simulation approach and the results from the system have been analyzed and compared with the classic 3-channel system. We have used two different evaluation metrics, RMSE and ΔE^*_{ab} for the evaluation purpose. As expected, the spectral metric:RMSE and the colorimetric metric:ΔE^*_{ab} metric have chosen different pair of filters as the best combination. The choice of filters depends on the estimation methods also. The results from the well known estimation methods: Maloney and Wandell, Imai and Berns, Least Square-Wiener, Polynomial and Neural network methods have been investigated. Even though these methods pick different filter combinations in some cases, there are not much differences in filter types and in the estimation results also. The results also show that the performance of the estimation methods are comparable with more or less similar results based on both the metrics. This indicates more or less the similar performances of these methods. However, the polynomial and neural network methods have advantages that they don't require the camera sensitivities and they provide optimization capability. Yet another observation from the simulation results is that optimizing one metric slightly increases the other metric values, which is obviously expected.

In all the cases of simulation, proposed 6-channel multispectral system outperforms classical 3-channel camera systems, both spectrally and colorimetrically. The improvement is significant with about 2% decrease in RMSE and about $3\Delta E^*_{ab}$ on the average. The simulation results, thus, show promising results in favor of our proposed approach of multispectral imaging with two RGB cameras or a stereo camera. We can see that sensitivities of the cameras are modified in a sensible way and that the dominant wavelengths are quite nicely spread out in the spectrum. This clearly shows that the use of two RGB cameras or a stereo camera with appropriate use of filters can function well as a multispectral system. We can expect further improvement in the results with the selection from a larger set of filters from other manufacturers as well.

A study on influence of noise level on the performance of 3-channel and 6-channel systems has been made and it shows stronger influence of noise level (see Figure 56) on 6-channel than on the 3-channel system. This indicates importance of noise consideration in the multispectral system.

6 Multispectral Acquisition with Mulstispectral-Stereo System: An Experimental Approach

This chapter discusses all the experimental works carried out. It then presents our proposed multispectral acquisition with two digital cameras/a stereo camera based on real experiments.

As discussed in Section 3.7, a stereo camera can be constructed with two digital cameras by joining them together. It takes bit of time and effort to construct, as stereo configuration requires precision work. In the mean time the first new digital stereo camera *Fujifilm FinePix REAL 3D W1* from Fujifilm has recently been out in the market. So, we decided to use it instead of creating a stereo camera with two digital cameras, with the interest of investigating the technology and at the same time avoide hassle of stereo rigging.

The camera has been characterized; the spectral power distribution and the reflectances of the color charts have been measured, and the images of the color charts have been captured. We will first discuss on the devices and instruments used along with their configurations, and then present the experiments carried out with the results obtained from the experiments.

6.1 Devices, Instruments and Tools Used

The devices, instruments and tools used in the experimental studies are described here.

6.1.1 Fujifilm FinePix REAL 3D W1 (Fujifilm3D) Stereo Camera

This camera has been used as the primary device for image acquisition. Figure 57 shows the front and back views of the camera. It has two lenses like our two eyes and twin CCD sensors making it capable to acquire 3D stereo images. It allows to capture both image as well as video in either 2D or 3D mode. Our interest here is in 3D still imaging. Details of the product can be found in Fujifilm (2010*b*).

The camera has some limitations that it doesn't allow to capture image in RAW format and some of its functions are automatic. These We will see these limitations and how these issues are handled in Section 6.1.1.1.

(a) Front view (b) Back view

Figure 57: Fujifilm FinePix REAL 3D W1 camera

6.1.1.1 Camera Configuration

For the study on multispectral image acquisition, we need to have known state of the camera. So the camera is set into manual mode; digital zoom and flash have been turned off all the time. As it provides manual exposure setting, different exposure settings have been tested and for minimal noise and to avoid the saturation, minimum ISO sensitivity of 100, exposure time of 1/50s, and aperture of F3.7 have been selected as the fixed setting for the whole experiment.

There is no way to turn the white balancing off. However, it allows to select certain predefined options and we select *FINE* as the fixed white balance setting, since we normally capture images in the lighting booth in the lab under daylight D50. Since there is no manual focusing, we use the auto focus, and capture the images, as recommended, by pressing the shutter button halfway for sometime to focus the object being captured and then pressing it fully. The camera captures stereo images into MPO (multiple-picture) file format. It allows to select either MPO or a JPG and a MPO files. We choose MPO only option. Software tools can be used to extract the two image files from MPO file. Best image size (L 4:3) of 3648×2736 resolution has been selected for the best possible image quality.

The camera setup and configurations we used are summarized and enlisted below:

Operation mode: **MANUAL** Flash: **OFF**
Exposure: **MANUAL**
ISO Sensitivity: **100**
Exposure time: **1/50s**
Aperture of lens: **F3.7** White balance: **FINE**
3D Recording file format: **MPO**
Image size: 3648×2736 **(10M)**

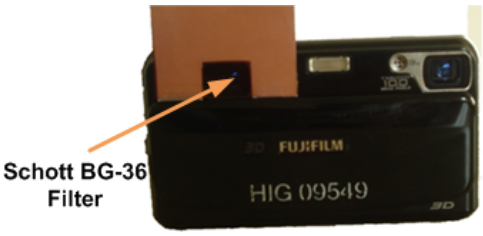


Figure 58: Schott BG-36 filter fitted on the camera

6.1.2 Optical Filter

In the simulation based approach in Chapter 5, a best combination of filters is selected from a set of 265 filters from Omega. We were targeting to do experiments with, if not all, but some of the possible best combinations in the experiment to verify the results we obtained with the simulation. However, because of delays in the arrival of camera and also long time required by Omega for the delivery of filters, it couldn't be materialized. From the simulations with different real and simulated cameras, it showed that multiband filters are good choice for better spreading of the camera sensitivities along the visual spectrum. Based on this, we managed to have a multiband Schott BG-36 filter, and decided to perform experiment with the Fujifilm3D camera, with one lens covered with this filter, while the other without any filter. Figure 58 shows the picture of the Schott BG-36 filter fitted onto the camera.

6.1.3 Color Charts

Like in the simulation approach in Chapter 5, the Macbeth color checker DC (Figure 33) has been used as the training target, and the classic Macbeth color checker (Figure 34) as the test target. And in the same way, the outer surrounding gray patches, the glossy patches in the S-column of the DC chart and inner white patches have been omitted, and the central four white patches are averaged to obtain the representative white patch for the training dataset.

6.1.4 Minolta CS-1000 spectroradiometer

A spectroradiometer is an instrument used to measure the spectral power distributions of illuminants. Minolta CS-1000 spectroradiometer (Figure 59) has been used to measure spectral power distribution of the light source (Daylight D50) in the lighting booth under which the images were captured for our experimental studies.

Figure 59: Minolta CS-1000

6.1.5 X-Rite EyeOne Pro

Figure 60: EyeOne Pro

A brand new EyeOne Pro Spectroradiometer (also known as the i1 Pro) from X-Rite (Figure 60) has been used for measuring spectral reflectances of both the Macbeth color checker DC and classic. The manufacturer claimed it as the only hand-held device that provides the full accuracy of a professional spectrophotometer.

6.1.6 Bentham TMc300 Monochromator

Figure 61: Bentham TMc300 Monochromator

A monochromator is an optical device that transmits a mechanically selectable narrow band of wavelengths of light or other radiation chosen from a wider range of wavelengths available at the input. Bentham TMc300 Monochromator (Figure 61) has been used for spectral characterization of the Fujifilm3D camera. Section 6.2 discusses on the camera characterization. The monochromator slit width has been set to 0.2mm at both entry and exit ends with the use of micrometer. BenWin+ software that came with the monochromator is used to operate and control the monochromator.

6.1.7 Stereo Photo Maker

It is a freely available software in the Internet (SPM 2010). It has many features useful for making stereo photos. It has been used for extracting left and right images from the MPO files captured with the Fujifilm3D camera. It allows to do so in batch mode, unlike the FinePix Viewer software that comes with the Fujifilm3D camera.

6.2 Spectral Camera Characterization

Spectral characterization of the camera determines the sensitivities of the three channels of the camera. Among the two methods discussed in Section 2.3.3, the direct method has been followed for more accurate characterization. For this, an integrating sphere is attached to the exit end of the monochromator to allow to capture diffused output light. Relative sensitivity of the three channels of the camera can be determined by capturing the output from the exit port of the integrating sphere for a range of wavelengths, particularly in the visible range 380-780nm. The process has been carried out in a dark room. The procedure followed is as follows.

1. The monochromator has been turned on for about 20mins to allow it to stabilize. Camera is positioned in front of the exit port of the integrating sphere with a tripod stand. The camera characterization setup is shown in Figure 62.

2. Next, usin BenWin+ software, set a wavelength (starting with 380nm) and time interval of 15s and scan with *Stationary Scan Setup* and capture the output spectrum with the camera.

3. Repeat step 2 for all wavelengths form 380nm to 780nm in the interval of 5nm.

4. Extract the R, G and B values of the three channels from the central rectangular area within the circular region of the image captured for all the wavelengths and compute the average values. Figure 63 shows the pictures of the some of the captured spectrum images. Average value thus computed gives the corresponding relative sensitivity value for a particular channel for the selected wavelength.

5. Repeat step 4 for all the channels for all the wavelengths and for both left and right cameras.

The computed sensitivities of both left and right cameras of the Fujifilm3D camera are already given in Figure 31(b), Chapter 5. Appendix A lists all the camera sensitivity values obtained.

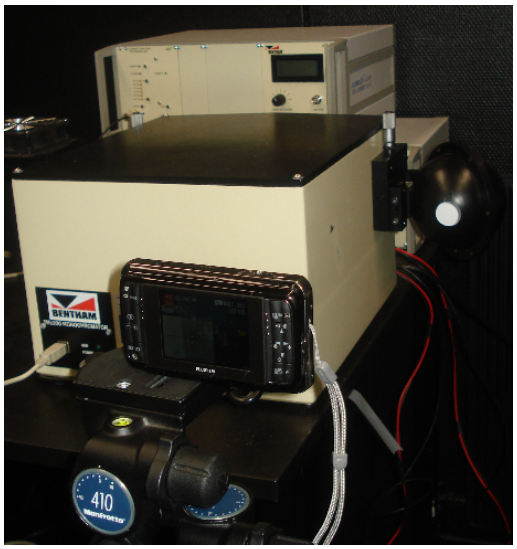


Figure 62: Setup for camera characterization

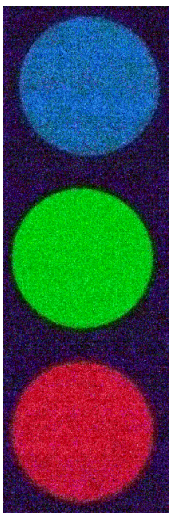


Figure 63: Some of the images of the output spectrum

6.3 Measurement of Spectral Power Distribution

In color science and radiometry, a spectral power distribution (SPD) describes the power per unit area per unit wavelength of an illumination (radiant exitance), or more generally, the per-wavelength contribution to any radiometric quantity (radiant energy, radiant flux, radiant intensity, radiance, irradiance, radiant exitance, or radiosity) (Wikipedia 2010c). SPD of a given light source describes, for every wavelength, how much of the light (or power) it radiates has that particular wavelength. In other words, how is its light distributed across the different wavelengths. An SPD curve is a fingerprint of the light source, depicting the actual distribution of the spectrum colors from 380 to 780nm in the visible light spectrum. The curves provide the user with a visual profile of the color characteristics of a light source. They show the radiant power emitted by the source at each wavelength or band of wavelengths over the visible region.

Minolta CS-1000 spectroradiometer has been used for measuring SPD of the light source (Daylight D50) in the lighting booth. Figure 64 shows the picture of the measurement setup in the lab. Spectroradiometer is set at a distance of 1m from the measurement target. Measurement has been done for the visible range 380-780nm and in $45^o/0^o$ configuration. The spectroradiometer was calibrated with the calibration white the come with the instrument. The spectral radiance measured with the calibration white corresponds to the SPD of the light source used. Figure 65 shows the SPD curve of the light source used in the experiment.

Figure 64: Setup for measurement of SPD

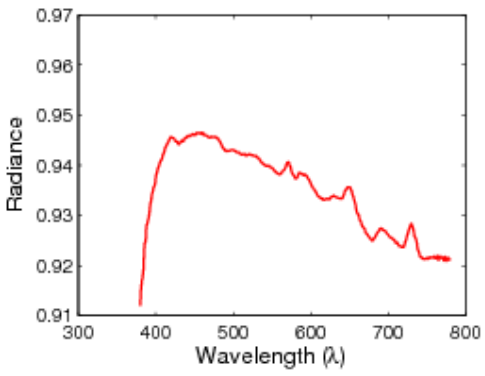


Figure 65: SPD of the light source

6.4 Measurement of Spectral Reflectances

Spectral reflectance of an object refers to how much it reflects the incident light in each wavelength of the visible spectrum. The reflectance factor of a sample is defined as the ratio of the radiant flux reflected in the directions delimited by the cone to that reflected in the same directions by a perfect reflecting diffuser identically irradiated. Spectral reflectances of the patches of both the Macbeth color checker DC and classic are measured using X-Rite EyeOne Pro spectroradiometer. Figure 66 shows the picture of measurement in action.

Figure 66: Measuring spectral reflectances

6.5 Multispectral Image Acquisition with Fujifilm3D Camera

After characterizing the camera and measuring the SPD and the spectral reflectances of the Macbeth color checker patches, the images of the color checker (both classic and DC) are acquired with the multispectral stereo camera: the Fujifilm3D camera with the right lens covered with the Schott BG-36 filter (Figure 58). The color of all the patches of the color charts are computed from the captured images.

The procedure for obtaining the RGB values of the color patches of the Macbeth color checkers are as follows:

1. A color chart is placed on the back wall of the lighting booth in the dark room under the Daylight D50 and the multispectral camera is positioned on a tripod stand to capture its images. Figure 67 shows a picture taken during the image capture session.

2. Capture 10 images of the same chart.

3. From each patch, RGB values of each pixel within the central rectangular area of certain size (100×100 for MCC24 and 40×40 for MCC240) within the patch area are extracted

4. Linearize and correct these extracted values for DC noise and non-uniformity (Sections 6.5.1, 6.5.2 and 6.5.3 discuss on these) and average them.

5. Average the RGB values thus computed from all 10 images of the chart and this averaged RGB value is taken as the value for the particular patch.

6. Perform all above steps for both the color charts: MCC24 and MCC240.

Figure 67: Multispectral image capture of a Macbeth color checker

6.5.1 Linearization

In general, the response of the camera, in particular of the sensor is assumed to be linear. However, in practice, it is not so. We, therefore, need to linearize the camera response, which is done by converting the camera responses to values that are linearly related to the camera input. The non-linear relationship between the light intensity (luminance, Y) input to the camera and the output digital count (C) is usually described by the power law:

$$C = Y^{\gamma}. \tag{6.1}$$

The exponent γ, is called the gamma. When $\gamma < 1$, the nonlinearity is called gamma compression and the gamma value is called an encoding gamma and when $\gamma > 1$, the nonlinearity is called gamma expansion and the gamma is called a decoding gamma.

Since our camera model is based on linearity, we need to correct the non-linearity (also known as gamma correction) and linearize the camera response. Nonlinearity behavior is corrected with $1/\gamma$. Figure 68 shows a typical example of gamma curve for $\gamma = 2.2$ and gamma correction.

In order to compute γ, we take log on both sides of the Equation 6.1 and obtain:

$$\gamma = \frac{\log(C)}{\log(Y)}. \tag{6.2}$$

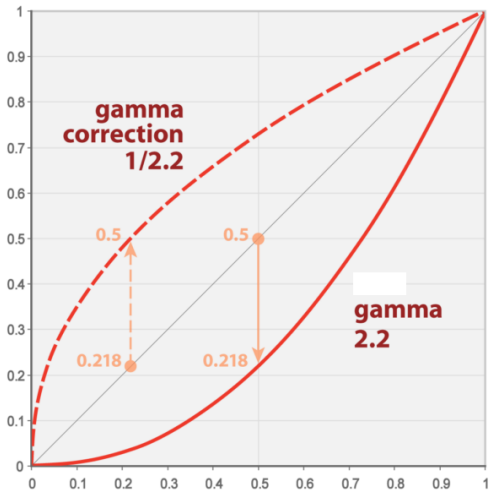


Figure 68: Illustration of gamma correction [Source: Wikipedia]

Knowing the luminances of the central gray patches of the MCC240 (Figure 69) and the corresponding camera outputs, we can estimate γ through optimization using any linear estimation model. The central gray patches whose area being smaller, non-uniformity variation within this area would not be so significant and hence would have lesser effect in gamma correction. Neural network method as discussed in Section 2.3.4 has been used to estimate the gamma values. The luminance values (Y) of the gray patches are computed from the spectral power distribution (l) of the light source, CIE XYZ 1964 color matching functions ($\bar{x}\bar{y}\bar{z}$) and the spectral reflectance (r) of the patch using the relation:

Figure 69: Graypatches used for linearization

$$Y = k \sum l r \bar{y}, \tag{6.3}$$

where $k = 100/\sum l\bar{y}$.

Normally, gamma is assumed to be the same in all three channels of the camera. However, for more accuracy we estimate the gamma values for each channel of both the left and right cameras separately as $\gamma = [\gamma_r, \gamma_g, \gamma_b]^T$. Table 9 gives the luminance values of the gray patches and corresponding camera responses from the left and right cameras.

Table 9: Luminance and camera responses of graypatches

Camera	Patch	I5	I6	I7	I8	J5	J8	K5	K8	L5	L6	L7	L8
	Y	83.57	28.16	16.88	6.99	65.04	4.37	51.82	3.77	41.39	24.18	11.41	3.35
Left	R	170.13	94.03	62.17	31.86	151.43	26.15	133.77	25.43	118.30	83.73	48.54	25.33
	G	177.11	100.54	68.08	34.23	159.47	28.37	141.50	27.26	125.15	89.72	53.05	27.27
	B	175.81	102.26	71.07	39.69	158.45	33.37	142.00	32.29	125.30	91.35	56.96	32.28
Right	R	170.86	93.84	61.98	33.14	152.29	28.56	135.16	28.32	119.66	83.90	50.62	28.97
	G	176.83	99.21	67.06	34.04	158.99	29.93	140.98	29.46	124.84	89.09	53.20	29.76
	B	175.12	100.24	69.48	39.49	158.16	35.12	142.56	34.66	125.95	90.70	58.31	35.55

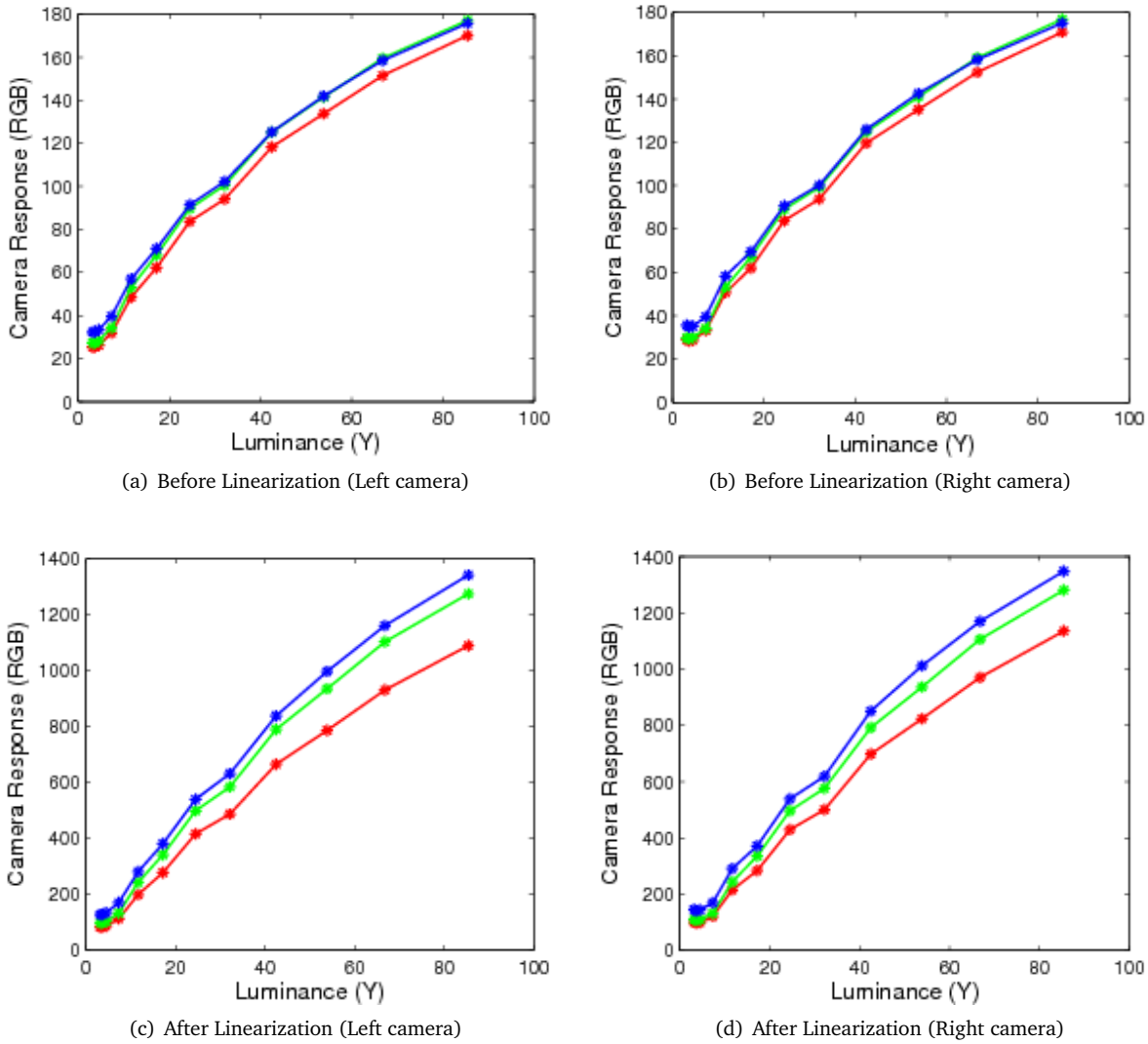
The gamma values of the three channels (RGB) for the left and right cameras have been obtained as follows:

$$\gamma_{(\text{left})} = \begin{pmatrix} 1.3616 \\ 1.3813 \\ 1.3933 \end{pmatrix} \text{ and } \gamma_{(\text{right})} = \begin{pmatrix} 1.3682 \\ 1.3825 \\ 1.3950 \end{pmatrix}.$$

All the images acquired are linearized by obtaining linearized RGB values $[R_L, G_L, B_L]^T$ from the captured RGB values and the gamma values using

$$\begin{pmatrix} R_L \\ G_L \\ B_L \end{pmatrix} = \begin{pmatrix} R^{\gamma_r} \\ G^{\gamma_g} \\ B^{\gamma_b} \end{pmatrix}. \tag{6.4}$$

Figure 70 shows the camera responses of the left and right cameras before and after linearization. Linearization result is also checked with the six gray patches of the MCC24. The camera responses of these patches before and after linearization are shown in the Figure 71. We can see that after linearization, the responses are more or less linear compared to the original responses.



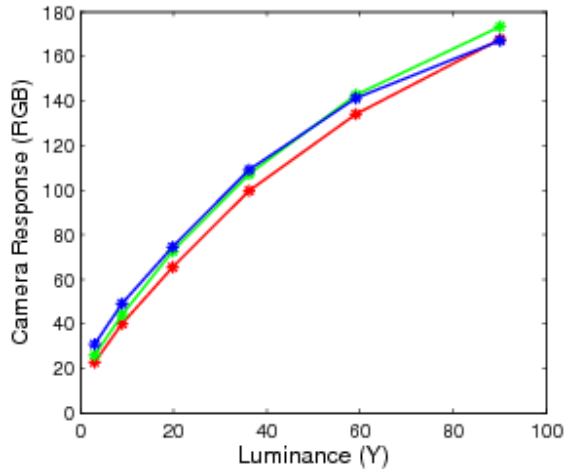
(a) Before Linearization (Left camera)

(b) Before Linearization (Right camera)

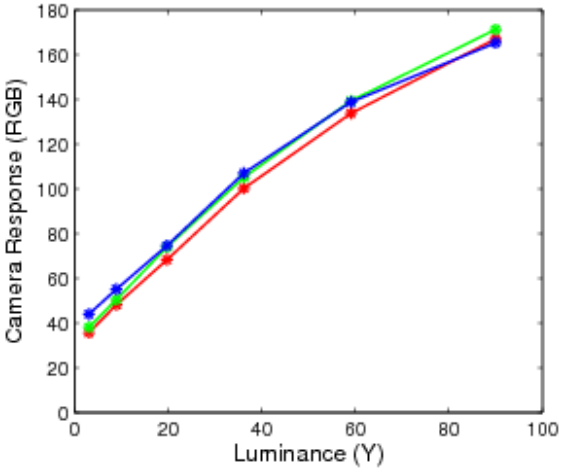
(c) After Linearization (Left camera)

(d) After Linearization (Right camera)

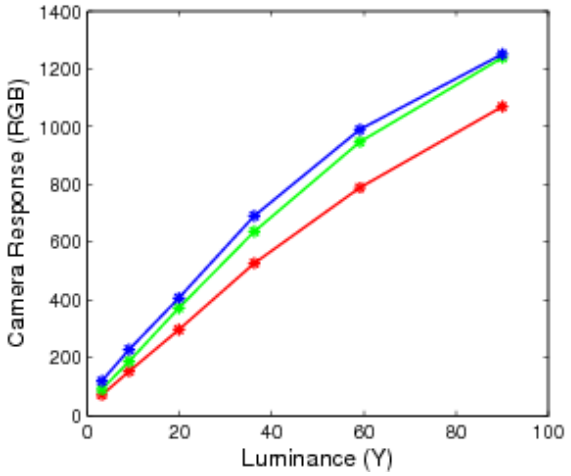
Figure 70: Linearization with graypatches of MCC240



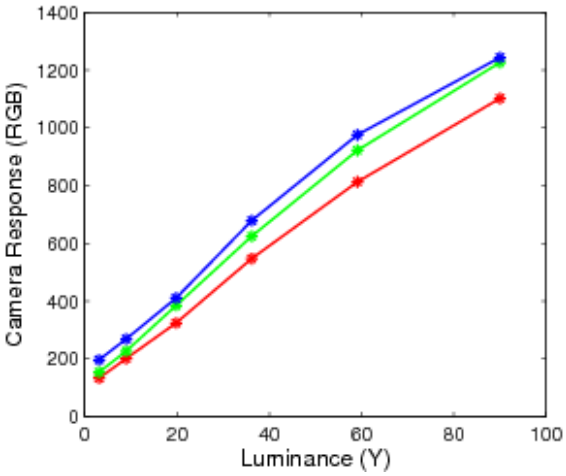
(a) Before Linearization (Left camera)



(b) Before Linearization (Right camera)



(c) After Linearization (Left camera)



(d) After Linearization (Right camera)

Figure 71: Linearization check with graypatches of MCC24

6.5.2 Dark Current (DC) Correction

The dark current (DC)is the relatively small electric current that flows through the camera sensor even when no photons are entering the device. As discussed in the Section 2.2.2, DC is one of the main sources for noise in sensors. Physically, dark current is due to the random generation of electrons and holes within the depletion region of the device that are then swept by the high electric field. DC causes additional offset from zero value in the camera responses which needs to be corrected for more accurate results. The magnitude of these values might be different for different pixels. Taking into consideration all these, the steps followed for computing and correcting DC are given below:

1. 10 images are captured in the dark room by closing both lenses of the camera so that no light enters into the camera.

2. Compute pixelwise average of these 10 images to obtain averaged DC black image (I_{bk}). This will compensate for the temporal variation in the DC noise. Figure 72 shows an averaged DC image with pixel values at some of the positions.

3. A DC corrected image is obtained as $I - I_{bk}$ by pixelwise subtraction of each channel value of the DC image (I_{bk}) from that of the given image (I) after the linearization.

4. The correction has been done separately for left camera and the right camera.

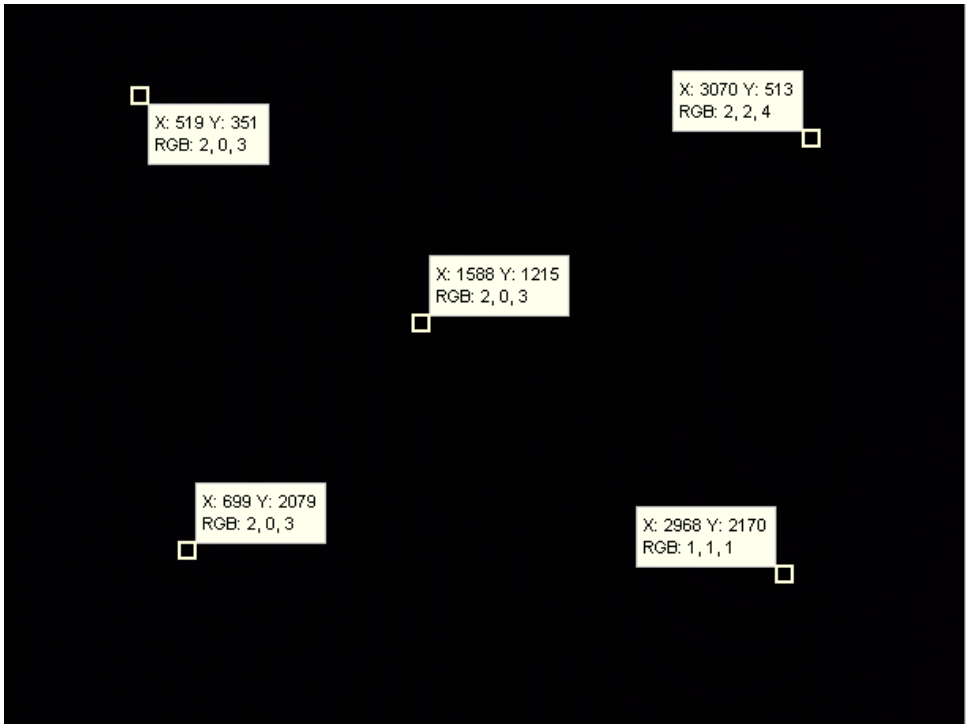


Figure 72: Averaged DC image

6.5.3 Non-Uniformity Correction

Yet another correction to be made is the spatial non-uniformity. The non-uniformity arises due to non-uniform illumination as well as the non-uniform camera sensitivity. Illumination non-uniformity is caused by the variation of light intensity imposed by the unevenness in lighting and optics of the system. In practice, sensitivity of the sensors of the camera also varies spatially. In order to minimize effects of any spatial non-uniformity caused by any of the two types, spatial uniformity correction has been performed. The following steps are carried out for this:

1. Like in DC correction, 10 images of the white diffuser sheet pasted on the back wall of the light booth (in the similar manner as the images were captured) are captured such that the whole image area is covered by the white sheet. Ideally, a perfect white diffuser sheet is necessary for this. However, the one available in the lab has been used even though it is not a perfect diffuser sheet.

2. Compute pixelwise average of these 10 images to obtain averaged WHITE image (I_{wt}). Figure 73 shows an averaged WHITE image with pixel values at some of the positions.

3. Linearize and DC correct it.

4. A non-uniformity corrected image I_{corr} is obtained from the linearized and DC corrected image I using:
$$I_{corr} = \frac{I}{I_{wt}}(\bar{I}_{wt} - \bar{I}_{bk}), \tag{6.5}$$
where $\bar{I}$ denotes the mean values from all the pixels in the image I.

Instead of performing DC correction and then non-uniformity correction one after another, both the corrections can be done from a linearized images in one shot using the equation:
$$I_{corr} = \frac{I - I_{bk}}{I_{wt} - I_{bk}}(\bar{I}_{wt} - \bar{I}_{bk}). \tag{6.6}$$

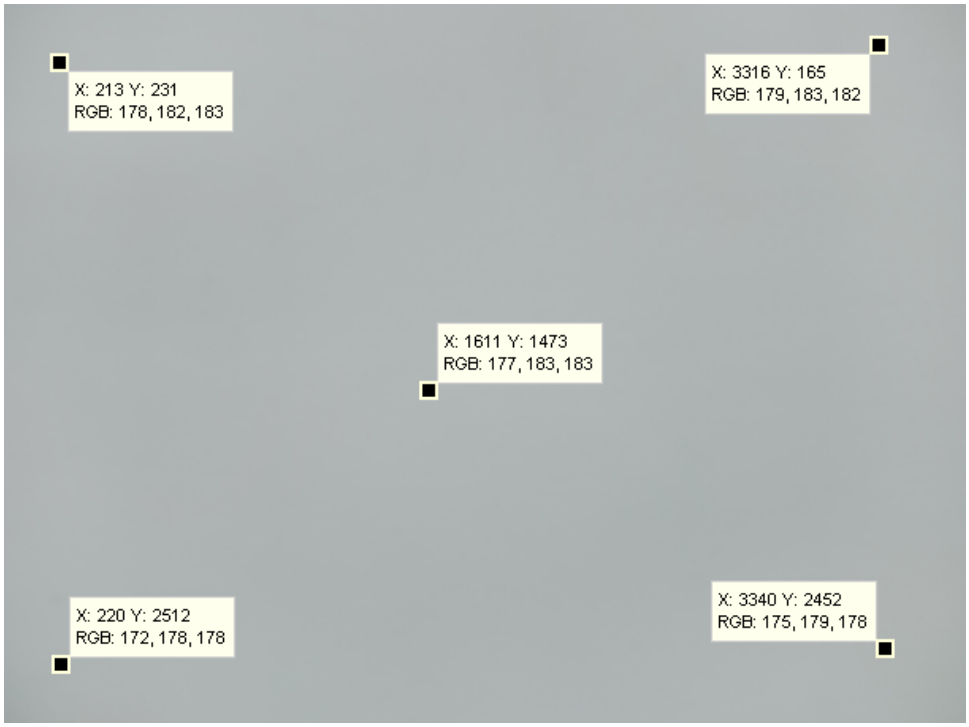


Figure 73: Averaged WHITE image

The corrected images thus obtained are then used in the experiment.

6.6 Experimental Results

The reflectances of the test target (MCC24) are estimated from the experimentally captured images from the multispectral camera. Among the several estimation methods described in Section 2.3.4, linear polynomial and neural network methods have been used, as they are capable to estimate even without knowing the camera sensitivities, which is generally the case in practice.

- *Polynomial Method:* The six RGB values of the color patches three from the left camera:$[R_l G_l B_l]^T$ and the three from the camera $[R_r G_r B_r]^T$ give the camera response of the six channel multispectral camera. Let $C_{(train)} = [R_l G_l B_l R_r G_r B_r]^T$ be the camera response matrix formed from the training target set and $C_{(test)} = [R_l G_l B_l R_r G_r B_r]^T$ the matrix formed from the test target set. Then, polynomials of n degree are formed from these response matrices to obtain $C_{p(train)}$ and $C_{p(test)}$ respectively. Let us denote $R_{(train)}$ as the matrix of reflectances of the training patches and $R_{(test)}$ as the matrix of the reflectances of the test patches.

  The polynomial model directly maps the reflectance from the polynomials of the camera responses by a transformation matrix M. This matrix M is determined from $R_{(train)}$ and $C_{p(train)}$ using the Equation 2.28 as

  $$M = R_{(train)} C^{+}_{p(train)}. \tag{6.7}$$

  To avoid over fitting, the Tikhonov Regularization (see Section 2.3.5 for details) is incorporated while computing the pseudo-inverse $C^{+}_{p(train)}$. A value of the regularization parameter α has been determined from within the range $10^{-1} - 10^{-10}$ that gives the minimum estimation error. The matrix M is then used to compute the estimated reflectance of the test target $\tilde{R}_{(test)}$ from the polynomial of the camera responses using the equation:

  $$\tilde{R}_{(test)} = M C_{p(test)}. \tag{6.8}$$

The degree of polynomials n has been determined by optimizing the estimation error. We have tested with different degrees of polynomials, and also with and without cross-terms. It has been found that the polynomial of degree 2, and without the cross terms give reasonably good results, and hence we have used these in our experiments. Higher order polynomials and with cross-terms are more prone to the influence of noise, particularly with the 6-channel system. Estimation of the spectral reflectances of the test target has been done for 3-channel as well as 6-channel multispectral systems. Since the left and right cameras are similar, the left camera is taken in the estimation as the 3-channel system.

The estimated reflectances of the test target thus obtained are evaluated with the measured reflectances (Section 6.4) using spectral metric RMSE and the colorimetric metric ΔE^*_{ab} (see Section 2.4 for details on these metrics). Figure 74 shows the plots of the estimated and measured reflectances of the 24 patches of the MCC24 for the 3-channel system and Figure 75 shows the plots for the 6-channel system. The plots also show the mean RMSE and ΔE^*_{ab} values.

- *Neural Network Method:* Hetereo-associative memory linear neural network method has been used. Section 2.3.4 discusses this model in details. Among the two parameters, learning rate (η) is determined through optimization such that the estimation error is minimized. The other parameter, number of iterations ($niter$) is set to a high value of 50000. Using these parameters, weight matrix w is obtained from the training reflectance $R_{(train)}$ and corresponding camera response $C_{(train)}$ from the neural network Delta rule using Equation 2.29.

  The w matrix is then used to compute the estimated reflectance of the test target $\tilde{R}_{(test)}$ from the camera responses $C_{(test)}$ using

$$\tilde{R} = wC_{(test)} \tag{6.9}$$

  The results of the estimation for 3-channel and 6-channel systems are shown in Figures 76 and 77.

Statistics of estimation errors for the two methods are given in Table 10.

Table 10: Statistics of estimation errors

Method	**3-Channel System**						**6-Channel System**					
	RMSE %			ΔE^*_{ab}			**RMSE %**			ΔE^*_{ab}		
	Max	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**	**Max**	**Mean**	**Std**
Polynomial (degree 2)	9.36	4.38	2.18	20.66	7.03	4.41	7.06	3.70	1.63	20.90	6.40	4.66
Neural Network	15.50	5.55	2.74	14.92	7.26	3.51	7.06	3.78	1.65	20.83	6.45	4.82

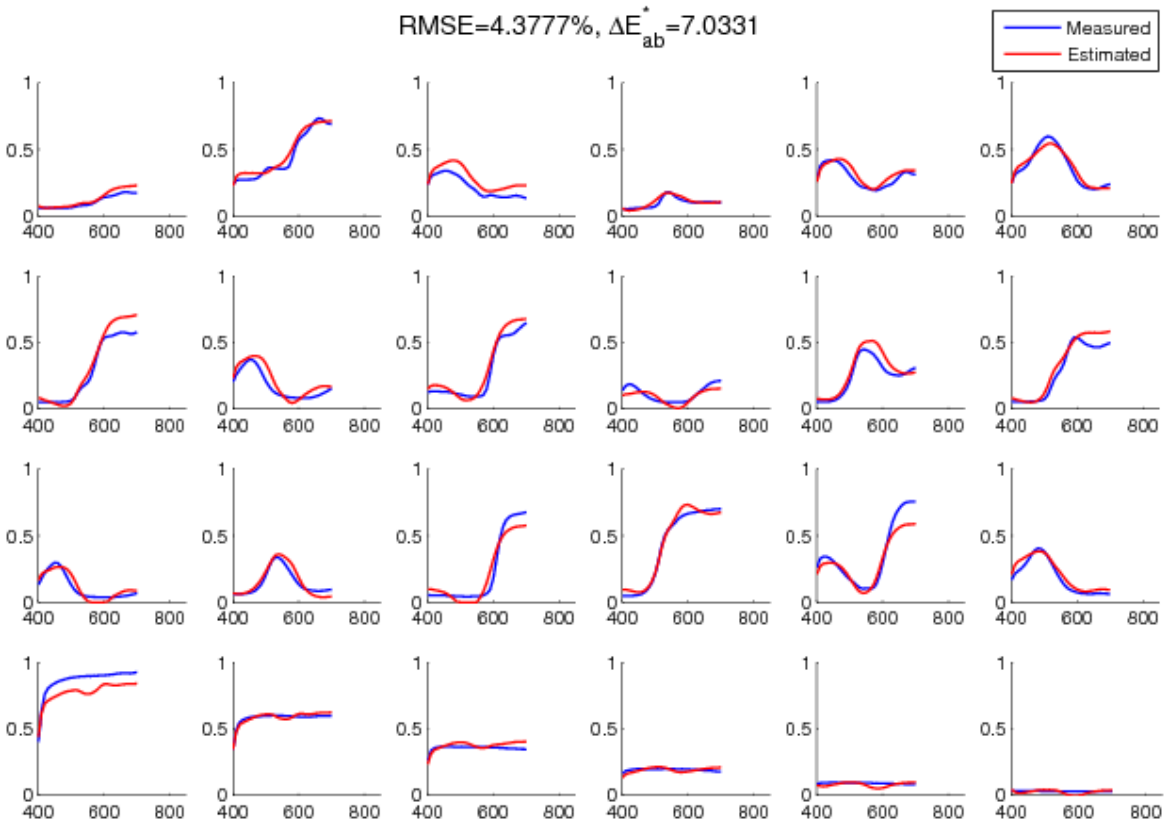


Figure 74: Reflectances with 3-channel system (Polynomial method)

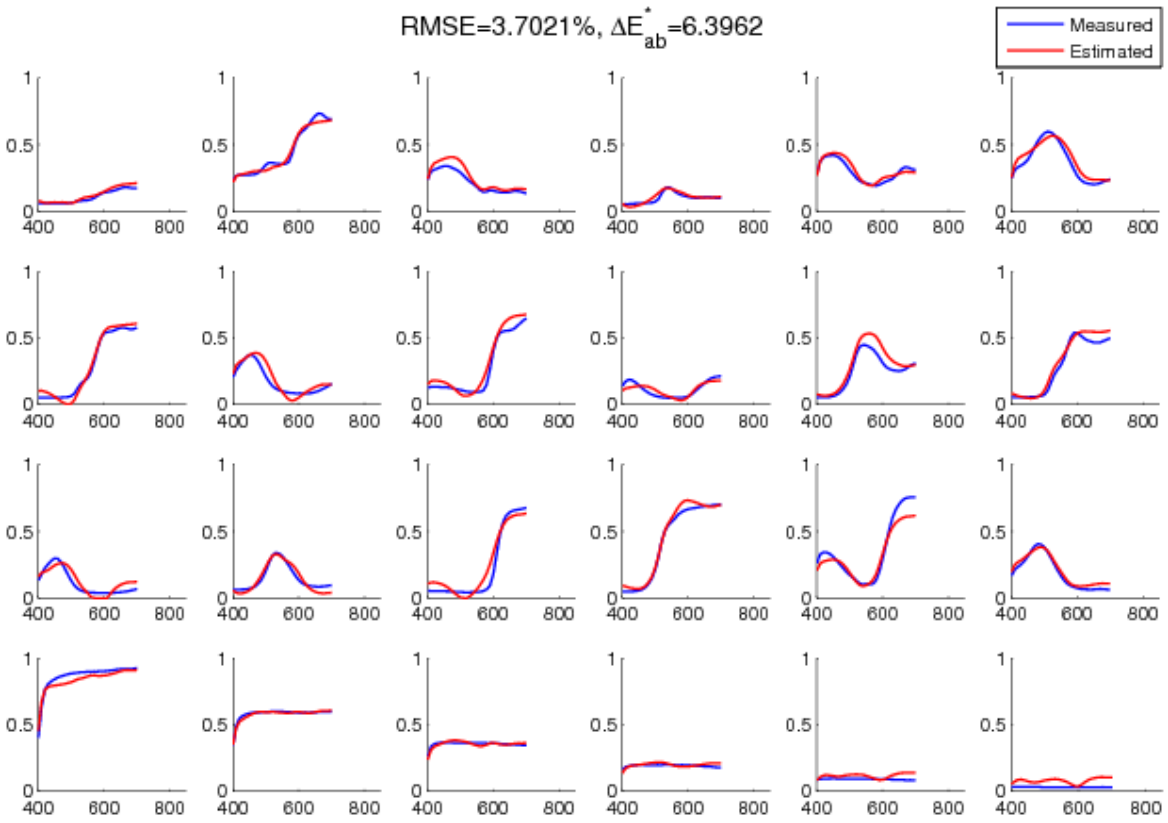


Figure 75: Reflectances with 6-channel system (Polynomial method)

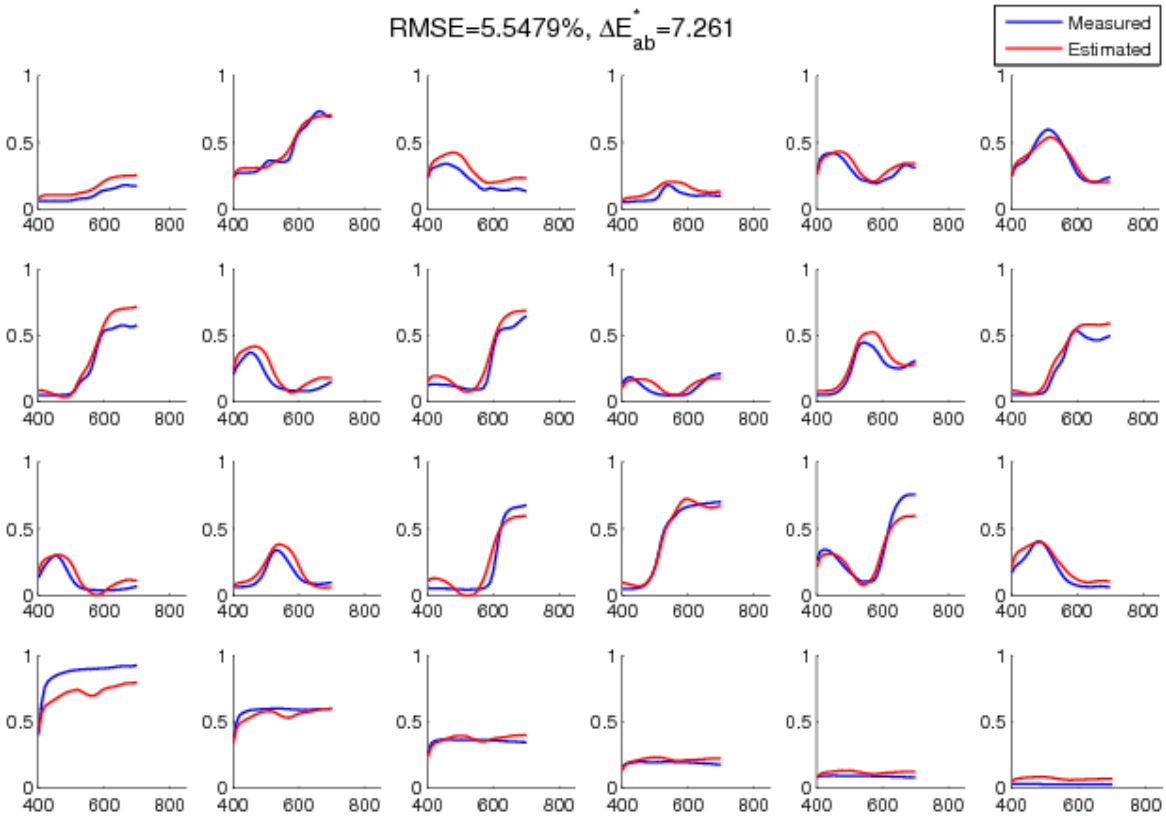


Figure 76: Reflectances with 3-channel system (Neural Network method)

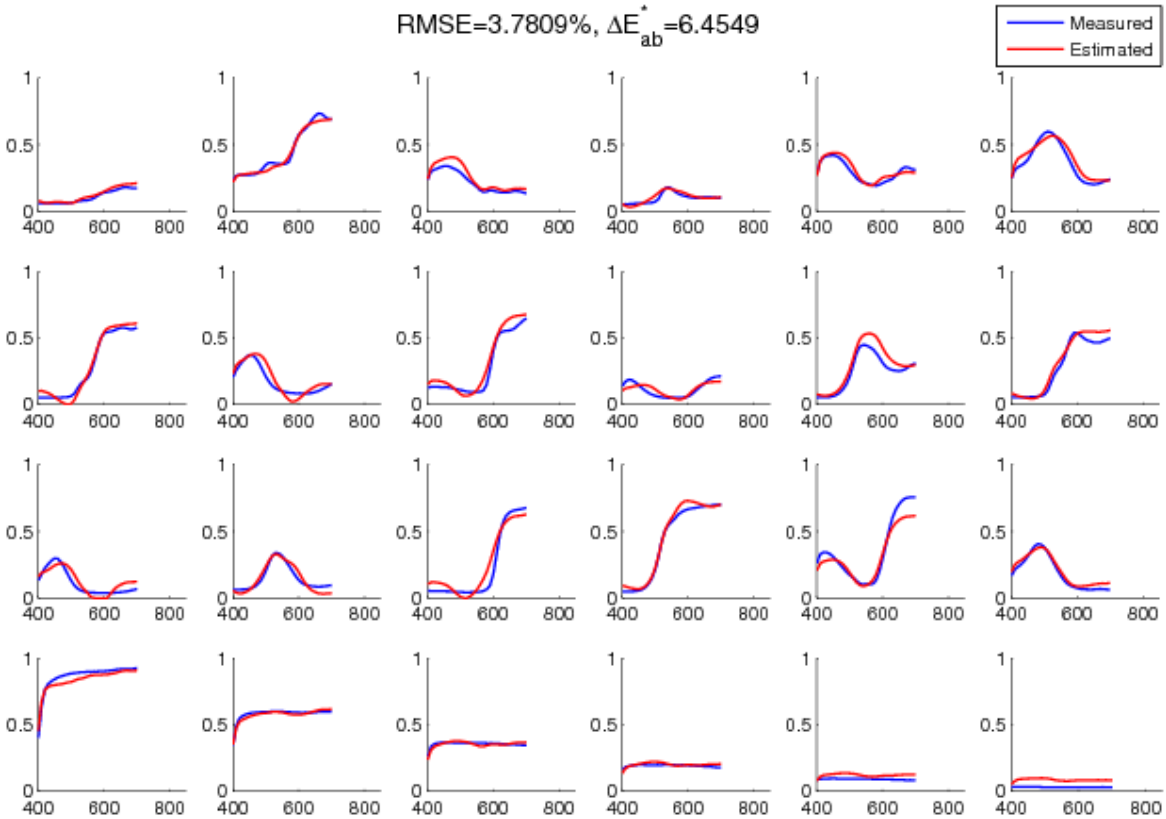


Figure 77: Reflectances with 6-channel system (Neural Network method)

6.7 Discussion

Here, we did all the measurements, spectral power distribution of the light source used during the experiment, spectral reflectances of the Macbeth color charts, and also characterized the Fujifilm 3D camera. Despite limited control with the camera as it has automatic white balancing, focusing and without raw data, we managed to carry out experiment by setting the camera into known, pre-defined configurations.

We have performed experiments and analyzed the results obtained with the proposed 6-channel multispectral system, constructed from Fujifilm 3D and a multiband Schott BG-36 filter, and compared with the 3-channel camera. Statistics of estimation errors from the results shows that the proposed 6-channels multispectral system performs better than the 3-channel system both spectrally (RMSE) and colorimetrically (ΔE^*_{ab}) in terms of mean metric values. Both polynomial and Neural network methods result in favor of proposed 6-channel multispectral-stereo system with improvement in error estimations of more than 1%RMSE and about $0.7\Delta E^*_{ab}$.

Previous work discussed in Section 4.1 mentioned the improvement of average and maximum ΔE^*_{ab} of HDTV camera from 4.1 and 8.2 down to 1.4 and 4.2 respectively. They obtained the result with specifically custom designed filters. Here, it is important to note that the filter we are using is not the optimal one, even so it still shows improved result with the 6-channel system. This clearly infers that the results we obtained are quite encouraging, and we can expect much better result with the optimal pair of filters and more controllable camera.

7 3D Stereo Image Acquisition with Mulstispectral-Stereo System

Stereo aspects of experimental works in this thesis work are presented in this chapter. We have used the 3D stereo camera (Fujifilm FinePix REAL 3D W1) itself from Fujifilm. So it can be assumed that the camera is in proper stereoscopic configuration. Left and right cameras are first calibrated and then using the resulting parameters, stereo calibration is performed to obtain the intrinsic and extrinsic parameters. Knowing these parameters, disparity maps for a given pair of stereo images are computed. With the disparity values, depth-map can be computed. We present how these steps are performed in this chapter. Disparity maps computed from the stereo images taken with the multispectral camera (with the Schott filter in front of the right lens) and with simply the stereo camera without the filter are compared. For faster and memory efficient processing, program has been written in Visual C++ 2008 using the OpenCV (**O**pen source **C**omputer **V**ision) version 2.1 (OpenCV 2010), a library of programming functions for real time computer vision. The functions used are listed in Appendix D with brief descriptions.

7.1 Stereo Image Capture and Camera Configuration

For a stereo image, the Fujifilm3D camera is set to 3D and image mode, and the 3D image is captured by pressing the shutter button. Shutter button is first pressed halfway by focusing the object to be imaged and once focused, the shutter button is fully pressed to capture the image. The camera allows to view the 3D image/video captured by automatic or manual parallax adjustment.

Most of the camera configurations to be set are discussed in Section 6.1.1.1. We use the same configuration for the experiment on stereo as well. 3D images captured are stored in MPO format and the left and right images are extracted from MPO file using Stereo Data Maker software tool.

Intrinsic and extrinsic parameters (see Section 3.4 for details) of the cameras are determined with the individual as well as stereo camera calibration. Section 3.4 discussed about theories on camera calibration and intrinsic and extrinsic parameters that can be determined through the calibration process. At first left and right cameras are calibrated separately and then the stereo camera calibration is performed. Section 7.2 below presents the calibration process.

7.2 Camera Calibration Process

Camera calibration is carried out by analyzing chessboard images taken with the camera and based on the theories presented in Section 3.4. A readymade checkerboard especially designed for camera calibration has been downloaded from the web (Bouguet 2006), printed with a high resolution printer and pasted on a hard board as shown in the Figure 78. Both sides of the squares are measured and found to be the same: 30mm. 30 images of the chessboard are taken with the camera in 3D mode with different orientations of the board. Among them, 20 images are used in the camera calibration. The thumbnails of the 20 images used are given in Figure 79.

The steps carried out in the calibration process are as follows.

- *Extraction of Corners:* For all the images from both left and right images, the inner corners of the chessboard are first extracted using OpenCV function `cvFindChessboardCorners()`, and then the accuracy of thus obtained coordinates is improved to about 0.1 pixel using the function `cvFindCornerSubPix()`. Figure 80 shows an example of successfully extracted corners.

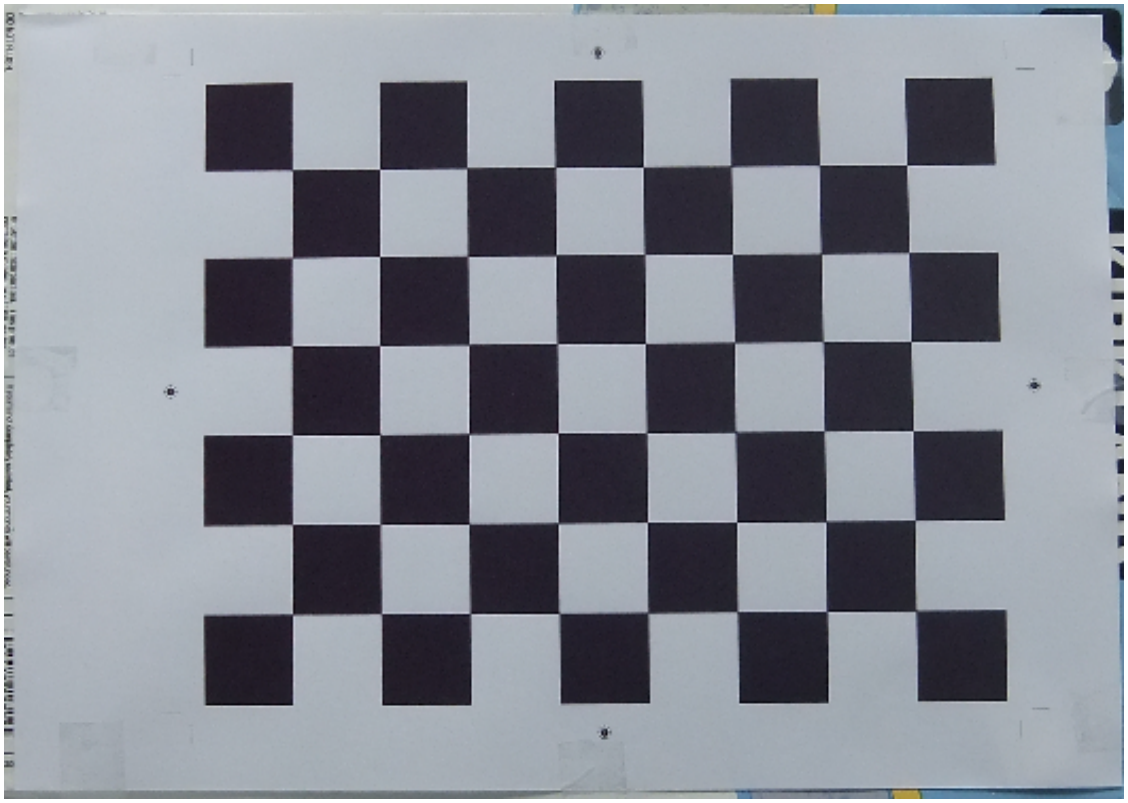


Figure 78: Chessboard used for camera calibration

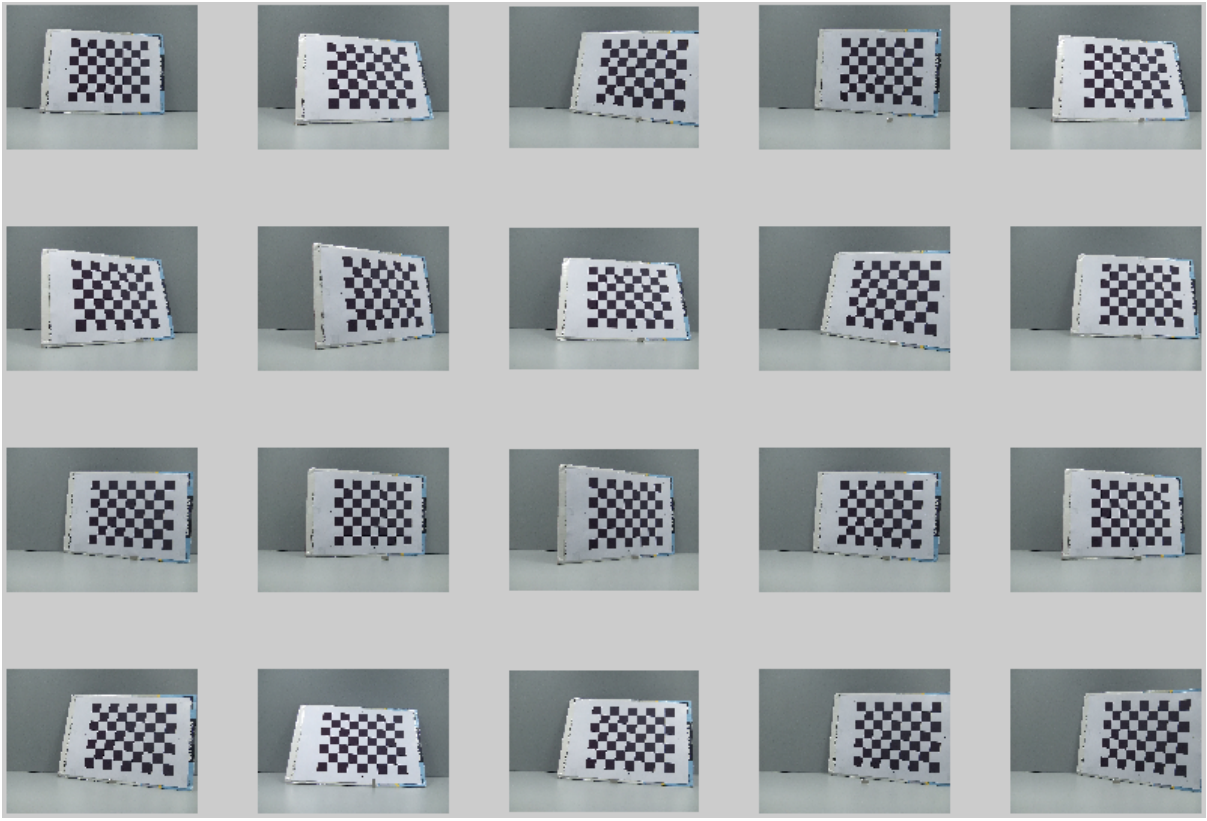


Figure 79: 20 chessboard images used for camera calibration

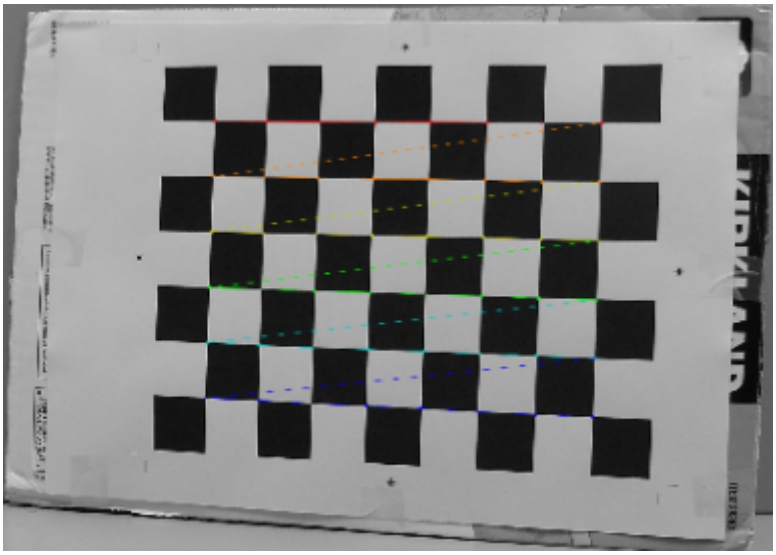


Figure 80: Extracted inner corners

- *Calibration of Camera:* Then both the left and right cameras are calibrated separately using the OpenCV function `cvCalibrateCamera2()`. It outputs the camera matrix, distortion coefficients, rotation vector and translation vector for the two cameras.

  Camera matrix (M) contains the focal length ($fc = [f_x, f_y]^T$) and the principal point ($cc = [c_x, c_y]^T$) as shown below:

$$M = \begin{pmatrix} f_x & 0 & c_x \\ f_y & 0 & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

- *Checking Reprojection Error:* Using the parameters thus computed for the two cameras, reprojection errors are computed using the OpenCV function `compute_reprojection_error()`. Reprojection errors are normally about 0.1 of a pixel. If in case reprojection errors are high, the process should be repeated by removing the erroneous image and adding a new image until a reasonably lower reproduction errors are obtained.

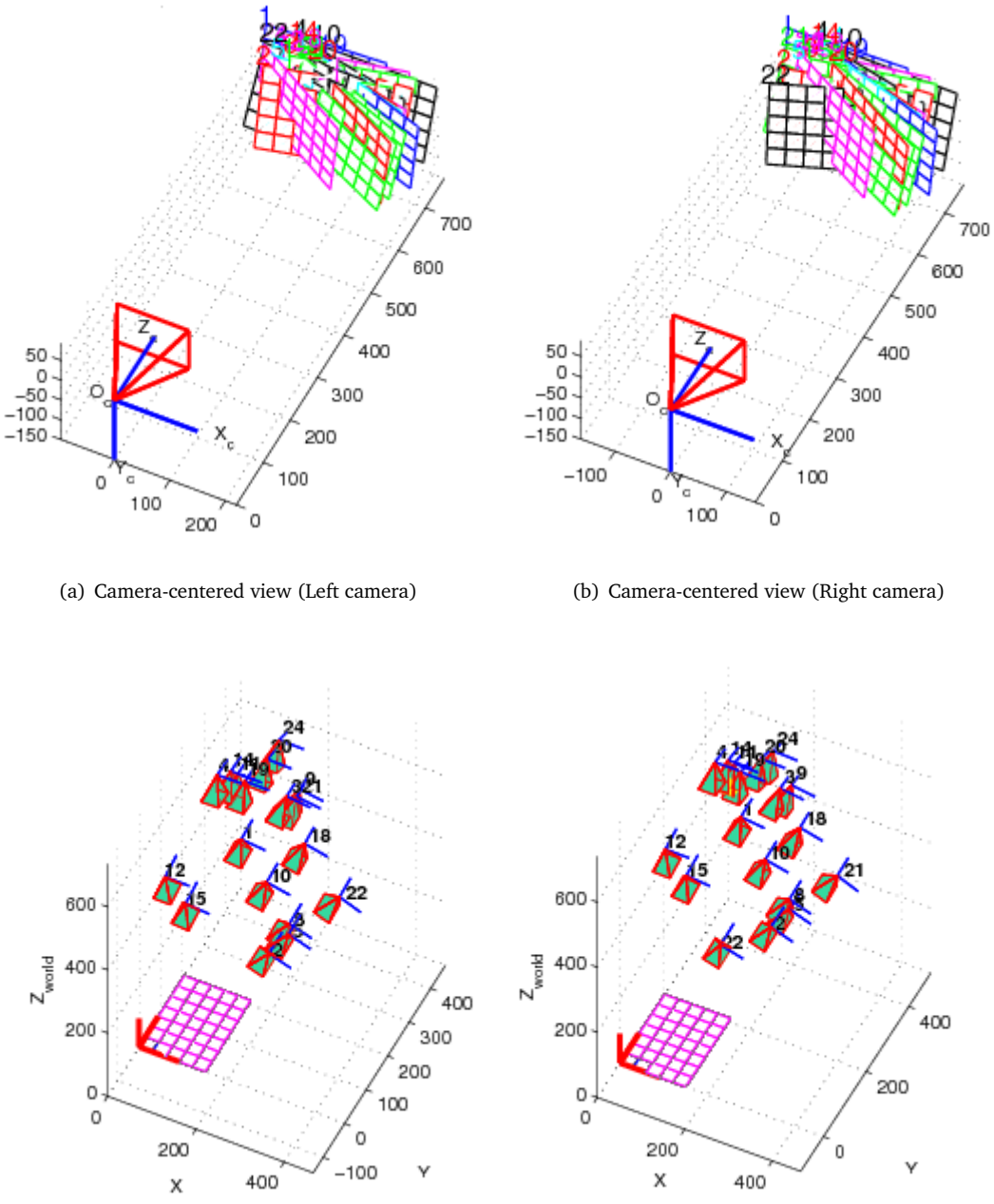
The camera matrices, M_l and M_r for the left and the right cameras resepectively and the corresponding distortion coefficients, d_l and d_r obtained from the camera calibrations are given below:

$$M_l = \begin{pmatrix} 4257.4 & 0 & 1877.3 \\ 4304.0 & 0 & 1302.4 \\ 0 & 0 & 1 \end{pmatrix} \qquad M_r = \begin{pmatrix} 4268.84 & 0 & 1916.69 \\ 4319.29 & 0 & 1281.41 \\ 0 & 0 & 1 \end{pmatrix}$$

$$d_l = (-0.1486, 0.3264, 0.0017, 0.0031) \quad d_r = (-0.1448, 0.5735, -0.0036, -0.0007)$$

All the parameter values and the average reprojection errors will be saved into a .YML file. The calibration result for the left camera is saved in **calibleft.yml** and the result for the right camera is saved in **calibright.yml**. Contents of these calibration output files are given in Appendix C.

The extrinsic parameters for both left and right cameras for 20 images are shown in a form of 3D plots in camera-centered and world-centered views in Figure Figure 81. In Figures 81(a) and 81(b) (camera-centerd views), the frame (O_c, X_c, Y_c, Z_c) is the camera reference frame. In Figures 81(c) and 81(d) (world-centered views), every camera position and orientation is represented by a green pyramid.



(a) Camera-centered view (Left camera)

(b) Camera-centered view (Right camera)

(c) World-centered view (Left camera)

(d) World-centered view (Right camera)

Figure 81: Extrinsic parameters

7.3 Stereo Calibration

As discussed in Section 3.5, stereo calibration uses the intrinsic and extrinsic parameters computed for individual left and right cameras as discussed in the above Section 7.2 and applies geometric transformation of the extrinsic parameters to find out the geometry of the stereo setting.

The stereo calibration is performed using OpenCV function `cvStereoCalibrate()`. It takes object points output from corner extraction, camera matrices and distortion coefficients of the left and right cameras among others as inputs and outputs the rotation matrix and the translation vector between the camerasŠ coordinate systems. Besides the stereo-related information, the function can also perform full calibration of each of the 2 cameras. However, because of the high dimensionality of the parameter space and noise

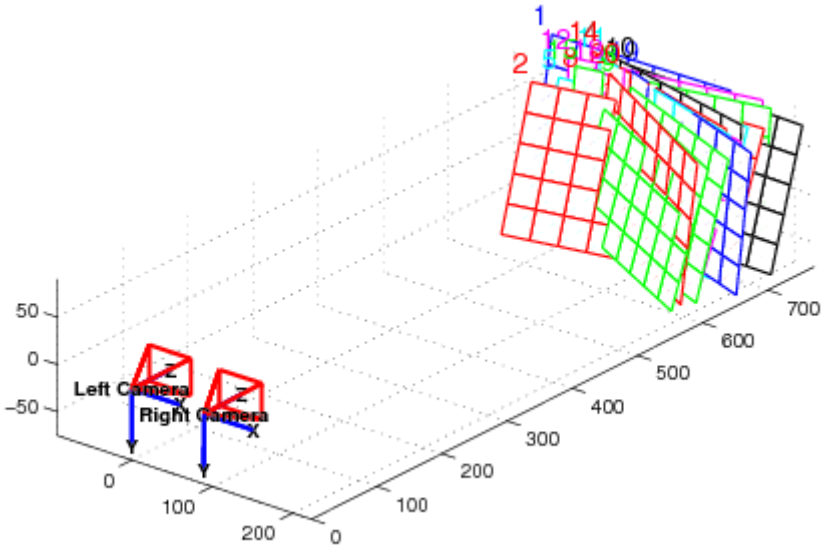


Figure 82: 3D view of extrinsic parameters of 20 calibration stereo image pairs

in the input data the function can diverge from the correct solution. Thus, if intrinsic parameters can be estimated with high accuracy for each of the cameras individually, it is recommended to do so and then pass `CV_CALIB_FIX_INTRINSIC` flag to the function along with the computed intrinsic parameters.

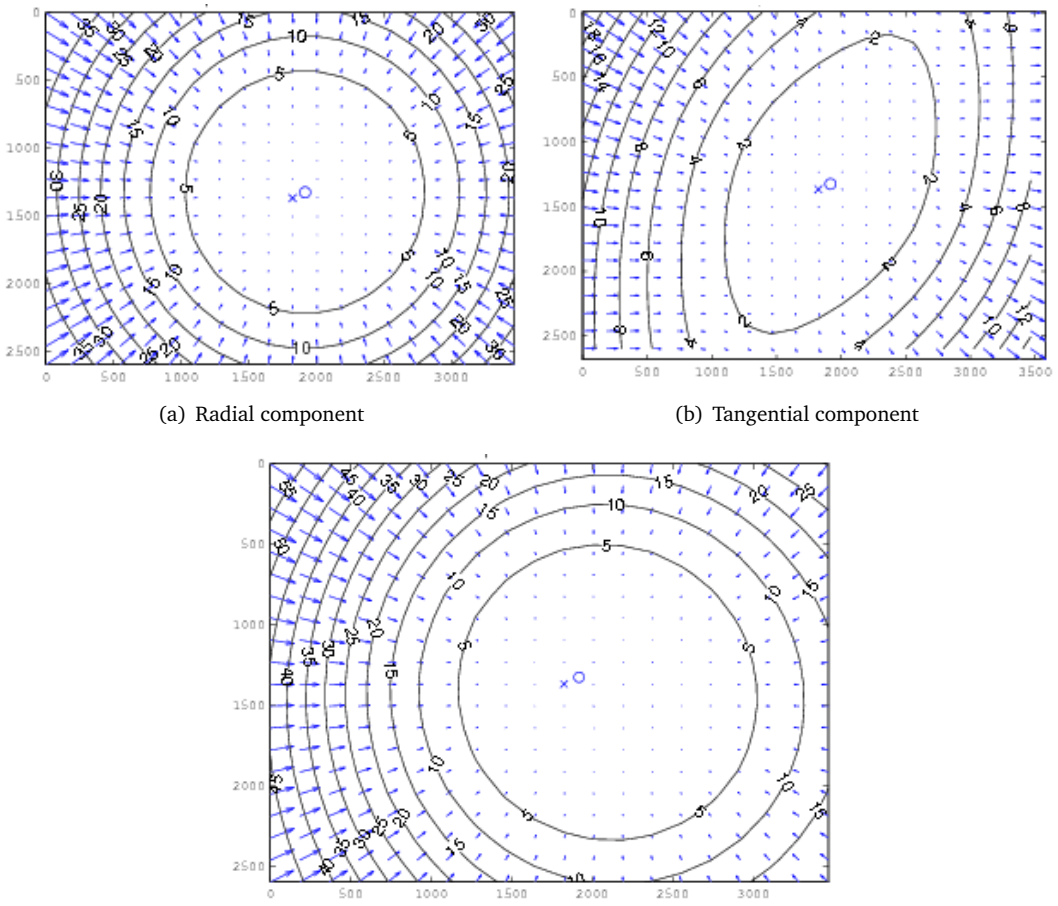
The stereo calibration gives the intrinsic and extrinsic parameters with smaller uncertainty. The camera matrices (M_l and M_r), distortion coefficients (d_l and d_r) of the left and the right cameras and the rotation matrix R, and the translation vector T obtained from the stereo calibration are given below in Table 11.

Table 11: Intrinsic and extrinsic parameters from stereo calibration

$M_l = \begin{pmatrix} 4253.2 & 0 & 1865.3 \\ 0 & 4301.3 & 1302.4 \\ 0 & 0 & 1 \end{pmatrix}$	$M_r = \begin{pmatrix} 4268.5 & 0 & 1923.7 \\ 0 & 4318.7 & 1288.5 \\ 0 & 0 & 1 \end{pmatrix}$
$d_l = (-0.22897, 1.505, 0.0013403, 0.0050117)$	$d_r = (-0.167, 1.7223, -0.0032161, -.000042)$
$R = \begin{pmatrix} 0.99963 & 0.0090 & 0.02566 \\ -0.00917 & 0.9999 & 0.00473 \\ -0.02562 & -0.0050 & 0.99966 \end{pmatrix}$	$T = \begin{pmatrix} -89.5480 \\ -1.2500 \\ 2.0085 \end{pmatrix}$

Note that the focal lengths and the principal points in the camera matrics are in pixel units, whereas the translation values in the translation vector T are in millimeters. Figure 82 shows 3D view of extrinsic parameters for 20 stereo pairs of calibration images. Figure 83 shows the radial, tangential components and the complete distortion plots for the left camera of the Fujifilm 3D as obtained from the camera calibration. Plots for the right camera are similar as the two cameras are more or less identical.

The intrinsic and extrinsic parameters output from the stereo calibration are saved into **intrinsics.yml** and **extrinsics.yml** files respectively. The contents of these files are given in Appendix C.



(a) Radial component

(b) Tangential component

(c) Complete distortion model

Figure 83: Distortion plots of the left camera

7.4 Stereo Rectification

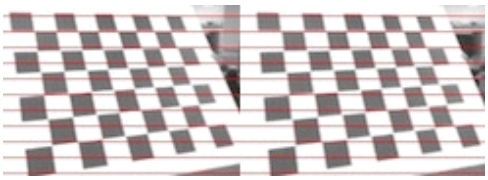


Figure 84: Illustration of stereo rectification

After stereo calibration, rectification transforms are computed for each head of a calibrated stereo camera using the function `cvStereoRectify()`. The function computes the rotation matrices for each camera that (virtually) make both camera image planes the same plane, and consequently, make all the epipolar lines parallel and thus simplifies the dense stereo correspondence problem. It takes the matrices computed by `stereoCalibrate()` as input and outputs two rotation matrices, one rotation and one translation matrices between 1st and the 2nd camera's coordinate systems, and also 2 projection matrices in the new coordinates. It also outputs disparity-to-depth mapping matrix. Any pair of stereo images are stereo rectified using the rectification transforms obtained with `cvStereoRectify()` and rectified images are then used for stereo matching. Figure 84 illustrates the rectified stereo image pair. It shows some red horizontal lines pass through the corresponding image regions, i.e. the images are well rectified.

7.5 Evaluation of the Stereo System

The stereo sub-system of the proposed multispectral-stereo system has been evaluated in two ways. The first approach of evaluation is by observing the 3D view of the images acquired by the system. The other approach is through quantitative analysis of the disparity maps obtained from the stereo images. The following subsections discuss on these.

7.5.1 3D View from Captured Stereo Images

When we view the stereo images captured from our experimental Fujifilm3D stereo camera through its screen or through a 3D viewer, we can see the 3D image. Now it is time to investigate if we can have 3D view with our multispectral system constructed with the Schott BG-36 filter covering one of its lens (the right one). And yes we can view the 3D image well except that the brightness is slightly diminished, which is expected as the introduction of filter in front of a camera lens lowers the intensity of the light to some extent, which can be seen from an example stereo images (left and right views) given in Figure 85. Having the stereo image pair, we can view the 3D image using any of the viewing methods discussed in Section 3.9.

Figure 85: An example stereo image pair taken with our multispectral camera system

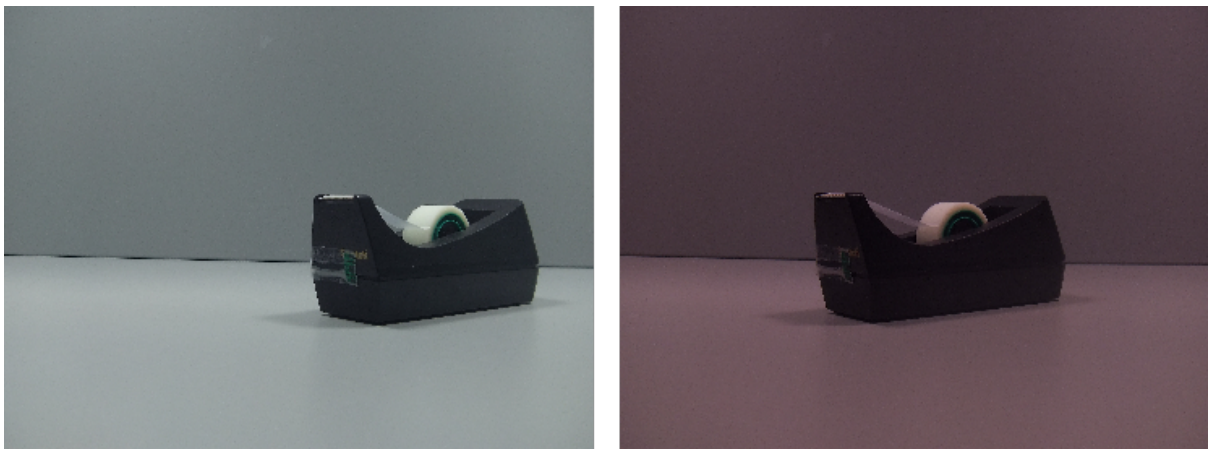
7.5.2 Disparity Mapping

The disparity maps from the stereo images acquired with the proposed multispectral-stereo system and from the pure stereo system (i.e. without the filter) are compared, taking the later one as the reference. We analyze how close the maps to the reference maps.

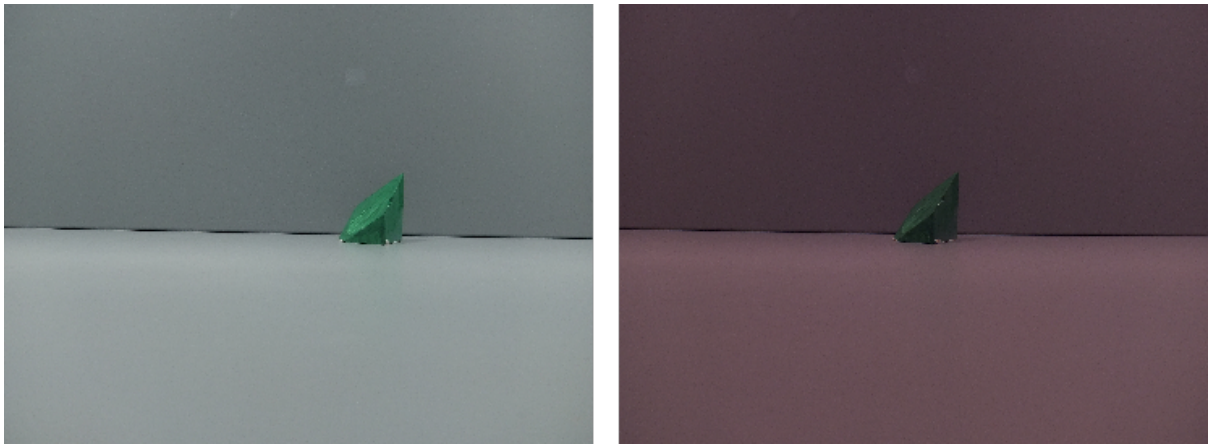
In order to compute disparity map, we need to match the corresponding points in the left and right images, the problem known as stereo matching or correspondence problem. Section 3.6 discusses stereo matching in details. Most commonly used stereo matching algorithm: Block Matching has been used. In order to facilitate faster search, the stereo image pair is rectified using the rectificatin transformation matrices as discussed above in Section 7.4. With rectified images, algorithms need to search for the corresponding blocks in one direction (horizontal or vertical) only.

Block Matching (BM) is the most widely used method for disparity estimation, and is simple and effective to implement. Tao et al. (2008) proposed a fast block matching algorithm for stereo correspondence. The basic idea of block matching is to segment the target image into fixed size blocks, and find for each block the corresponding block that provides the best match from the reference image. In general, the block minimizing the estimation error is usually selected as the matching block. However, block matching with a simple error measure may not yield smooth disparity fields, and thus may result in increased entropy of the disparity field and therefore increased bit rate of the disparity field.

Several images are taken to investigate the disparity maps from the stereo camera with and without filter. Figure 86 shows two different images from both left and right cameras for illustrations. Disparity maps of these images computed from BM algorithm are given in Figure 87. Left column images are disparity maps for the images captured with the multispectral-stereo camera, and the right column images are disparity maps for images captured with the pure stereo camera.



(a)



(b)

Figure 86: Stereo images (Left and right views)

We now make comparisons between the disparity maps thus obtained.

7.5.2.1 Comparison of Disparity Maps

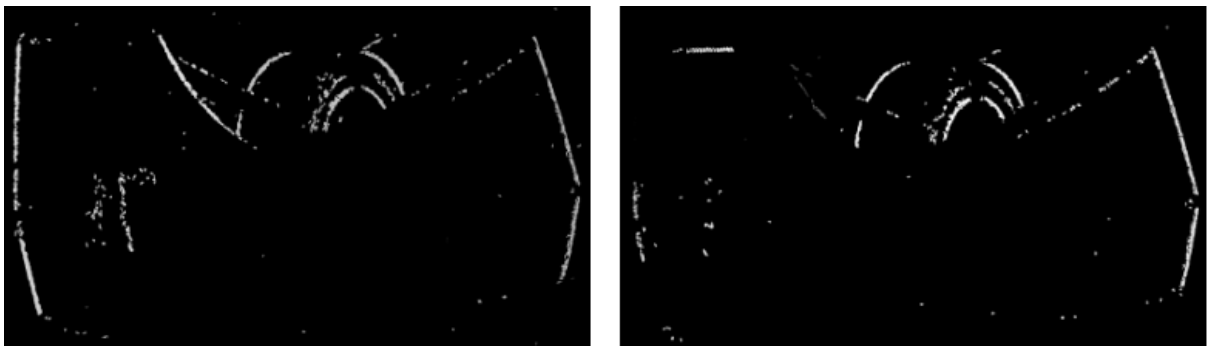
We evaluate the multispectral-stereo system based on how close the disparity maps from the stereo images acquired from it (Say, D_{ms}) with the maps created from the pure stereo system, D_{ps}. It has been found that there are some shifts in the disparity maps in the multispectral-stereo system due to the change in optics caused by the introduction of the filter. This suggests that it is not a good idea to compare direct pixel-to-pixel comparison. So, we come up with the idea of using precision and recall, two widely used statistical metrics, as the evaluation metrics of the system.

Precision can be defined as the number of relevant pixels generated as valid pixels in the disparity map of a system divided by the total number of valid pixels in the disparity map.

Recall can be defined as the number of relevant pixels generated divided by the total number of relevant pixels (which should have been generated and this is given by the valid pixels from D_{ps}). Precision can be seen as a measure of exactness, whereas Recall is a measure of completeness.

To accommodate for the pixel shifts, a threshold of shift tolerance, τ can be set. A valid pixel in D_{ms} is considered as relevant if it is in the same pixel or within τ neighbors. We have used 10 pixels tolerance based on the shifts check from the disparity images. Precision and recall values computed in this way from the disparity maps shown in Figure 87 are given in Table 12.

The results show that, on the average, the precision is about 65% and the recall is about 53%.



(a)

(b)

Figure 87: Disparity maps with BM method(Left - Stereo, Right - Multispectral stereo)

Table 12: Precision and recall from disparity maps

Image	**No. of Pixels**			**Precision (%)**	**Recall (%)**
	relevant (all)	**generated (all)**	**generated (relevant)**		
a	62575	45768	32382	70.75	51.75
b	27386	24929	14827	59.48	54.14

7.6 Depth Map

Knowing the disparity map, we can compute a 3-channel image representing 3D surface using a projection transformation matrix Q obtained with `cvStereoRectify()` function. That is, for each pixel (x, y) and the corresponding disparity $d = \text{disparity}(x, y)$ it computes:

$$\begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = Q \begin{pmatrix} x \\ y \\ d \\ 1 \end{pmatrix}, \tag{7.1}$$

$$\text{3Dimage}(x, y) = \begin{pmatrix} X/W \\ Y/W \\ Z/W \end{pmatrix}. \tag{7.2}$$

There is an OpenCV function `cvReprojectImageTo3D()` for this computation which outputs a 3D point cloud matrix. A point cloud is a set of un-gridded 3-D points corresponding to a 2-D image. It is still under active research to reconstruct smooth 3D surface from these point clouds.

7.7 Discussion

In this chapter, we have studied the 3D stereo image acquisition with our proposed multispectral-stereo system. Our intention here is to investigate how the introduction of the filter affects the stereo system qualitatively by practical observation of the stereo images, and quantitatively by means of disparity maps compared with the reference maps from the pure stereo system.

Our observation of the views of the 3D images acquired with the multispectral-stereo system shows clear 3D images. However there is slight reduction in the brightness, due to the introduction of the filter. Filter reduces the intensity of the light entering into the camera, thus causing the reduced brightness. However, it doesn't affect much in the 3D view.

We have used the most commonly used method: Block Matching for computing the disparity maps. Existing state of the art stereo matching methods are not yet capable enough to have a very good disparity maps as the problems of occlusion, discontinuity, regularity and repetitivity as discussed in Section 3.6.3 are yet to be solved completely. However, block matching method serves well for our purpose of comparing disparity maps.

From the resulting disparity maps in our experiment and comparisons, we see that the precision is about 65% on the average. This means that the 65% disparity points are at the same positions as in the disparity map from the pure stereo camera system. Recall value is about 53% on the average. This means, only 53% of the disparity points that are in the disparity maps from the pure stereo images are available in the disparity maps from the multispectral-streo systems. This lower value of recall is obviously due to the reduction in the intensity of the light because of the introduction of the filter. This can be clearly seen in the disparity images that the lines formed on the edges are thinner comparatively. However, the points are enough to identify the shape of the objects and in turn enough to give the clear 3D images.

For accurate results, we need to make sure that the focus of the cameras are not changed after calibration or during calibration, since the focus is an important factor in calibration. Focusing will change the distance between the center of the lens and the projection plane (the sensor). Although this variation might seem relatively small, it may be enough to void the calibration. Auto-focusing would introduce and extra unknown and unstable factor into the system. However, this could not be maintained in our experiment as the camera doesn't support manual focus. So, this might have introduced some errors in the experiment.

8 Conclusion and Future Perspectives

8.1 Conclusion

This thesis has attempted to conceive a fast and practical multispectral system with capability of 3D imaging at the same time and hence named the system as **Multispectral-Stereo System**. The proposed multispectral-stereo system has been investigated thoroughly with simulation as well as experimental approaches. Both multispectral as well as stereo aspects of the system have been studied.

Simulations with different pairs of cameras and a large number of filters have shown promising results. Selecting a best pair from among the available filters for one each for the pair of cameras lead to a six channel multispectral system which give rise to better estimation of spectral reflectances of the target. We have investigated not only with pair of similar cameras as normally the case with stereo cameras, but also with two different types of cameras. The two cameras need to be operated in the same resolution. To make the simulation more realistic, simulated noise (random shot noise and quantization noise) has been introduced in the camera outputs. The optimal exhaustive filter selection process has been enhanced with additional constraints so as to reduce the processing time. Several well known spectral reflectance estimation methods: Maloney and Wandell, Imai and Berns, least square Wiener, neural network and polynomial, have been investigated. Among them, the polynomial and neural network methods do not require the sensitivities of the cameras to be known, so they are used in experimental studies where we normally have such situation. Performances of all these methods are found to be more or less similar, however polynomial and neural network methods provides possibility of optimization for better results.

With such encouraging results from simulation, we investigated our proposed system further with an experimental approach. Among the two options: creating a stereo camera by joining two cameras or using a readymade stereo camera, we have used, arguably the world's first digital 3D stereo camera recently produced from Fujifilm, the Fujifilm FinePix REAL 3D W1. Rather than going through unnecessary hassle for already proven technique of creating a stereo camera from two cameras, we decided to use the Fujifilm 3D camera, also to explore the new technology behind it. Due to the timely unavailability of the optimal filters as indicated by simulation, we carried out the experiment with a single multiband filter from Schott (Schott BG-36) used in front of the right lens of the camera and the left camera left as it is. We assume that if the proposed 6-channel multispectral-stereo system performs better with this setup than classic 3-channel system, it clearly infers that the result would be much better with the optimal filter combination. And to support our proposition, the experimental results have indicated significantly better performance with our proposed system.

On the stereo side, our multispectral-stereo system outputs 3D stereo images in addition to multispectral information. Having pair of stereo images, 3D images were clearly viewable with any stereo viewer. However, brightness was slightly reduced due to the presence of the filter. Further stereo analysis has been done with the disparity maps obtained from our multi-spectral stereo system. The results from the most widely used block matching algorithm have shown that the information in the disparity maps with the system was reduced compared to the maps with pure stereo system without the filter. However, the information available is quite enough to have 3D image.

It is worth noting that we have simply placed the filter in front of the lens from outside. There are precision and optical issues that might have played role in the accuracy of the results. Moreover, the Fujifilm3D camera does automatic (predefined) white balancing, no manual focusing, and provides no raw image format. So, we had limited control on it. Despite these, still getting encouraging results shows that with more controllable system, the system can be modeled more accurately and we can expect even better result in favor of the proposed system.

Summing up from the whole thesis work, we can conclude that our proposed multi-spectral system could be a viable two-in-one system acting as both multispectral and stereo imaging system. By choosing a best pair of filters, it can be realized easily. And after all, it is fast, one shot multispectral as well as 3D imaging, and practical. We, therefore, believe that this thesis work will be a milestone towards the development of commercial cheap, fast and practical multispectral systems for general, and widespread use.

Finally, we hope that this thesis work will have great value in both the multispectral as well as the 3D imaging world.

8.2 Future Perspectives

Due to time and resource limitations, we couldn't do some of the works we wished to. Furthermore, there are possibilities of extending this thesis work with many interesting further works. All these works can be considered as future perspectives to add to this thesis. Some of these possible future works have been listed here.

- We have performed the experiment on multispectral acquisition with the proposed multispectral-stereo system, using a single multiband Schott BG-36 filter. Further experiment can be done with several filter combinations, especially with those selected by simulation and make comparative analysis.

- We managed to obtain filter data from Omega only. Simulation work can be extended further with more filters from other manufacturers.

- The readymade Fujifilm 3D stereo camera has been used as a primary acquisition device. It has got limitations: no manual focus, no RAW data and no way to turn white balance off. Further study can be done with more controllable cameras.

- Experiment with manually constructed stereo rig from quite different types of cameras can be done and analyze results of multispectral and stereo systems.

- An interesting further work could be to study the accuracy of depth estimation from the stereo system with and without the filters using ground truth 3D objects. We have used the most common Block Matching algorithm for stereo matching. Different stereo matching algorithms can also be investigated.

- The proposed multispectral-stereo system could be tested in real applications like: acquisition of 3D artworks and resulting images with virtual illuminants.

- Yet another interesting further research work would be to relate the disparity/depth estimation with the physical property of the surface, the reflectance data obtained from the multispectral system.

A Camera Sensitivities

Normalized sensitivity of the left and right camera of the Fujifilm 3D as measured during characterization using monochromator is given in the following table. Figure 31(b) shows the plots of these sensitivities.

Table 13: Normalized sensitivity of the Fujifilm 3D camera

	Left Camera			**Right Camera**		
λ	**R**	**G**	**B**	**R**	**G**	**B**
380	0.0694	0.0140	0.1004	0.0357	0.0096	0.0171
385	0.0597	0.0167	0.0905	0.0371	0.0109	0.0165
390	0.0592	0.0147	0.0878	0.0376	0.0119	0.0179
395	0.0742	0.0161	0.1161	0.0421	0.0112	0.0207
400	0.0741	0.0157	0.1216	0.0390	0.0104	0.0197
405	0.0783	0.0177	0.1427	0.0456	0.0118	0.0286
410	0.0705	0.0172	0.1415	0.0503	0.0155	0.0366
415	0.0874	0.0235	0.1677	0.0444	0.0125	0.0425
420	0.0973	0.0189	0.2042	0.0565	0.0149	0.0935
425	0.0859	0.0220	0.2181	0.0554	0.0159	0.0975
430	0.1040	0.0205	0.2678	0.0640	0.0177	0.1352
435	0.1078	0.0216	0.3060	0.0699	0.0196	0.1499
440	0.1153	0.0264	0.3349	0.0698	0.0216	0.1549
445	0.1251	0.0316	0.3549	0.0869	0.0249	0.1843
450	0.1237	0.0329	0.3588	0.0896	0.0267	0.2119
455	0.1359	0.0433	0.3877	0.0909	0.0282	0.2424
460	0.1381	0.0533	0.4200	0.0986	0.0352	0.2871
465	0.1429	0.0807	0.4670	0.0861	0.0632	0.3290
470	0.1262	0.1126	0.4750	0.0875	0.0947	0.3589
475	0.1339	0.1347	0.4945	0.0913	0.1219	0.3540
480	0.1159	0.1654	0.4668	0.0763	0.1447	0.3457
485	0.1142	0.1809	0.4634	0.0713	0.1717	0.3331
490	0.0825	0.2174	0.4655	0.0531	0.2088	0.3020
495	0.0542	0.2543	0.4058	0.0324	0.2451	0.3043
500	0.0246	0.3116	0.3333	0.0178	0.2922	0.2430
505	0.0275	0.3583	0.2690	0.0095	0.3489	0.1549
510	0.0120	0.3981	0.1371	0.0058	0.4300	0.0493
515	0.0106	0.4832	0.0704	0.0050	0.4727	0.0158
520	0.0039	0.5689	0.0274	0.0017	0.5409	0.0039
525	0.0014	0.5885	0.0134	0.0018	0.5627	0.0015
530	0.0015	0.5955	0.0104	0.0017	0.5668	0.0010
535	0.0017	0.6030	0.0107	0.0016	0.5629	0.0009

Table 13 contd..

λ	Left Camera			Right Camera		
	R	**G**	**B**	**R**	**G**	**B**
540	0.0016	0.6030	0.0074	0.0014	0.5558	0.0007
545	0.0014	0.5974	0.0075	0.0014	0.5525	0.0007
550	0.0009	0.5832	0.0063	0.0013	0.5535	0.0006
555	0.0007	0.5711	0.0049	0.0013	0.5555	0.0004
560	0.0008	0.5749	0.0049	0.0013	0.5565	0.0003
565	0.0009	0.5757	0.0051	0.0009	0.5427	0.0003
570	0.0047	0.5668	0.0048	0.0036	0.5361	0.0005
575	0.2134	0.4930	0.0097	0.1977	0.4661	0.0009
580	0.6279	0.4141	0.0085	0.5749	0.4095	0.0015
585	0.8060	0.3804	0.0141	0.8395	0.3602	0.0024
590	0.9286	0.3286	0.0210	0.9502	0.2622	0.0039
595	0.9554	0.2427	0.0437	1.0000	0.1986	0.0083
600	0.9676	0.1861	0.0564	0.9924	0.1594	0.0081
605	1.0000	0.1112	0.0701	0.9563	0.1077	0.0100
610	0.9721	0.0758	0.0672	0.9445	0.0783	0.0100
615	0.9744	0.0480	0.0927	0.9416	0.0534	0.0167
620	0.9135	0.0397	0.0961	0.9300	0.0449	0.0204
625	0.8705	0.0288	0.0994	0.9290	0.0356	0.0220
630	0.8636	0.0247	0.1014	0.9166	0.0277	0.0232
635	0.8619	0.0205	0.1158	0.8881	0.0154	0.0204
640	0.7813	0.0256	0.0957	0.7336	0.0173	0.0172
645	0.6439	0.0168	0.0843	0.5517	0.0200	0.0194
650	0.4074	0.0321	0.0925	0.3190	0.0239	0.0299
655	0.2972	0.0327	0.0924	0.2147	0.0196	0.0332
660	0.2189	0.0190	0.1059	0.1463	0.0114	0.0279
665	0.1298	0.0163	0.0955	0.0619	0.0123	0.0225
670	0.1066	0.0152	0.1045	0.0544	0.0130	0.0250
675	0.0952	0.0163	0.0936	0.0430	0.0113	0.0200
680	0.0916	0.0169	0.0985	0.0491	0.0138	0.0238
685	0.0931	0.0157	0.1049	0.0481	0.0126	0.0240
690	0.0734	0.0182	0.0934	0.0427	0.0141	0.0219
695	0.0781	0.0160	0.0964	0.0411	0.0119	0.0196
700	0.0658	0.0176	0.0935	0.0401	0.0129	0.0199
705	0.0779	0.0186	0.0970	0.0389	0.0120	0.0181
710	0.0792	0.0154	0.0899	0.0426	0.0112	0.0168
715	0.0732	0.0148	0.0869	0.0456	0.0124	0.0190
720	0.0831	0.0195	0.0998	0.0448	0.0128	0.0198

B Filters Used

The table below lists all the two hundred and sixty five filters of 3 categories: exciter, dichroic and emitter from OmegaOmega (2010) used in the simulation of 6-channel multispectral system in Chapter 5.

Table 14: List of all 265 filters of 3 categories from Omega, used in the simulation

Exciter			**Dichroic**		**Emitter**		
XF1000	XF1046	XF1093	XF2000	XF2036	XF3000	XF3056	XF3094
XF1001	XF1048	XF1094	XF2001	XF2037	XF3001	XF3057	XF3095
XF1004	XF1049	XF1095	XF2002	XF2038	XF3002	XF3058	XF3096
XF1005	XF1050	XF1096	XF2003	XF2039	XF3003	XF3059	XF3097
XF1006	XF1051	XF1097	XF2004	XF2040	XF3004	XF3060	XF3099
XF1007	XF1052	XF1098	XF2005	XF2041	XF3005	XF3061	XF3100
XF1008	XF1053	XF1100	XF2006	XF2042	XF3006	XF3063	XF3104
XF1009	XF1054	XF1101	XF2007	XF2043	XF3007	XF3064	XF3105
XF1010	XF1055	XF1103	XF2008	XF2044	XF3008	XF3065	XF3106
XF1011	XF1057	XF1104	XF2009	XF2045	XF3009	XF3066	XF3108
XF1012	XF1058	XF1111	XF2010	XF2046	XF3011	XF3067	XF3113
XF1013	XF1059	XF1201	XF2011	XF2047	XF3012	XF3068	XF3114
XF1014	XF1060	XF1202	XF2012	XF2048	XF3014	XF3069	XF3115
XF1015	XF1062	XF1203	XF2013	XF2050	XF3015	XF3070	XF3116
XF1016	XF1063	XF1204	XF2014	XF2051	XF3016	XF3071	XF3118
XF1017	XF1064	XF1206	XF2015	XF2053	XF3017	XF3073	XF3119
XF1018	XF1065	XF1207	XF2016	XF2054	XF3018	XF3074	XF3120
XF1019	XF1067	XF1208	XF2017	XF2055	XF3019	XF3075	XF3301
XF1020	XF1068	XF1210	XF2018	XF2056	XF3020	XF3076	XF3302
XF1021	XF1069	XF1301	XF2019	XF2057	XF3021	XF3077	XF3303
XF1022	XF1070	XF1402	XF2020	XF2058	XF3022	XF3078	XF3304
XF1023	XF1071	XF1403	XF2021	XF2059	XF3023	XF3079	XF3305
XF1024	XF1072		XF2022	XF2060	XF3024	XF3080	XF3307
XF1025	XF1073		XF2023	XF2061	XF3025	XF3081	XF3308
XF1026	XF1074		XF2024	XF2062	XF3027	XF3082	XF3309
XF1027	XF1075		XF2025	XF2065	XF3028	XF3083	XF3401
XF1028	XF1076		XF2026	XF2072	XF3029	XF3084	XF3403
XF1029	XF1077		XF2027	XF2075	XF3030	XF3085	
XF1030	XF1078		XF2028	XF2077	XF3031	XF3086	
XF1031	XF1079		XF2029	XF2082	XF3034	XF3087	
XF1032	XF1080		XF2030	XF2083	XF3035	XF3088	
XF1038	XF1082		XF2031	XF2084	XF3042	XF3089	
XF1042	XF1085		XF2032	XF2090	XF3043	XF3090	
XF1043	XF1087		XF2033	XF2091	XF3049	XF3091	
XF1044	XF1090		XF2034	XF2203	XF3054	XF3092	
XF1045	XF1091		XF2035	XF2589	XF3055	XF3093	

C Calibration Output Files

Output of the calibration with OpenCV functions are saved into YML files. YML is a human-readable data serialization format. The results of the calibration of the left and right cameras are saved in **intrinsics-left.yml** file and **intrinsics-right.yml** file respectively. The intrinsic and extrinsic parameters from the stereo calibration are saved in **intrinsics.yml** and **extrinsics.yml** files. The contents of these files are given below.

C.1 Individual Camera Calibration Output Files

1. **intrinsics-left.yml**

```
%YAML:1.0
calibration_time: "05/23/10 18:55:27"
image_count: 20
image_width: 3648
image_height: 2736
board_width: 8
board_height: 6
square_size: 30.
flags: 0
camera_matrix: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 4.2573583268527309e+003, 0., 1.8773356796229127e+003, 0.,
       4.3020359956892053e+003, 1.3024059318657828e+003, 0., 0., 1. ]
distortion_coefficients: !!opencv-matrix
   rows: 1
   cols: 4
   dt: d
   data: [ -1.4861250383011643e-001, 3.2637673720935850e-001,
            1.7354102578749399e-003, 3.1359656284123986e-003 ]
avg_reprojection_error: 7.3037707010904951e-001
per_view_reprojection_errors: !!opencv-matrix
   rows: 1
   cols: 20
   dt: d
   data: [ 5.5138905843098962e-001, 1.0744539896647136e+000,
           5.6956736246744788e-001, 3.5890324910481769e-001,
           9.6725718180338538e-001, 1.1059595743815105e+000,
           1.1592470804850261e+000, 9.0493901570638025e-001,
           7.4378840128580725e-001, 6.7945988972981775e-001,
           4.6361414591471356e-001, 6.5709940592447913e-001,
           9.6058781941731775e-001, 3.9113362630208331e-001,
           7.1965662638346350e-001, 6.7188517252604163e-001,
```

```
            8.5716374715169275e-001, 6.7819849650065100e-001,
            4.4362258911132813e-001, 6.4961496988932288e-001 ]
```

2. **intrinsics-right.yml**

```
%YAML:1.0
calibration_time: "05/23/10 19:10:54"
image_count: 20
image_width: 3648
image_height: 2736
board_width: 8
board_height: 6
square_size: 30.
flags: 0
camera_matrix: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 4.2688389650466370e+003, 0., 1.9166917064515162e+003, 0.,
           4.3192877923639599e+003, 1.2814083430457515e+003, 0., 0., 1. ]
distortion_coefficients: !!opencv-matrix
   rows: 1
   cols: 4
   dt: d
   data: [ -1.4477588865624766e-001, 5.7353781419877270e-001,
           -3.5731452233667663e-003, -7.3779788261235982e-004 ]
avg_reprojection_error: 7.0582936604817703e-001
per_view_reprojection_errors: !!opencv-matrix
   rows: 1
   cols: 20
   dt: d
   data: [ 4.3363316853841144e-001, 9.3661626180013025e-001,
           6.6733296712239587e-001, 4.2369333902994794e-001,
           8.8064193725585938e-001, 9.6989440917968750e-001,
           1.0675404866536458e+000, 7.7111434936523438e-001,
           8.5469182332356775e-001, 6.1827723185221350e-001,
           4.8569742838541669e-001, 5.9260559082031250e-001,
           8.5469945271809900e-001, 4.3922932942708331e-001,
           6.1047490437825525e-001, 7.7059046427408850e-001,
           7.7830123901367188e-001, 6.6384760538736975e-001,
           5.1025772094726563e-001, 7.8744761149088538e-001 ]
```

C.2 Stereo Calibration Output Files

1. **intrinsics.yml**

```
%YAML:1.0
M1: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 4.2532475204278753e+003, 0., 1.8653319793692649e+003, 0.,
       4.3013127964913956e+003, 1.3023827220978151e+003, 0., 0., 1. ]
D1: !!opencv-matrix
   rows: 1
   cols: 4
   dt: d
   data: [ -2.2896895407389070e-001, 1.5049579552551959e+000,
            1.3403388538534013e-003, 5.0116628693841403e-003 ]
M2: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 4.2684504066644622e+003, 0., 1.9237113469666192e+003, 0.,
           4.3187078694229967e+003, 1.2885073551037951e+003, 0., 0., 1. ]
D2: !!opencv-matrix
   rows: 1
   cols: 4
   dt: d
   data: [ -1.6699878382873154e-001, 1.7223394604688445e+000,
           -3.2161370832869381e-003, -4.1527684659167750e-005 ]
```

2. **extrinsics.yml**

```
%YAML:1.0
R: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 9.9962973067876348e-001, 9.0433411807287357e-003,
           2.5663583603848954e-002, -9.1676911537198914e-003,
           9.9994678000579063e-001, 4.7318701335785788e-003,
          -2.5619425872037319e-002, -4.9653938756137649e-003,
           9.9965943694922788e-001 ]
T: !!opencv-matrix
   rows: 3
   cols: 1
   dt: d
   data: [ -8.9548149042163161e+001, -1.2500163615581947e+000,
            2.0084957471571312e+000 ]
```

```
R1: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 9.9972757320996342e-001, 2.3105060328523813e-002,
           3.3068944544005583e-003, -2.3112552854938736e-002,
           9.9973034693817964e-001, 2.2457318171549261e-003,
          -3.2541149710687608e-003, -2.3215507925077577e-003,
           9.9999201053692066e-001 ]
R2: !!opencv-matrix
   rows: 3
   cols: 3
   dt: d
   data: [ 9.9965121846586136e-001, 1.3954284843405707e-002,
          -2.2421403930793289e-002, -1.4005458843351380e-002,
           9.9989965655256630e-001, -2.1269576034944311e-003,
           2.2389473917577566e-002, 2.4402378099212937e-003,
           9.9974634617843194e-001 ]
P1: !!opencv-matrix
   rows: 3
   cols: 4
   dt: d
   data: [ 4.8599018656139697e+003, 0., 1.9252109375000000e+003, 0., 0.,
           4.8599018656139697e+003, 1.3065872650146484e+003, 0., 0., 0., 1.,0. ]
P2: !!opencv-matrix
   rows: 3
   cols: 4
   dt: d
   data: [ 4.8599018656139697e+003, 0., 1.9252109375000000e+003,
          -4.3525227754524612e+005, 0., 4.8599018656139697e+003,
           1.3065872650146484e+003, 0., 0., 0., 1., 0. ]
Q: !!opencv-matrix
   rows: 4
   cols: 4
   dt: d
   data: [ 1., 0., 0., -1.9252109375000000e+003, 0., 1., 0.,
          -1.3065872650146484e+003, 0., 0., 0., 4.8599018656139697e+003, 0.,
           0., -1.1165712659846483e-002, 0. ]
```

D OpenCV Functions

Stereo analysis has been carried out with programs developed in Visual C++ 2008 using the OpenCV version 2.1 library. The OpenCV functions used are listed here along with brief descriptions. Detail explanation of these functions can be found on the web (OpenCV 2010).

1. `cvFindChessboardCorners()`

   The function attempts to determine whether the input image is a view of the chessboard pattern and locate the internal chessboard corners (i.e., points, where the black squares touch each other). The function returns a non-zero value if all of the corners have been found, and they have been placed in a certain order (row by row, left to right in every row), otherwise, if the function fails to find all the corners or reorder them, it returns 0. The coordinates detected are approximate, and to determine their position more accurately, the user may use the function `cvFindCornerSubPix()`.

2. `cvFindCornerSubPix()`

   This function is used after `cvFindChessboardCorners()` for more accurately locating corners. The function iterates to find the sub-pixel accurate location of corners, or radial saddle points.

3. `cvDrawChessboardCorners()`

   It renders the detected chessboard corners. The function draws the individual chessboard corners detected as red circles if the board was not found or as colored corners connected with lines if the board was found.

4. `cvCalibrateCamera2()`

   The function estimates the intrinsic camera parameters and extrinsic parameters for each of the views. The coordinates of 3D object points and their correspondent 2D projections in each view must be specified. That may be achieved by using an object with known geometry and easily detectable feature points. Chessboard images are used as calibration objects and the corners extracted from `cvFindChessboardCorners()` are used as 3D object points.

5. `cvStereoCalibrate()`

   The function estimates transformation between the two cameras making a stereo pair. Besides the stereo-related information, the function can also perform full calibration of each of the two cameras. However, because of the high dimensionality of the parameter space and noise in the input data, the function can diverge from the correct solution. Thus, if intrinsic parameters can be estimated with high accuracy for each of the cameras individually (e.g. using `cvCalibrateCamera2()`), it is recommended to do so and then pass `CV_CALIB_FIX_INTRINSIC` flag to the function along with the computed intrinsic parameters. Otherwise, if all the parameters are estimated at once, it makes sense to restrict some parameters, e.g. pass `CV_CALIB_SAME_FOCAL_LENGTH` and `CV_CALIB_ZERO_TANGENT_DIST` flags, which are usually reasonable assumptions.

6. `cvStereoRectify()`

   The function computes the rotation matrices for each camera that (virtually) make both camera image planes the same plane. Consequently, that makes all the epipolar lines parallel and thus simplifies the dense stereo correspondence problem. On input the function takes the matrices computed by `cvStereoCalibrate()` and on output it gives two rotation matrices and also two projection matrices in the new coordinates.

7. `cvReprojectImageTo3D()`

   It reprojects disparity image to 3D space by transforming 1-channel disparity map to 3-channel image representing a 3D surface.

Bibliography

Ajito, T., Obi, T., Yamaguchi, M. and Ohyama, N. (2000), Expanded color gamut reproduced by six-primary projection display, *in* 'Projection Displays 2000: Sixth in a Series', Vol. 3954 of *SPIE Proceedings*, pp. 130–137.

Alexander, B. F. and Ng, K. C. (1987), 3D shape measurement by active triangulation using an array of coded light stripes, *in* 'Optics, Illumination, and Image Sensing for Machine Vision II', Vol. 850, pp. 199–209.

Baker, H. H. (1981), Depth from edge and intensity based stereo, PhD thesis, University of Illinois, Champaign, IL, USA.

Barnard, K. and Funt, B. (2002), "Camera characterization for color research", *Color Research & Application* , Vol. 27, WILEY, pp. 152–163.

Barnard, S. and Thompson, W. (1980), "Disparity analysis of images", *PAMI* , Vol. 2(4), pp. 333–340.

Barron, J., Fleet, D. J. and Beauchemin, S. (1994), "Performance of optical flow techniques", *International Journal of Computer Vision* , Vol. 12(1), Springer Netherlands, pp. 43–77.

Batlle, J., Mouaddib, E. and Salvi, J. (1998), "Recent progress in coded structured light as a technique to solve the correspondence problem - image segmentation techniques", *Pattern Recognition* , Vol. 31, Elsevier, pp. 963–982.

Bay, H., Tuytelaars, T. and Gool, L. V. (2006), SURF: Speeded up robust features, *in* 'Computer Vision ECCV 2006', Vol. 3951 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 404–417.

Berns, R. S. (1998), Challenges for color science in multimedia imaging, *in* 'CIM'98: Colour Imaging in Multimedia', Derby, UK, pp. 123–133.

Berns, R. S., Imai, F. H., Burns, P. D. and Tzeng, D. Y. (1998), Multispectral-based color reproduction research at the munsell color science laboratory, *in* 'Electronic Imaging: Processing, Printing, and Publishing in Color', Vol. 3409 of *SPIE Proceedings*, pp. 14–25.

Berns, R. S., Taplin, L. A. and Liang, T. Z. (2003), 'Spectral color reproduction with six color output'. US Patent 2003/0098896 A1.

Bianco, S., Gasparini, F., Schettini, R. and Vanneschi, L. (2008), "Polynomial modeling and optimization for colorimetric characterization of scanners", *Journal of Electronic Imaging* , Vol. 17(04), SPIE.

Blasco, J., Aleixos, N., Gómez, J. and Moltó, E. (2007), "Citrus sorting by identification of the most common defects using multispectral computer vision", *Journal of Food Engineering* , Vol. 83(3), pp. 384–393.

Bouchard, M. B., Chen, B. R., Burgess, S. A. and Hillman, E. M. C. (2009), "Ultra-fast multispectral optical imaging of cortical oxygenation, blood flow, and intracellular calcium dynamics", *Opt. Express* , Vol. 17(18), OSA, pp. 15670–15678.

Bouguet, J. Y. (2006), 'Camera calibration toolbox for Matlab', Online, `http://www.vision.caltech.edu/bouguetj/calib_doc/`. Last Visit: May, 2010.

Brauers, J., Helling, S. and Aach, T. (2009), "Multispectral image acquisition with flash light sources", *Journal of Imaging Science and Technology* , Vol. 53(3), IS&T, p. 031103.

Brown, D. (1966), "Decentering distortion of lenses", *Photogrammetric Eng.* , Vol. 32(3), pp. 444–462.

Burger, R. E. and Sherman, D. (1994), Producing colorimetric data from film scanners using a spectral characterization target, *in* 'Device-Independent Color Imaging', Vol. 2170 of *SPIE Proceedings*, pp. 42–52.

Burns, P. D. (1997), Analysis of image noise in multispectral color acquisition, PhD thesis, Center for Imaging Science, Rochester Institute of Technology.

Burns, P. D. and Berns, R. S. (1996), Analysis of multispectral image capture, *in* 'Proceedings of the IS&T/SID Fourth Color Imaging Conference: Color Science, Systems, and Applications, Color Imaging Conference', IS&T/SID, Scottsdale, AZ, United States, pp. 19–22.

Carstensen, J. M. (2007), 'Multispectral imaging offers new tools', Online, `http://www.optoiq.com/index/machine-vision-imaging-processing/display/vsd-article-display/286653/articles/vision-systems-design/volume-12/issue-3/features/multispectral-imaging-offers-new-tools.html`. Last Visit: June, 2010.

Chatzis, I. S., Kappatos, V. A. and Dermatas, E. S. (2006), Filter selection for multi-spectral image acquisition using the feature vector analysis methods, *in* 'IPROM2006'.

Cheung, V., Westland, S., Li, C., Hardeberg, J. and Connah, D. (2005), "Characterization of trichromatic color cameras by using a new multispectral imaging technique", *J. Opt. Soc. Am. A* , Vol. 22(7), pp. 1231–1240.

Cohen, J. (1964), "Dependency of the spectral reflectance curves of the munsell color chips", *Psychonomic Sci.* , Vol. 1, pp. 369–370.

Connah, D., Alsam, A. and Hardeberg, J. Y. (2006), "Multispectral imaging: How many sensors do we need?", *Journal of Imaging Science and Technology* , Vol. 50(1), IS&T, pp. 45–52.

Connah, D. R. and Hardeberg, J. Y. (2005), Spectral recovery using polynomial models, *in* 'Color Imaging X: Processing, Hardcopy, and Applications', Vol. 5667 of *SPIE Proceedings*, pp. 65–75.

Cotte, P. and Dupouy, M. (2003), CRISATEL high resolution multispectral system., *in* 'PICS', IS&T, pp. 161–165.

Cotte, P. and Dupraz, D. (2007), 'Multispectral scanner with enlarged gamut, in particular a single-pass flat-bed scanner'. US Patent 20070223058A1.

Dannemiller, J. L. (1992), "Spectral reflectance of natural objects: how many basis functions are necessary?", *J. Opt. Soc. Am. A* , Vol. 9(4), OSA, pp. 507–515.

Day, D. C. (2003*a*), Filter selection for spectral estimation using a trichromatic camera, Master Thesis, Rochester Institute of Technology, Center for Imaging Science, Rochester, New York, United States.

Day, E. A. (2003*b*), The effects of multi-channel visible spectrum imaging on perceived spatial image quality and color reproduction accuracy, M.S. Thesis, Rochester Institute of Technology, Center for Imaging Science, Rochester, New York, United States.

Day, E. A., Berns, R. S., Taplin, L. A. and Imai, F. H. (2003), A psychophysical experiment evaluating the color accuracy of several multispectral image capture techniques, *in* 'Proceedings of IS&T PICS Conference, PICS', IS&T, Springfield, Virginia, United States, pp. 199–204.

Deriche, Zhang, Deriche, R., Zhang, Z., Luong, Q. T. and Faugeras, O. (1994), Robust recovery of the epipolar geometry for an uncalibrated stereo rig, *in* 'Computer Vision – ECCV '94', Vol. 800 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 567–576.

Doi, M., Ohtsuki, R. and Tominaga, S. (2005), Spectral estimation of skin color with foundation makeup, *in* 'Image Analysis', Vol. 3540 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 95–104.

Dyas, B. (2000), Robust sensor response characterization, *in* 'The IS&T/SID Eighth Color Imaging Conference', pp. 144–148.

Ellrod, G. P., Connell, B. H. and Hillger, D. W. (2003), "Improved detection of airborne volcanic ash using multispectral infrared satellite data", *J. Geophys. Res., 108(D12), 4356* , Vol. 108 (D12), pp. 4356–4369.

Everdell, N. L., Styles, I. B., Claridge, E., Hebden, J. C. and Calcagni, A. S. (2009), Multispectral imaging of the ocular fundus using LED illumination, *in* 'Novel Optical Instrumentation for Biomedical Applications IV', Vol. 7371 of *SPIE Proceedings*.

Farina, B., Bartoli, C., Bono, A., Colombo, A., Lualdi, M., Tragni, G., Marchesini, R., Farina, B., Bartoli, C., Bono, A., Colombo, A., Lualdi, M., and Tragni, G. (2000), "Multispectral imaging approach in the diagnosis of cutaneous melanoma: potentiality and limits", *Physics in Medicine and Biology* , Vol. 45(5), pp. 1243–1254.

Farkas, D. L., Ballou, B. T., Fisher, G. W., Fishman, D., Garini, Y., Niu, W. and Wachman (1996), Microscopic and mesoscopic spectral bio-imaging, *in* 'Optical Diagnostics of Living Cells and Biofluids', Vol. 2678 of *SPIE Proceedings*.

Farrell, J. E. and Wandell, B. A. (1993), "Scanner linearity", *Journal of Electronic Imaging* , Vol. 3, SPIE, pp. 147–161.

Faugeras, O. (1993), *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, Cambridge, MA, USA.

Finlayson, G. D. (1997), "Constrained least-squares regression in color spaces", *Journal of Electronic Imaging* , Vol. 6(4), SPIE, pp. 484–493.

Fisher, R. B. (2007), 'CVonline: The evolving, distributed, non-proprietary, on-line compendium of computer vision', Online, `http://homepages.inf.ed.ac.uk/rbf/CVonline/`. Last Visit: June, 2010.

Fitzgibbon, A. W. and Zisserman, A. (1998), Automatic 3D model acquisition and generation of new images from video sequences, *in* 'Proceedings of European Signal Processing Conference (EUSIPCO '98), Rhodes, Greece', pp. 1261–1269.

Fleet, D. J. and Jepson, A. D. (1993), "Stability of phase information", *IEEE Trans. Pattern Anal. Mach. Intell.* , Vol. 15(12), IEEE Computer Society, Washington, DC, USA, pp. 1253–1268.

Frisby, J. P. (1998), "Stereo correspondence and neural networks", *The handbook of brain theory and neural networks* , MIT Press, Cambridge, MA, USA, pp. 937–941.

Fryer, J. and Brown, D. (1986), "Lens distortion for close-range photogrammetry", *Photogrammetric Eng.* , Vol. 52(1), pp. 51–58.

Fujifilm (2010*a*), 'Fujifilm FinePix REAL 3D V1 viewer', Online, `http://www.fujifilm.com/products/3d/viewer/finepix_real3dv1/`. Last Visit: June, 2010.

Fujifilm (2010*b*), 'Fujifilm FinePix REAL 3D W1 camera', Online, `http://www.fujifilm.com/products/3d/camera/finepix_real3dw1/`. Last Visit: June, 2010.

Garcia, F., Araiza, R. and Rzycki, B. (2000), Towards optimal mosaicking of multi-spectral images, *in* 'Proceedings of the 2nd Nat'l NASA Student Conference', pp. 261–266.

Goulermas, J. Y. and Liatsis, P. (2000), "A new parallel feature-based stereo-matching algorithm with figural continuity preservation, based on hybrid symbiotic genetic algorithms", *Pattern Recognition* , Vol. 33(3), pp. 529–531.

Griesser, A., Koninckx, T. P. and Gool, L. V. (2004), Adaptive real-time 3D acquisition and contour tracking within a multiple structured light system, *in* D. Azada, ed., '12th Pacific Conference on Computer Graphics and Applications (PG2004)', IEEE Computer Society, pp. 361–370.

Grimson, E. L. W. (1981), *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, MA, USA.

Gulliksson, M. and Wedin, P.-A. (2000), "The use and properties of Tikhonov filter matrices", *SIAM Journal on Matrix Analysis and Applications* , Vol. 22(1), SIAM, pp. 276–281.

Haneishi, H., Hasegawa, T., Hosoi, A., Yokoyama, Y., Tsumura, N. and Miyake, Y. (2000), "System design for accurately estimating the spectral reflectance of art paintings", *Appl. Opt.* , Vol. 39(35), OSA, pp. 6621–6632.

Haneishi, H., Hasegawa, T., Tsumura, N. and Miyake, Y. (1997), Design of color filters for recording art works, *in* 'IS&T 50th Annual Conference', pp. 369–372.

Hannah, M. (1985), SRI's baseline stereo system, *in* 'DARPA85', pp. 149–155.

Hardeberg, J. Y. (1999), Acquisition and Reproduction of Colour Images: Colorimetric and Multispectral Approaches, Doctoral dissertation, École Nationale Supérieure des Télécommunications de Paris.

Hardeberg, J. Y. (2004), "Filter selection for multispectral color image acquisition", *Journal of Imaging Science and Technology* , Vol. 48(2), IS&T, pp. 105–110.

Hardeberg, J. Y., Schmitt, F. and Brettel, H. (2002), "Multispectral color image capture using a liquid crystal tunable filter", *Optical Engineering* , Vol. 41(10), SPIE, pp. 2532–2548.

Hashimoto, M. and Kishimoto, J. (2008), Two-shot type 6-band still image capturing system using commercial digital camera and custom color filter, *in* 'IS&T Fourth European Conference on Colour in Graphics', Terrassa, Spain, p. 538.

Hauta-Kasari, M., Miyazawa, K., Toyooka, S. and Parkkinen, J. (1999), "Spectral vision system for measuring color images", *J. Opt. Soc. Am. A* , Vol. 16(10), OSA, pp. 2352–2362.

Healey, G. E. and Kondepudy, R. (1994), "Radiometric CCD camera calibration and noise estimation", *IEEE Trans. Pattern Anal. Mach. Intell.* , Vol. 16(3), pp. 267–276.

Heikkila, J. and Silven, O. (1997), A four-step camera calibration procedure with implicit image correction, *in* 'Conference on Computer Vision and Pattern Recognition (CVPR '97)', IEEE Computer Society, Washington, DC, USA, p. 1106.

Henkel, R. D. (1997), Fast stereovision by coherence detection, *in* 'Computer Analysis of Images and Patterns', Vol. 1296 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 297–304.

Henkel, R. D. (1998), A simple and fast neural network approach to stereovision, *in* 'Proceedings of NIPS'97', MIT Press, pp. 808–814.

Henkel, R. D. (2000), 'Synchronization, coherence-detection and three-dimensional vision'.

Herbig, G. P. (2005), Commentary to the review of the stereo calculators, *in* 'Stereoscopy', number 25 *in* 'Series 2'.

Hill, B. (1998), Multispectral color technology: a way toward high-definition color image scanning and encoding, *in* 'Electronic Imaging: Processing, Printing, and Publishing in Color', Vol. 3409 of *SPIE Proceedings*, pp. 2–13.

Hill, B. and Vorhagen, F. W. (1994), 'Multispectral image pick-up system'. US Patent 5,319,472.

Hirschmuller, H. (2006), Stereo vision in structured environments by consistent semi-global matching, *in* 'Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference', Vol. 2, pp. 2386–2393.

Horaud, R. (1989), "Stereo correspondence through feature grouping and maximal cliques", *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 11, pp. 1168–1180.

Horman, M. H. (1976), "Temperature analysis from multispectral infrared data", *Appl. Opt.* , Vol. 15(9), OSA, pp. 2099–2104.

Huang, H. H. (2004), Acquisition of multispectral images using digital cameras, *in* 'Asian Association on Remote Sensing (ACRS)'.

Hubel, P. M., Sherman, D. and Farrell, J. E. (1994), A comparison of methods of sensor spectral sensitivity estimation, *in* 'IS&T and SID's 2nd Color Imaging Conference: Color Science, Systems', Scottsdale, Arizona, pp. 45–48.

Hung, Y. P., Chen, C. S., Hung, K. C., Chen, Y. S. and Fuh, C. S. (1998), "Multipass hierarchical stereo matching for generation of digital terrain models from aerial images", *Machine Vision and Applications* , Vol. 10(5-6), Springer Berlin / Heidelberg, pp. 280–291.

Imai, F. H. (1998), Multi-spectral image acquisition and spectral reconstruction using a trichromatic digital camera system associated with absorption filters, Technical report, Munsell Color Science Laboratory Technical Report, Rochester, NY.

Imai, F. H. (2000), "A comparative analysis of spectral reflectance estimated in various spaces using a trichromatic camera system", *Journal of Imaging Science and Technology* , Vol. 44, IS&T, pp. 280–287.

Imai, F. H. and Berns, R. S. (1999), Spectral estimation using trichromatic digital cameras, *in* 'International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives', pp. 42–49.

Imai, F. H. and Berns, R. S. (2002), Spectral estimation of artist oil paints using multi-filter trichromatic imaging, *in* '9th Congress of the International Colour Association', Vol. 4421 of *SPIE Proceedings*, pp. 504–507.

Imai, F. H., Rosen, M. R. and Berns, R. S. (2000), Comparison of spectrally narrow-band capture versus wide-band with a priori sample analysis for spectral reflectance estimation, *in* 'Proceedings of Eighth Color Imaging Conference: Color Science and Engineering, Systems, Technologies and Applications, IS&T', pp. 234–241.

Imai, F. H., Rosen, M. R. and Berns, R. S. (2002), Comparative study of metrics for spectral match quality, *in* 'The First European Conference on Colour in Graphics, Imaging and Vision (IS&T, Springfield, Va., 2002)', pp. 492–496.

Imai, F. H., Taplin, L. A. and Day, E. A. (2003), Comparative study of spectral reflectance estimation based on broad-band imaging systems, Technical report, Rochester Institute of Technology, College of Science, Center for Imaging Science,Munsell Color Science Laboratory, Rochester, New York, United States.

ISU (2010), 'Making a start in stereo', International Stereoscopic Union, Online, `http://www.stereoscopy.com/isu/index.html`. Last Visit: June, 2010.

Jaumann, R., Neukum, G., Behnke, T., Duxbury, T., Eichentopf, K., Flohrer, J., Gasselt, S., Giese, B., Gwinner, K., Hauber, E., Hoffmann, H., Hoffmeister, A., Köhler, U., Matz, K., McCord, T., Mertens, V., Oberst, J., Pischel, R., Reiss, D., Ress, E., Roatsch, T., Saiger, P., Scholten, F., Schwarz, G., Stephan, K. and Wählisch, M. (2007), "The high-resolution stereo camera (HRSC) experiment on mars express: Instrument aspects and experiment conduct from interplanetary cruise through the nominal mission", *Planetary and Space Science* , Vol. 55(7-8), Elsevier, pp. 928–952.

Jenkin, M. R. M. and Jepson, A. D. (1994), "Recovering local surface structure through local phase difference measurements", *CVGIP: Image Understanding* , Vol. 59(1), Academic Press, Inc., Orlando, FL, USA, pp. 72–93.

Jon, H., Schmitt, F., Brettel, H., Crettez, J. P. and Matre, H. (1998), Multispectral imaging in multimedia, *in* 'Proceedings, CIM'98, Colour Imaging in Multimedia', pp. 75–86.

Kaarna, A. and Parkkinen, J. (2000), "Wavelet filter selection in multispectral image compression", *Pattern Recognition, International Conference on* , Vol. 3, IEEE Computer Society, Los Alamitos, CA, USA, p. 3246.

Keusen, T. (1996), "Multispectral color system with an encoding format compatible with the conventional tristimulus model", *Journal of Imaging Science and Technology* , Vol. 40(6), IS&T, pp. 510–515.

Kleynen, O., Leemans, V. and Destain, M. F. (2005), "Development of a multi-spectral vision system for the detection of defects on apples", *Journal of Food Engineering* , Vol. 69, Elsevier, pp. 41–49.

Koenig, F. and Praefcke, W. (1998), Practice of multispectral image acquisition, *in* 'Electronic Imaging: Processing, Printing, and Publishing in Color', Vol. 3409 of *SPIE Proceedings*, pp. 34–41.

Koo, H.-S. and Jeong, C.-S. (2001), An area-based stereo matching using adaptive search range and window size, *in* 'Computational Science - ICCS 2001', Vol. 2074 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 44–53.

Kuffar, H. and Takaya, K. (2007), Depth measurement and 3D metric reconstruction from two uncalibrated stereo images, *in* 'Canadian Conference on Electrical and Computer Engineering, 2007 (CCECE 2007)', pp. 1460–1463.

Kumar, R. and Hanson, A. R. (1994), "Robust methods for estimating pose and a sensitivity analysis", *CVGIP: Image Underst.* , Vol. 60(3), Academic Press, Inc., Orlando, FL, USA, pp. 313–342.

Kurazume, R. K., Nishino, K., Zhang, Z. and Ikeuchi, K. (2002), Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute, *in* 'Proceedings of Fifth Asian Conference on Computer Vision', pp. 99–106.

Kurland, E. (2009), 'How to build your own 3D camera rig for under $20', Online, `http://www.maximumpc.com/article/howtos/how_build_your_own_3d_camera_rig`. Last Visit: June, 2010.

Levenson, R. M., Cronin, P. J. and Pankratov, K. K. (2003), Spectral imaging for brightfield microscopy, *in* 'Spectral Imaging: Instrumentation, Applications, and Analysis II', Vol. 4959 of *SPIE Proceedings*, pp. 27–33.

Li, G., Genc, Y. and Zucker, S. W. (2007), Multi-view edge-based stereo by incorporating spatial coherence, *in* '3DIM '07: Proceedings of the Sixth International Conference on 3-D Digital Imaging and Modeling', IEEE Computer Society, Washington, DC, USA, pp. 341–348.

Lim, H. S. and Binford, T. O. (1987), Stereo correspondence: A hierarchical approach, *in* 'Image Understanding Workshop', pp. 234–241.

Lowe, D. G. (2004), "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision* , Vol. 60(2), Springer Netherlands, pp. 91–110.

Luo, A., Burkhardt, H. and I, T. I. (1995), 'An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions'.

Maitre, H., Schmitt, F., Crettez, J. P., Wu, Y. and Hardeberg, J. Y. (1996), Spectrophotometric image analysis of fine art paintings, *in* 'IS&T and SID's 4th Color Imaging Conference: Color Science, Systems and Applications', Scottsdale, Arizona, pp. 50–53.

Maloney, L. T. (1992), *Evaluation of linear models of surface spectral reflectance with small numbers of parameters*, Jones and Bartlett Publishers, Inc., USA.

Maloney, L. T. and Wandell, B. A. (1986), "Color constancy: a method for recovering surface spectral reflectance", *J. Opt. Soc. Am. A* , Vol. 3(1), OSA, pp. 29–33.

Mansouri, A., Marzani, F. S. and Gouton, P. (2005), Neural networks in two cascade algorithms for spectral reflectance reconstruction, *in* 'IEEE International Conference on Image Processing (IEEE, 2005)', pp. 2053–2056.

Marapane, S. B. and Trivedi, M. M. (1994), "Multi-primitive hierarchical (MPH) stereo analysis", *IEEE Trans. Pattern Anal. Mach. Intell.* , Vol. 16(3), IEEE Computer Society, Washington, DC, USA, pp. 227–240.

Marr, D. and Poggio, T. (1979), A computational theory of human stereo vision, *in* 'Proc. Roy. Soc. London, Series B, Biological Sciences', Vol. 204, pp. 301–328.

Martinez, K., Cupitt, J., Saunders, D. and Pillay, R. (2002), Ten years of art imaging research, *in* 'Proceedings of the IEEE', pp. 28–41.

Matas, J., Chum, O., Urban, M. and Pajdla, T. (2004), "Robust wide-baseline stereo from maximally stable extremal regions", *Image and Vision Computing* , Vol. 22(10), pp. 761–767. British Machine Vision Computing 2002.

Mehl, P. M., Chen, Y.-R., Kim, M. S. and Chan, D. E. (2004), "Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations", *Journal of Food Engineering* , Vol. 61(1), Elsevier, pp. 67–81.

Miao, L. and Qi, H. (2004), A generic method for generating multi-spectral filter array, *in* 'IEEE International Conference on Image Processing, ICIP'04', Vol. 5, pp. 3343–3346.

Miller, P. J. and Hoyt, C. C. (1995), Multispectral imaging with a liquid crystal tunable filter, *in* 'Optics in Agriculture, Forestry, and Biological Processing', Vol. 2345 of *SPIE Proceedings*, pp. 354–365.

Miyake, Y. and Yokoyama, Y. (1998), Obtaining and reproduction of accurate color images based on human perception, *in* G. B. Beretta and R. Eschbach, eds, 'Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III', Vol. 3300 of *SPIE Proceedings*, pp. 190–197.

Mourad, S. and Kornfeld, C. (2006), Doubling the color gamut volume of ink jet prints using a simple post-processing, *in* 'CGIV '06: 3rd European Conference on Colour in Graphics, Imaging, and Vision'.

Murakami, Y., Ishii, J., Obi, T., Yamaguchi, M. and Ohyama, N. (2004), "Color conversion method for multi-primary display for spectral color reproduction", *Journal of Electronic Imaging* , Vol. 13(4), SPIE, pp. 701–708.

Novati G., P. P. and R., S. (2004), Selection of filters for multispectral acquisition using the filter vectors analysis method, *in* 'Color Imaging IX: Processing, Hardcopy, and Applications', Vol. 5293 of *SPIE Proceedings*, pp. 20–26.

Ohsawa, K., Ajito, T., Komiya, Y., Fukuda, H., Hanelshi, H., Yamaguchi, M. and Ohyama, N. (2004), "Six band HDTV camera system for spectrum-based color reproduction", *Journal of Imaging Science and Technology* , Vol. 48; PART 2, IS&T, pp. 85–92.

Omega (2010), 'Omega filters', Online, `https://www.omegafilters.com/curvo2/`. Last Visit: May, 2010.

Ononye, A. E., Vodacek, A. and Saber, E. (2007), "Automated extraction of fire line parameters from multispectral infrared images", *Remote Sensing of Environment* , Vol. 108(2), pp. 179–188.

OpenCV (2010), 'Open Source Computer Vision', Online, `http://opencv.jp/opencv-2.1_org/c/camera_calibration_and_3d_reconstruction.html`. Last Visit: June, 2010.

Park, J., Lee, M., Grossberg, M. D. and Nayar, S. K. (2007), Multispectral imaging using multiplexed illumination, *in* 'IEEE International Conference on Computer Vision (ICCV)'.

Park, S. O., Kim, H. S., Park, J. M. and Eem, J. K. (1995), Development of spectral sensitivity measurement system of image sensor devices, *in* 'IS&T and SID's 3rd Color Imaging Conference: Color Science, Systems and Applications', Scottsdale, Arizona, pp. 115–118.

Parkkinen, J. P. S., Hallikainen, J. and Jaaskelainen, T. (1989), "Characteristic spectra of munsell colors", *J. Opt. Soc. Am. A* , Vol. 6(2), OSA, pp. 318–322.

Peercy, M. S. (1993), Linear color representations for full speed spectral rendering, *in* 'SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques', ACM, New York, NY, USA, pp. 191–198.

Pilu, M. (1997), Uncalibrated stereo correspondence by singular value decomposition, Technical report, HP Laboratories Bristol.

Pratt, W. K. and Mancill, C. E. (1976), "Spectral estimation techniques for the spectral calibration of a color image scanner", *Appl. Opt.* , Vol. 15(1), OSA, pp. 73–75.

Ross, W. P. (1993), A practical stereo vision system, *in* 'IEEE Conference on Computer Vision and Pattern Recognition', pp. 148–153.

Rosselet, A. C., Graff, W., Wild, U. P., Keller, C. U. and Gschwind, R. (1995), Persistent spectral hole burning used for spectrally high-resolved imaging of the sun, *in* 'Imaging Spectrometry', Vol. 2480 of *SPIE Proceedings*, pp. 205–212.

Sanger, T. D. (1988), "Stereo disparity computation using gabor filters", *Biological Cybernetics* , Vol. 59(6), Springer Berlin / Heidelberg, pp. 405–418.

Scharstein, D. and Szeliski, R. (2003), High-accuracy stereo depth maps using structured light, *in* 'Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003,', Vol. 1, pp. 195–202.

Sharma, G. and Trussell, H. (1993), Characterization of scanner sensitivity, *in* 'IS&T and SID's Color Imaging Conference: Transforms & Transportability of Color', IS&T, pp. 103–107.

Sharma, G. and Trussell, H. (1996), "Set theoretic estimation in color scanner characterization", *J. Electronic Imaging* , Vol. 5, SPIE, pp. 479–489.

Sherman, D. and Farrell, J. E. (1994), When to use linear models for color calibration, *in* 'IS&T and SID's 2nd Color Imaging Conference: Color Science, Systems and Application', Scottsdale, Arizona, pp. 33–36.

Shi, M. and Healey, G. (2002), "Using reflectance models for color scanner calibration", *J. Opt. Soc. Am. A* , Vol. 19(4), OSA, pp. 645–656.

SPM (2010), 'Stereo Photo Maker', Online, `http://stereo.jpn.org/eng/stphmkr/index.html`. Last Visit: June, 2010.

Stefano, L. D., Marchionni, M. and Mattoccia, S. (2004), "A fast area-based stereo matching algorithm", *Image and Vision Computing* , Vol. 22(12), pp. 983–1005. Proceedings from the 15th International Conference on Vision Interface.

Stereoeye.jp (2010), 'How to view ...', Online, `http://www.stereoeye.jp/index_e.html`. Last Visit: June, 2010.

Sugiura, H., Kuno, T. and Ikeda, H. (1998), Methods of measurement for color reproduction of digital cameras, *in* G. M. Williams and Jr., eds, 'Digital Solid State Cameras: Designs and Applications', Vol. 3302 of *SPIE Proceedings*, pp. 113–122.

Sun, C. (2002), "Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques", *Int. J. Comput. Vision* , Vol. 47(1-3), Kluwer Academic Publishers, Hingham, MA, USA, pp. 99–117.

Swain, P. H. and Davis, S. M. (1978), *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York.

Tang, B., Ait-Boudaoud, D., Matuszewski, B. J. and Shark, L. K. (2006), An efficient feature based matching algorithm for stereo images, *in* 'GMAI '06: Proceedings of the conference on Geometric Modeling and Imaging', IEEE Computer Society, Washington, DC, USA, pp. 195–202.

Tao, T., Koo, J. C. and Choi, H. R. (2008), A fast block matching algorthim for stereo correspondence, *in* 'Cybernetics and Intelligent Systems, 2008 IEEE Conference on', pp. 38 –41.

Taxt, T. and Lundervold, A. (1994), "Multispectral analysis of the brain using magnetic resonance imaging", *Medical Imaging, IEEE Transactions on* , Vol. 13(3), pp. 470 –481.

Toivanen, P. J., Ansamäki, J., Parkkinen, J. P. S. and Mielikäinen, J. (2003), "Edge detection in multispectral images using the self-organizing map", *Pattern Recognition Letters* , Vol. 24(16), Elsevier, pp. 2987–2994.

Tominaga, S. (1996), "Multichannel vision system for estimating surface and illumination functions", *J. Opt. Soc. Am. A* , Vol. 13(11), OSA, pp. 2163–2173.

Tominaga, S. (1999), "Spectral imaging by a multichannel camera", *Journal of Electronic Imaging* , Vol. 8(4), SPIE, pp. 332–341.

Tsai, R. (1986), An efficient and accurate camera calibration technique for 3D machine vision, *in* 'Proceedings of IEEE Conference on Computer Vision and Pattern Recognition', pp. 364–374.

Tsai, R. Y. (1987), "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses", *IEEE J. Robotics Automat.* , Vol. Vol. RA-3, pp. 323–344.

Tsuchida, M., Sakaguchi, Y., Arai, H., Nishiko, M., Fujikawa, N., Yamaguchi, M., Haneishi, H. and Ohyama, N. (2005), High-fidelity color reproduction based on multi-channel BTF/BRDF acquisition, rendering and display, *in* 'SIGGRAPH '05: ACM SIGGRAPH 2005 Sketches', ACM, New York, NY, USA, p. 42.

Tsumura, N. (2006), "Appearance reproduction and multispectral imaging", *Color Research and Application* , Vol. 31(4), pp. 270–277.

Valero, E. M., Nieves, J. L., Nascimento, S. M. C., Amano, K. and Foster, D. H. (2007), "Recovering spectral data from natural scenes with an RGB digital camera", *Color Research & Application* , Vol. 32, pp. 352–360.

Verdu, M., F., Pujol, J. and Capilla, P. (1998), Designing a tristimulus colorimeter from a conventional machine vision system, *in* 'CIM'98, Colour Imaging in Multimedia', Derby, UK, pp. 319–333.

Wang, W., Hauta-Kasari, M. and Toyooka, S. (1997), Optimal filters design for measuring colors using unsupervised neural network, *in* '8th Congress of the International Colour Association, AIC Color 97', Vol. I, Color Science Association of Japan, Kyoto, Japan, pp. 419–422.

Wang, Z. F. and Zheng, Z. G. (2008), "A region based stereo matching algorithm using cooperative optimization", *Computer Vision and Pattern Recognition, IEEE Computer Society Conference* , Vol. 0, IEEE Computer Society, Los Alamitos, CA, USA, pp. 1–8.

Weng, J., Ahuja, N. and Huang, T. (1988), Two-view matching, *in* 'Second International Conference on Computer Vision', pp. 64–73.

Wikipedia (2010*a*), 'Gradient descent', Online, `http://en.wikipedia.org/wiki/Gradient_descent`. Last Visit: June, 2010.

Wikipedia (2010*b*), 'Rodrigues' rotation formula', Online, `http://en.wikipedia.org/wiki/Rodrigues'_rotation_formula`. Last Visit: June, 2010.

Wikipedia (2010*c*), 'Spectral power distribution', Online, `http://en.wikipedia.org/wiki/Spectral_power_distribution`. Last Visit: June, 2010.

Yamaguchi, M., Haneishi, H. and Ohyama, N. (2008), "Beyond Red–Green–Blue (RGB): Spectrum-based color imaging technology", *Journal of Imaging Science and Technology* , Vol. 52(1), IS&T, pp. 1–15.

Yamaguchi, M., Iwama, R., Ohya, Y., Obi, T., Ohyama, N., Komiya, Y. and Wada, T. (1997), "Natural color reproduction in the television system for telemedicime", *Medical Imaging 1997: Image Display* , Vol. 3031(1), SPIE, pp. 482–489.

Yamaguchi, M., Teraji, T., Ohsawa, K., Uchiyama, T., Motomura, H., Murakami, Y. and Ohyama, N. (2002), Color image reproduction based on the multispectral and multiprimary imaging: Experimental evaluation, *in* 'Color Imaging: Device-Independent Color, Color Hardcopy, and Applications VII', Vol. 4663 of *SPIE Proceedings*, pp. 15–26.

Yokohama, Y. (1997), "A new color management system based on human perception and its application to recording and reproduction of art printing", *Proceedings of IS&T/SID's 5th Color Imaging Conference, Color Science, Systems and Appl.* , IS&T, pp. 169–172.

Zhang, X. and Xu, H. (2008), "Reconstructing spectral reflectance by dividing spectral space and extending the principal components in principal component analysis", *J. Opt. Soc. Am. A* , Vol. 25(2), OSA, pp. 371–378.

Zhang, Z. (2000), "A flexible new technique for camera calibration", *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , Vol. 22(11), pp. 1330–1334.

Zhao, Y. and Berns, R. S. (2007), “Image-based spectral reflectance reconstruction using the matrix R method”, *Color Research and Application* , Vol. 32, pp. 343–351.

Zitnick, C. and Kanade, T. (1999), A cooperative algorithm for stereo matching and occlusion detection, Technical Report CMU-RI-TR-99-35, Robotics Institute, Pittsburgh, PA.