

2017

## Robust Specularity Removal from Hand-held Videos

PriyaankaDevi Guggilapu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Guggilapu, PriyaankaDevi, "Robust Specularity Removal from Hand-held Videos" (2017). *Graduate Theses, Dissertations, and Problem Reports*. 5722.

<https://researchrepository.wvu.edu/etd/5722>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# Robust Specularity Removal from Hand-held Videos

PriyaankaDevi Guggilapu

Thesis submitted to the  
Benjamin M. Statler College of Engineering and Mineral Resources  
at West Virginia University  
in partial fulfilment of the requirements for the degree of

Master of Science  
in  
Electrical Engineering

Xin Li, Ph.D., Chair  
Mark Tseytlin, Ph.D., (Co-Chair)  
Matthew C. Valenti, Ph. D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia

2017

Keywords: Specular reflection, diffuse reflection, Specularity removal, Kanade-Lucas-Tomasi (KLT) algorithm, YUV color space, Augmented Lagrange Multiplier (ALM), Alternating Direction Minimizing (ADM), Homography transformation.

Copyright © 2017 PriyaankaDevi Guggilapu

# Abstract

## Robust Specularity Removal from Hand-held Videos

PriyaankaDevi Guggilapu

Specular reflection exists when one tries to record a photo or video through a transparent glass medium or opaque surfaces such as plastics, ceramics, polyester and human skin, which can be well described as the superposition of a transmitted layer and a reflection layer. These specular reflections often confound the algorithms developed for image analysis, computer vision and pattern recognition. To obtain a pure diffuse reflection component, specularity (highlights) needs to be removed. To handle this problem, a novel and robust algorithm is formulated. The contributions of this work are three-fold.

First, the smoothness of the video along with the temporal coherence and illumination changes are preserved by reducing the flickering and jagged edges caused by hand-held video acquisition and homography transformation respectively.

Second, this algorithm is designed to improve upon the state-of-art algorithms by automatically selecting the region of interest (ROI) for all the frames, reducing the computational time and complexity by utilizing the luminance (Y) channel and exploiting the Augmented Lagrange Multiplier (ALM) with Alternating Direction Minimizing (ADM) to facilitate the derivation of solution algorithms.

Third, a quantity metrics is devised, which objectively quantifies the amount of specularity in each frame of a hand-held video. The proposed specularity removal algorithm is compared against existing state-of-art algorithms using the newly-developed quantity metrics. Experimental results validate that the developed algorithm has superior performance in terms of computation time, quality and accuracy.

To My Parents and Sister

## Acknowledgments

It gives me an immense pleasure and pride to express my deep sense of gratitude and respect for my teacher and guide Dr. Xin Li for his evergreen expertise and inspiring guidance throughout the period of my work. I am indebted to him for enlightening me on the finer skills of dealing with problems. He is the one behind giving me a decisive turn and significant boost to my career. Every meeting with him motivated me towards pragmatic thinking to solve research problems. His insights into every problem proved to be remarkable in successful completion of my master's study. I consider myself one of the fortunate people to be associated with him.

I thank Dr. Mark Tseytlin, for accepting to serve as the committee member. I am extremely fortunate to work with Dr. Mark in his project where I have honed my MATLAB skills to a very large extent in addition to learning the key concepts of EPR (Electron Paramagnetic Resonance). He explains every problem in the easiest way and encourages me to find the solution for them. I always get feedback from him which has helped me to improve my skills. I am very thankful for his encouragement and guidance.

I would like to offer my sincere thanks to Dr. Matthew Valenti, for his valuable suggestions towards this thesis. Despite his busy schedule, Dr. Valenti always responded immediately to any of my questions and encouraged me by providing valuable guidance. He has been an excellent teacher and his course has helped immensely to improve my problem-solving skills. He also took great concern about my class performance and made sure that I understood the key details.

I would like to express my love, respect and gratitude to my lovely parents, Sravanthi (sister) without whose invaluable love, care, help, sacrifice and belief, this journey so far would never have been achieved. I thank God who is the most merciful for giving me energy and his blessings for pursuing this project.

As a final word, I would like to extend my sincere thanks to every individual whose name has slipped from my mind have been a source of support and helped me to complete my dissertation work successfully.

Priyaanka Devi Guggilapu

# Table of Contents

List of Tables .....	vii
List of Algorithms.....	viii
1. Introduction .....	1
1.1 Importance of Specularity Removal.....	1
1.2 Thesis Framework .....	3
1.3 Video Data Collection .....	4
1.4 Challenges in Separation.....	5
1.5 Contributions.....	7
1.6 Thesis Outline .....	8
2. Background .....	9
2.1 Illumination and Surface Reflectance .....	9
2.1.1 Specular Reflection .....	12
2.1.2 Diffuse Reflection .....	12
2.2 Color and Color Space .....	13
2.2.1 RGB Color Space .....	14
2.2.2 YUV Color Space.....	15
2.2.3 RGB to YUV Conversion.....	17
2.2.4 YUV to RGB Conversion .....	18
2.3 Saliency Detection .....	19
2.3.1 Graph based visual saliency & Itti, Koch, Niebur Algorithm .....	19
2.4 Planar Homography Transformation .....	20
2.4.1 RANSAC method for Homography estimation .....	22
2.4.2 Limitations for Homography Transformation.....	23
3. Automatic ROI detection .....	24
3.1 Harris Corner detection .....	24
3.1.1 Mathematical formulation .....	24
3.1.2 Limitations of Harris Corner Detector .....	28
3.2 Tracking using KLT Algorithm with Error detection .....	28
3.2.1 Template Image $T(x)$ .....	28

3.2.2 Optimization problem and Solution Algorithm .....	29
4. Specularity Removal Algorithm .....	32
4.1 Problem formulation .....	32
4.1.1 First Structural Prior.....	32
4.1.2 Second Structural Prior .....	32
4.2 Optimization Problem .....	33
4.3 Solution Algorithm.....	34
4.3.1 Solution Algorithm for specularity removal .....	34
4.3.2 Determining the optimum number of frames.....	37
4.4 Final image to Video sequence conversion.....	38
5. Quantity Metrics.....	39
5.1 Quality Evaluation metrics .....	41
6. Experimental Results .....	44
6.1 Results of Automatic detection of ROI.....	44
6.2 Results of tracking using KLT algorithm .....	46
6.3 Results of Specularity Removal.....	47
6.4 Results for Comparison with State of Art Algorithms.....	49
6.5 Implementation Details .....	50
7. Conclusion .....	52
7.1 Future Scope .....	52
8. Electron Paramagnetic Resonance Imaging.....	53
8.1 Introduction to EPR .....	53
8.2 Resonance Phenomenon.....	54
8.3 Bloch Equation.....	55
8.4 Why do we need EPR? .....	57
8.5 Radio Frequency (RF) Electronic design.....	59
8.6 Audio Frequency modulation and Scans design.....	59
8.7 Magnet and gradient system design.....	60
8.8 EPR Imaging.....	61
Bibliography .....	64

# List of Figures

Figure 1.1 Specular reflections through glass, opaque surface [19], human skin [1] and iris respectively.....	1
Figure 1.2 (a) Colonoscope generated specular highlight [2], (b) Iris detection, (c) Number plate recognition.....	2
Figure 1.3 Main schema of the thesis .....	3
Figure 1.4 (a) Book placed on an opaque surface, (b) Photo frame, (c) Book in high illumination condition. ....	5
Figure 2.1 Natural and artificial sources of illumination.....	9
Figure 2.2 Bidirectional Reflectance Distribution Function.....	10
Figure 2.3 Mechanism of Surface Reflection. ....	11
Figure 2.4 Specular Reflection and Diffuse Reflection.....	12
Figure 2.5 (a) Additive color mixing, (b) Subtractive color mixing [25].....	13
Figure 2.6 RGB Color Space .....	14
Figure 2.7 R, G, B channels for an Image .....	15
Figure 2.8 (a) Original RGB image, (b) Y channel, (c) U channel, (d) V channel.....	16
Figure 2.9 (a) Color based visual saliency, (b) Orientation based visual saliency .....	19
Figure 2.10 (a) GBVS saliency map, (b) Itti, Koch, Niebur saliency map.....	20
Figure 2.11 Planar Homography transformation .....	21
Figure 2.12 Mapping of points during planar homography.....	21
Figure 2.13 Non-homogeneous coordinates .....	21
Figure 2.14 Projective transformation Example .....	22
Figure 2.15 Jagged edges due to Projective transform .....	23
Figure 3.1 Basic idea of a “flat”, “edge” and “corner” .....	24
Figure 3.2 Classification of the points based on Eigen values .....	25
Figure 3.3 Corner Response Map .....	26
Figure 3.4 Tracking a point (x,y).....	29
Figure 3.5 Distinct Warp functions for different transformations [33] .....	30
Figure 3.6 Illustration of Algorithm 1.....	31
Figure 4.1 Plot of number of input frames with respect to energy in transmission layer and reflection layer .....	37
Figure 5.1 Quantity metrics for specularity measurement Frame-by-Frame basis.....	39



Figure 5.2 Quantity metrics for specularity measurement.....	40
Figure 5.3 GUI for Specularity measurement (Low Specularity).....	40
Figure 5.4 GUI for Specularity measurement (High Specularity).....	41
Figure 5.5 Quality Metrics for Amount of Specularity.....	42
Figure 6.1 Results from Automatic detection of the ROI in (a) Book placed on an opaque surface, (b) Photo frame, (c) Book in high illumination condition.....	45
Figure 6.2 Results from ROI tracking using KLT in (a) Book placed on an opaque surface, (b) Photo frame, (c) Book in high illumination condition.....	47
Figure 6.3 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by our method for different video sequences .....	48
Figure 6.4 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by [21] .....	49
Figure 6.5 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by [36] .....	49
Figure 6.6 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by [37] .....	50
Figure 6.7 Computational time for different number of frames for Y channel and RGB channel .....	51
Figure 8.1 Electron Spinning States.....	53
Figure 8.2 Splitting of Energy levels .....	54
Figure 8.3 Precession of a Top .....	54
Figure 8.4 Electro-Magnetic Wave.....	55
Figure 8.5 Bloch Equation .....	56
Figure 8.6 Absorption and dispersion Graph.....	57
Figure 8.7 Survey by National Cancer Institute.....	58
Figure 8.8 Survey by Centers for disease control and Prevention.....	58
Figure 8.9 Rapid scan coil (White colored) and Permanent Magnet (black Colored).....	60
Figure 8.10 GUI for Control and Data Acquisition .....	61
Figure 8.11 4-D spectral-spatial images in vivo .....	62
Figure 8.12 Fitted Absorption + Dispersion .....	62

## List of Algorithms

Algorithm 1 Automatic detection of the ROI of first frame .....	27
Algorithm 2 ROI Tracking (Specularity Robust) .....	31
Algorithm 3 Specularity Removal .....	36
Algorithm 4 Images to Video Sequence .....	38
Algorithm 5 Video Quantity Evaluation (Specularity).....	43

# 1. Introduction

## 1.1 Importance of Specularity removal

The algorithms developed for computer vision tasks, image analysis, and pattern recognition assumes that the surface of the objects is purely diffuse. When recording a photo or video from a transparent or opaque surface one can experience the occlusion of a part of image or video, where these algorithms become erroneous. To get only pure diffuse reflection, the specularity component needs to be completely removed.

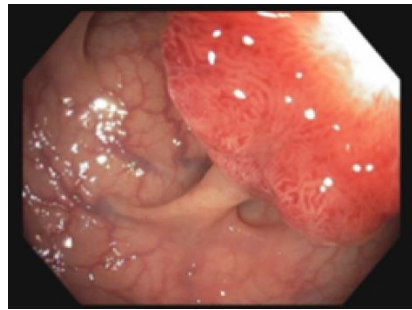
Figure 1.1 contains the specular reflection when capturing the image through the inhomogeneous material such as human skin, glass, plastics, opaque surfaces etc.



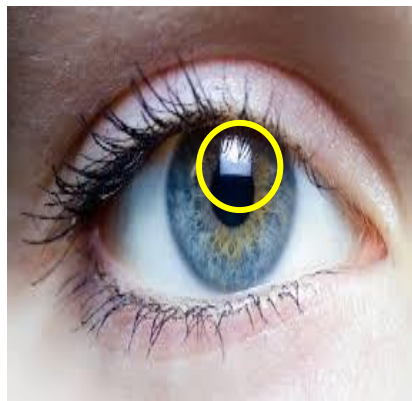
Figure 1.1 Specular reflections through glass, opaque surface [19], human skin [1] and iris respectively.

Due to growth in the application of the computer vision, many specular reflection separation algorithms have been formulated. Specularity elimination requires no camera calibration or other a priori information regarding the scene.

Figure 1.2 shows some of the domains in which specular reflection obstructs the computer vision tasks.



(a)



(b)



(c)

**Figure 1.2 (a) Colonoscope generated specular highlight [2], (b) Iris detection, (c) Number plate recognition**

## 1.2 Thesis Framework

Figure 1.3 shows the process scheme of this thesis. It proposes an ideal way for specular reflection separation from a video. Given an input video sequence, at the end of the process, we'll have an output video sequence free from the specularity. To achieve this goal, the input video sequence is processed through many stages as shown in the flowchart. Specular reflection separation has been mostly performed on either multiple images or single images and is a highly ill-posed problem as the number of unknowns to be recovered are twice as many as that of the given inputs.

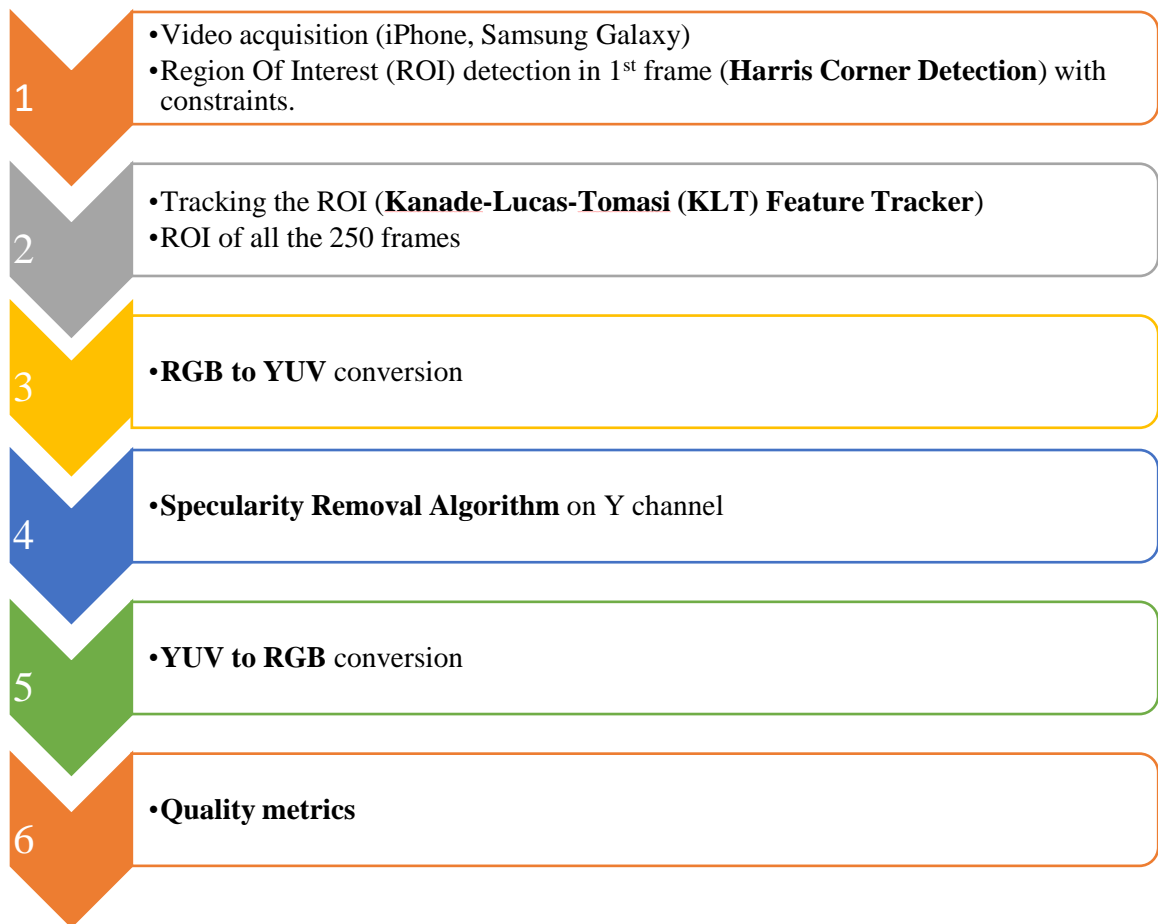
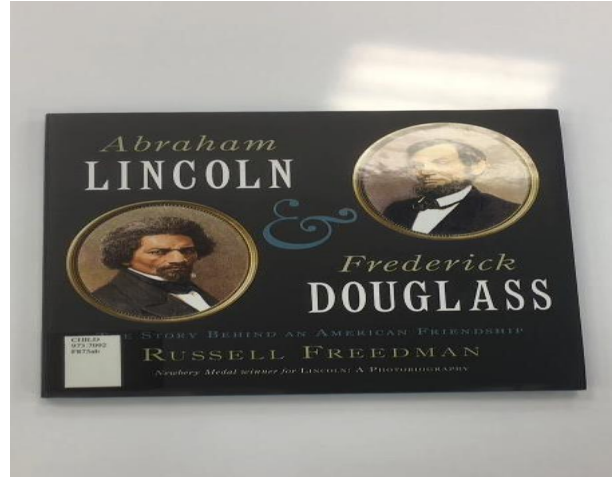
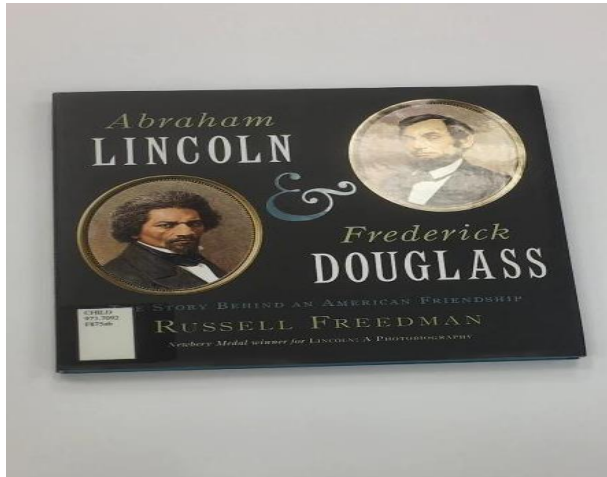


Figure 1.3 Main schema of the thesis

### 1.3 Video Data Collection

Since the specular reflection separation from a video is a novel idea, there weren't any available datasets. So, I have collected the video data using iPhone 6S which are in .MOV format. These data are recorded in our university library. Figure 1.4 shows the video data collected, for the understanding of the reader, two frames from each data are displayed.

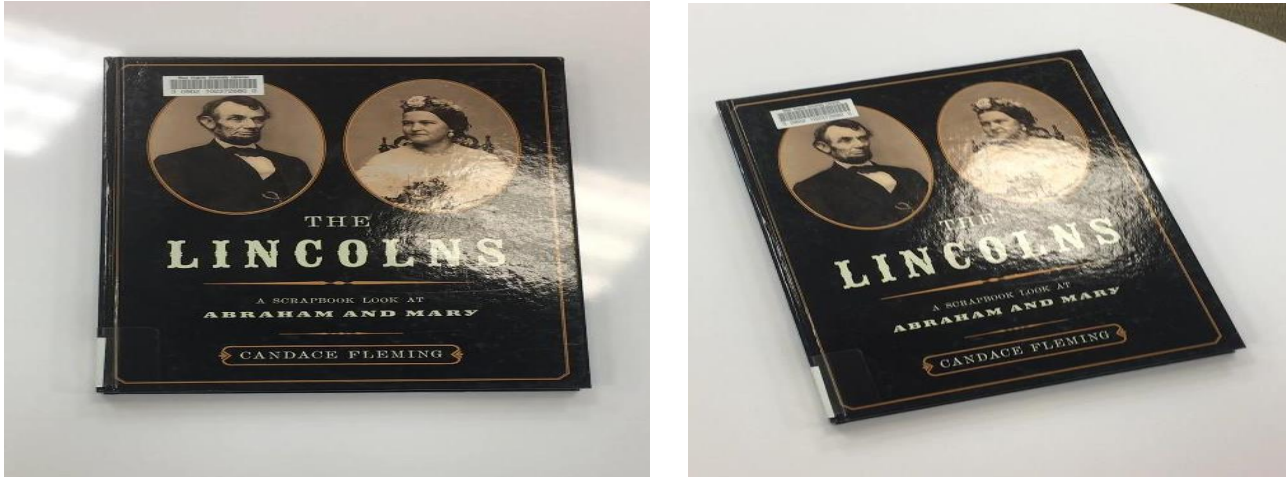


(a)



(b)





(c)

Figure 1.4 (a) Book placed on an opaque surface, (b) Photo frame, (c) Book in high illumination condition

## 1.4 Challenges in Separation

Many algorithms have been formulated to separate the transmission layer and the reflection layers which can be grouped into two methods

1. Single-Image Methods
2. Multi-Image Methods

**Single-Image Methods:** These methods use only a single image to separate the layers in an image. They do not require several images of the identical part taken at diverse angles, which needs time and effort in the image acquisition process. Papers based on these methods can be clustered into two different categories based on their approach.

Firstly, **Color space analysis**, in this approach the analysis of color space aids in the separation of the highlights from object colors. **Klinker et al.** [4] classified the color pixels as matte (evince only body reflection), highlight (both body and specular reflections) and clipped (this behaves as a specularly if the light reflection exceeds the dynamic range of camera) [3]. Whereas, **Two-dimensional diagram approach**, which is superior in terms of computational time, runs mainly in the UV-space. In this approach, the RGB data is converted to UV-space in the first step and then to h-space in the succeeding step, later the h-space is processed by a morphological filter [5,6]. In **Bajcsy et al. approach**, a novel S-color space was introduced [7]. Klinker et al. approach, Two-

dimensional diagram approach and Bajcsy et al. approach together contribute towards Color space analysis.

Secondly, **Neighborhood analysis**, in this approach some local operations are performed and color information is utilized in the separation process. Specular-free image approach, PDE approach, Inpainting techniques, color information and classifier, Fresnel reflection coefficient methods deploy the use of Neighborhood analysis in their formulation. **Specular-free image approach** generates a pseudo-diffuse component (specular-free image) [8,9,10,20, 21] and then categorize the pixels as specular or diffuse. If the pixel is labeled specular then the specular-to-diffuse algorithm is applied to produce more diffuse component. Hui Liang Shen and Zhi-Huan Zheng [19] proposed an efficient method that generates a pseudo-chromaticity image and then categorizes the pixels by using chromaticity threshold ( $T_c$ ) and percentile threshold ( $T_p$ ). In **PDE approach**, the main concept is to iteratively erode the specular component at each pixel value [11,12]. While **Inpainting techniques** are built on the process of reconstructing the lost or deteriorated parts of an image. Tan et al. [13] observed that highlight pixels contain useful information for guiding the inpainting process. Using **color information and classifier**, Tappen et al. [14] assumed that the input image is the product of the shading and reflectance images and a logarithm is applied on both sides to make it additive. These derivatives are then classified by grey-scale classifier into diffuse and reflection components. Specularity detection based on **Fresnel reflection coefficient** [15], observed that the Fresnel coefficient changes with the wavelength, which affects the color of the specular highlight. So, a segmentation step is performed by making use of mean-shift segmentation algorithm.

**Multiple-Image methods:** These methods uses the information contained in an image sequence of the same scene at different angles or with varying light information. Lin et al. [16] presented a technique that uses the sequence of images to identify the specular pixels by **color histogram difference** and then finding the stereo correspondences for the specular pixels. Once the specular pixels are identified, they are removed by associating them to their corresponding diffuse points in other images. The **Multi-Flash methods** are exploited by Feris et al. [17] and Agarwal et al. [18], which utilizes the multiple snapshots of the same image at the same point but by varying the input variable i.e. flash-light position.



Although many algorithms have been formulated they are limited by the conditions of their applicability. Most of the techniques rely on a specific reflection model and assume that the specularity varies inconsiderably with the wavelength. Sparse blind separation algorithm with spatial shifts [22,11] are under the assumption of the motions in uniform translations. RASL [23] has the drawback that the visual quality of the layers is not guaranteed. Xiaojie Guo et al. [24] proposed a promising method for decomposition of superimposed images but fails when the superimposed image is blurred.

## 1.5 Contributions

A robust specular reflection separation method requires that all the specular reflection components are removed, the diffuse reflection part to be recovered with an algorithm having less computational complexity, optimized implementation and has greater accuracy than the state of art algorithms. The contributions of this work are as delineated below

- This work extends the algorithms prevailing for the images or image sequences, to a video having camera motion and illumination changes. This algorithm is formulated to have less computational complexity and higher accuracy.
- An automatic Region of Interest detection and tracking algorithm is formulated for the 4-side polygon region of interest.
- Since in video there are more than 250 frames, we need to find an optimum number of frames and the best frames to work on. This is determined by using the energy in the transmitted layer, reflection layer and the computation time for the number of frames (5,7,10,12,14,15,16,17,18,19,20 and 21). A threshold level is reached which shows that 15 frames are the optimum number of frames to obtain accurate separation of the specular reflection from the given video sequence.
- Quantity metrics is devised for the videos to determine the amount of specularity and based on that Quality evaluation is performed.

Every frame of the acquired video should contain the 4-sided polygon region of interest since the correspondence between the frames is exploited to remove the highlights. In future, this algorithm and quantity metrics can aid for videos that have no correspondence.

## 1.6 Thesis outline

The thesis is outlined in the flowchart on Figure 1.3. Background Chapter 2 provides the basic physical and mathematical background. This chapter discusses the Specular reflection, diffuse reflection, color, color spaces (mainly algorithms for RGB to YUV and YUV to RGB) and about the saliency detection. Chapter 3 to Chapter 5 delineate the proposed method for Specular reflection separation. Chapter 3 presents the Automatic detection and tracking the of region of interest, Chapter 4 presents the most important algorithm for separation of the layers (Superimposed Image Decomposition) in detail, along with homography transformation, limitations and final image sequence to video conversion. Chapter 5 is regarding the novel Quantity metrics and the GUI developed for the side by side evaluation of the specularity. Chapter 6 is about Experimental results, Implementation Details, and Quality evaluation metrics. Chapter 7 concludes the work by presenting the limitations of this work and future research scope along this direction.

## 2. Background

This chapter provides the basic physical and mathematical background concerning to this thesis. Sec.2.1 refers to illumination, the main influential part of computer vision. On it depends the surface reflectance explained in Sec.2.1 which has two major classes, specular reflection in Sec.2.1.1 and diffuse reflection in Sec.2.1.2. Sec 2.2 introduces color and color space which has two subsections Sec.2.2.1 about the RGB color space and Sec.2.2.2 about the YUV color space. Next, the RGB to YUV and YUV to RGB conversion algorithms are explained in Sec.2.2.3 and Sec.2.2.4 respectively.

### 2.1 Illumination and Surface Reflectance

Illumination is demarcated as the quantity of source light incident on the scene. A scene without illumination turns vision system to useless. There are two types of illumination. Firstly, the natural illumination provided by the sun, which is varying with the time of the day and secondly, the artificial illumination by the artificial light sources. The illumination of a scene is defined as  $i(x,y)$  which satisfies  $0 < i(x,y) < \mathbf{inf}$  condition.



Figure 2.1 Natural and artificial sources of illumination

Intensity of an image depends on illumination and reflectance. Reflectance is the amount of the light reflected by the object in the scene denoted by  $r(x,y)$ .

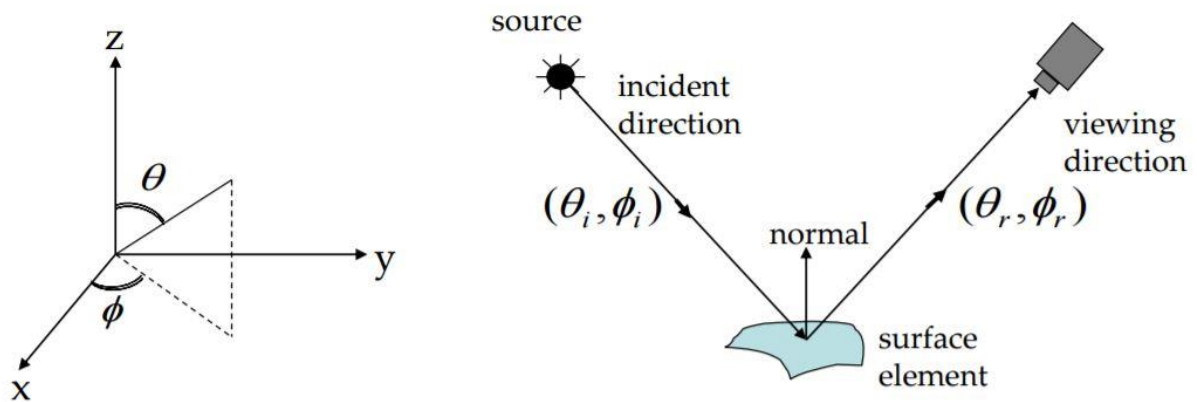
$$f(x,y) = i(x,y) r(x,y)$$

where,  $0 < i(x,y) < \infty$

and  $0 < r(x,y) < 1$ , where 1 corresponds to total reflectance and 0 to a total absorption

$f(x,y)$  is the image intensity.

Figure 2.2 shows a basic case of surface reflectance, where a beam of light strikes a surface with an angle  $\Theta_i$  respect to the normal, and outgoing beam with an angle of  $\Theta_r$ .



$E^{surface}(\theta_i, \phi_i)$  Irradiance at Surface in direction  $(\theta_i, \phi_i)$

$L^{surface}(\theta_r, \phi_r)$  Radiance of Surface in direction  $(\theta_r, \phi_r)$

$$\text{BRDF} : f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{L^{surface}(\theta_r, \phi_r)}{E^{surface}(\theta_i, \phi_i)}$$

Figure 2.2 Bidirectional Reflectance Distribution Function<sup>1</sup>

<sup>1</sup><https://www.cs.cmu.edu/afs/cs/academic/class/15462-f09/www/lec/lec8.pdf>

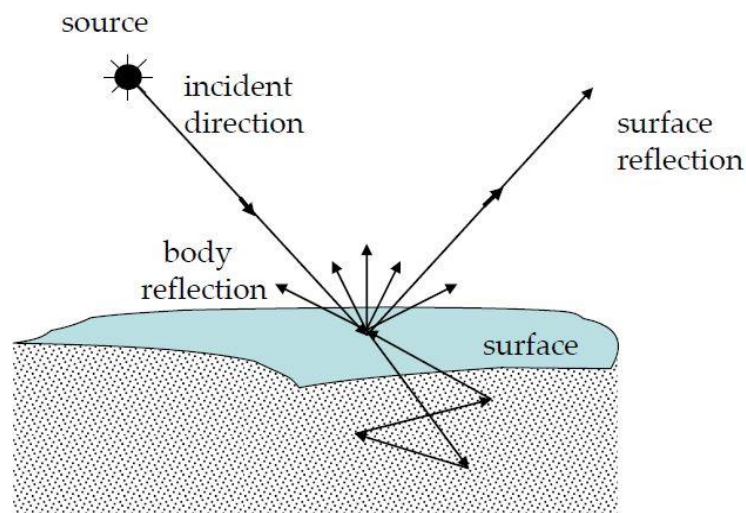
Two main points can be considered

- (a) The outgoing beam does not have the same energy and wavelength as that of the incoming beam.
- (b)  $\Theta_i$  may or may not be equal to  $\Theta_r$ .

Surfaces are divided into two main classes; specular surfaces and diffuse or Lambertian surfaces.

Figure 2.3 shows the mechanism of surface reflection; the image intensity is the additive sum of body reflection and surface reflection.

## Mechanisms of Surface Reflection



Body Reflection:

Diffuse Reflection  
Matte Appearance  
Non-Homogeneous Medium  
Clay, paper, etc

Surface Reflection:

Specular Reflection  
Glossy Appearance  
Highlights  
Dominant for Metals

Image Intensity = Body Reflection + Surface Reflection

**Figure 2.3 Mechanism of Surface Reflection<sup>1</sup>**

<sup>1</sup><https://www.cs.cmu.edu/afs/cs/academic/class/15462-f09/www/lec/lec8.pdf>

### 2.1.1 Specular Reflection

Specular reflection is due to the bouncing of light from a shiny surface such as a mirror which is highly polished or something which is like it, where parallel rays of light bounce off at the same angle. This is the main phenomenon due to which we could see our face in the mirror.

### 2.1.2 Diffuse Reflection

Diffuse reflections occur on rough surfaces such as paper, clothes, asphalt etc. These reflections behave in accordance with the laws of reflections, but due to the roughness of the surfaces, there will be variation in the normal along the surface. The normals are not parallel as that in case of the specular reflection. Diffuse reflection contributes mostly in identifying the object when compared to specular reflection.

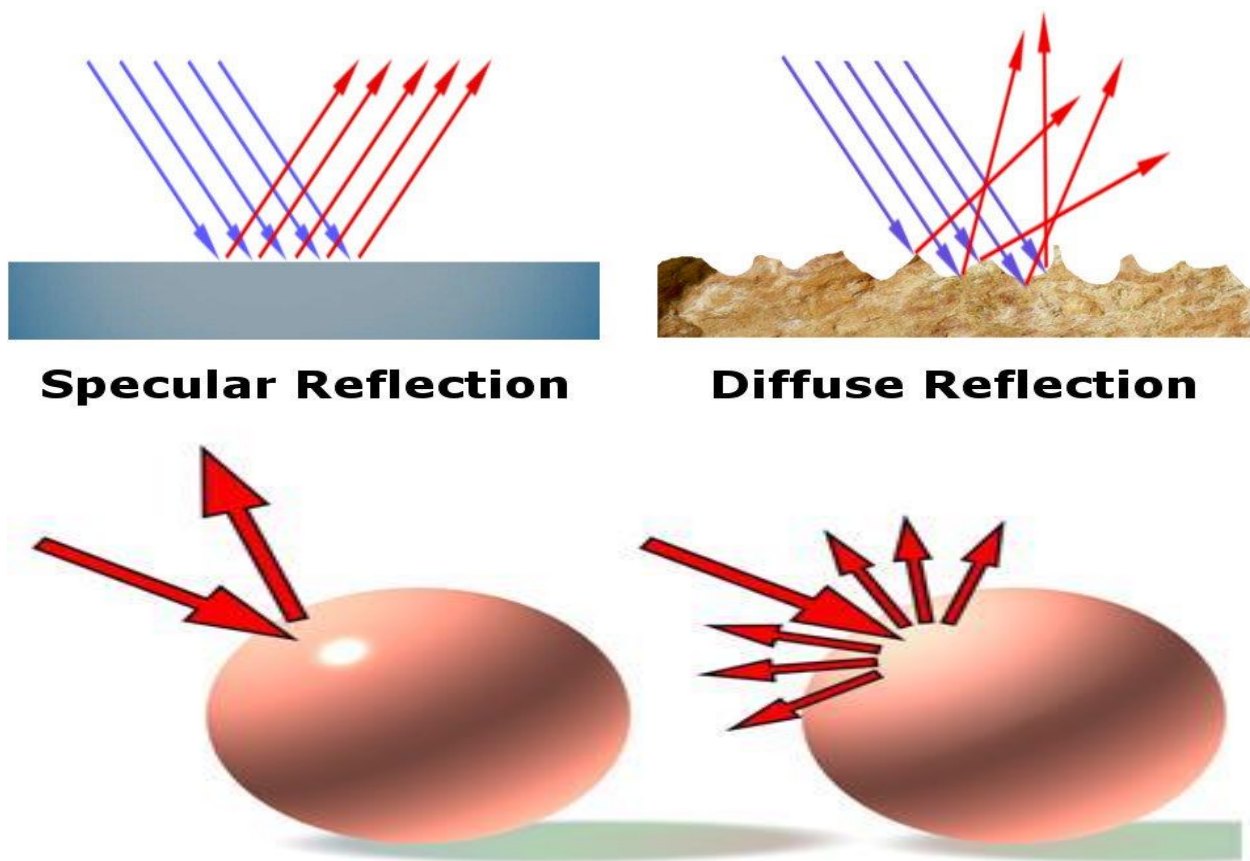


Figure 2.4 Specular Reflection and Diffuse Reflection<sup>1</sup>

<sup>1</sup><http://help.autodesk.com/>

Figure 2.4 shows the specular and diffuse reflections, the impact of these reflections on a surface.

## 2.2 Color and Color Space

Color is an important descriptor for human visual perception. Achromatic light is the light that appears on a black and white television, chromatic light spans approximately 400 to 700nm of the electromagnetic spectrum. Chromatic colors have hue, saturation, and intensity. These colors are identified due to the stimulation of the cone cells in the human eye [25]. Roughly 65% of all cones are delicate to red, 33% to green light and 2% are delicate to blue light. Based on this red, green and blue are ascertained to be the primary colors. The **primary colors** can be summed to yield the **secondary colors** of light- cyan (green + blue), magenta (red + blue), and yellow (red + green). Mixing all the primaries produces a white light and mixing of all the secondary colors produces a black light [26].

Figure 2.5 (a) shows additive color mixing, (b) shows subtractive color mixing.

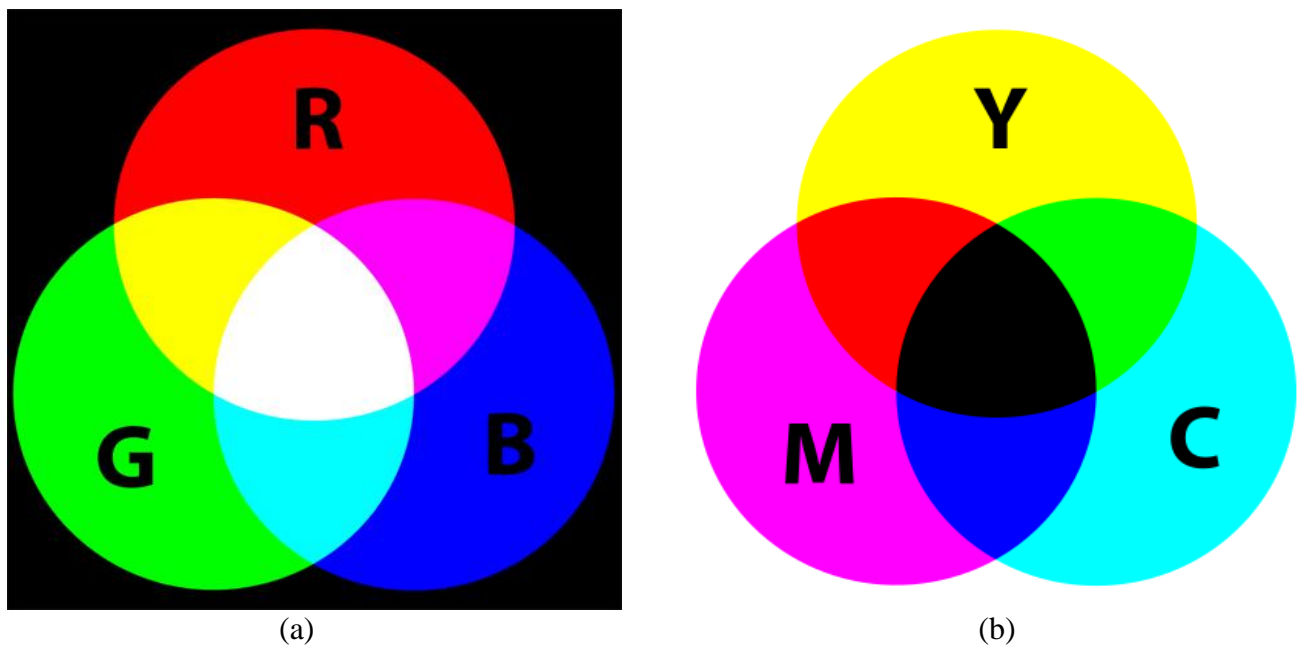


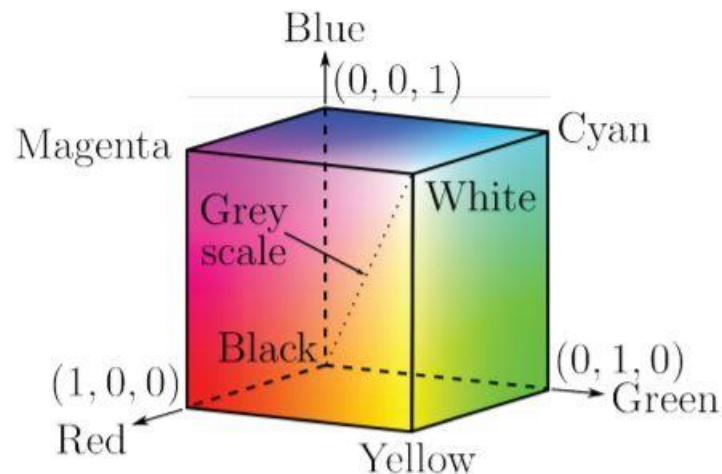
Figure 2.5 (a) Additive color mixing, (b) Subtractive color mixing [25]

The main purpose of the color space is to specify the colors in some standards. Most colors are oriented either towards hardware (color monitors and printers) or towards application where color manipulation (such as computer graphics for animation) [26]. There are innumerable color spaces which are used based on their application.

- (a) RGB color space for color monitors and video cameras
- (b) CMY and CMYK for color printing
- (c) HSI in human perception
- (d) YUV for digital television

### 2.2.1 RGB Color Space

This color space is based on a Cartesian coordinate system. In this model, each color appears as a subset of the three primary colors (red, green and blue). Figure 2.6 is RGB color space. Red, Green, and Blue are at the corners of the cube, the other three corners are covered by Cyan, Magenta and Yellow. The white color is at the coordinate (1,1,1) and black is at the origin (0,0,0).



**Figure 2.6 RGB Color Space<sup>1</sup>**

Images represented using the RGB color space consists of three component image, one for each primary color which are then combined to produce the image. Figure 2.7 shows the three channels for a bird.

<sup>1</sup>[http://eprints.grf.unizg.hr/2318/1/Z617\\_Komugovi%C4%87\\_Ana.pdf](http://eprints.grf.unizg.hr/2318/1/Z617_Komugovi%C4%87_Ana.pdf)



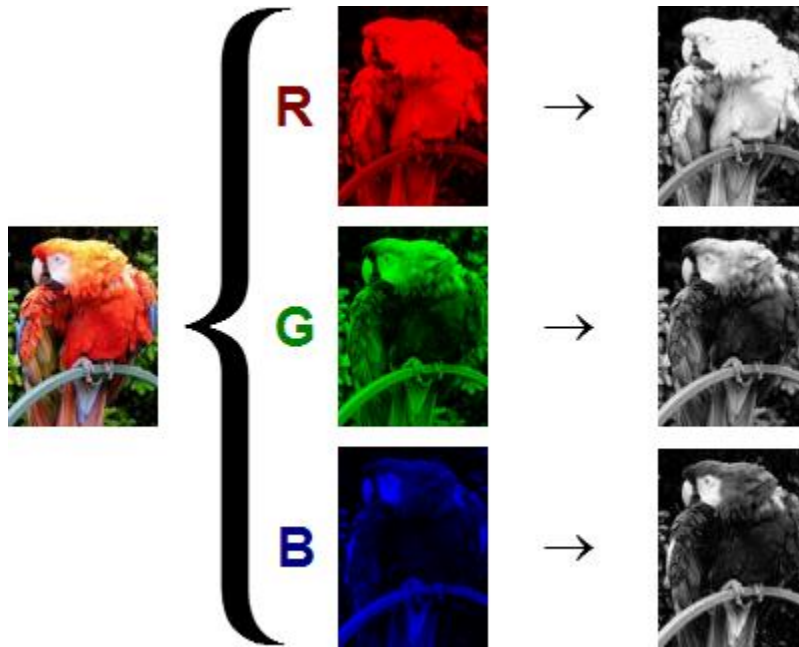


Figure 2.7 R, G, B channels for an Image<sup>1</sup>

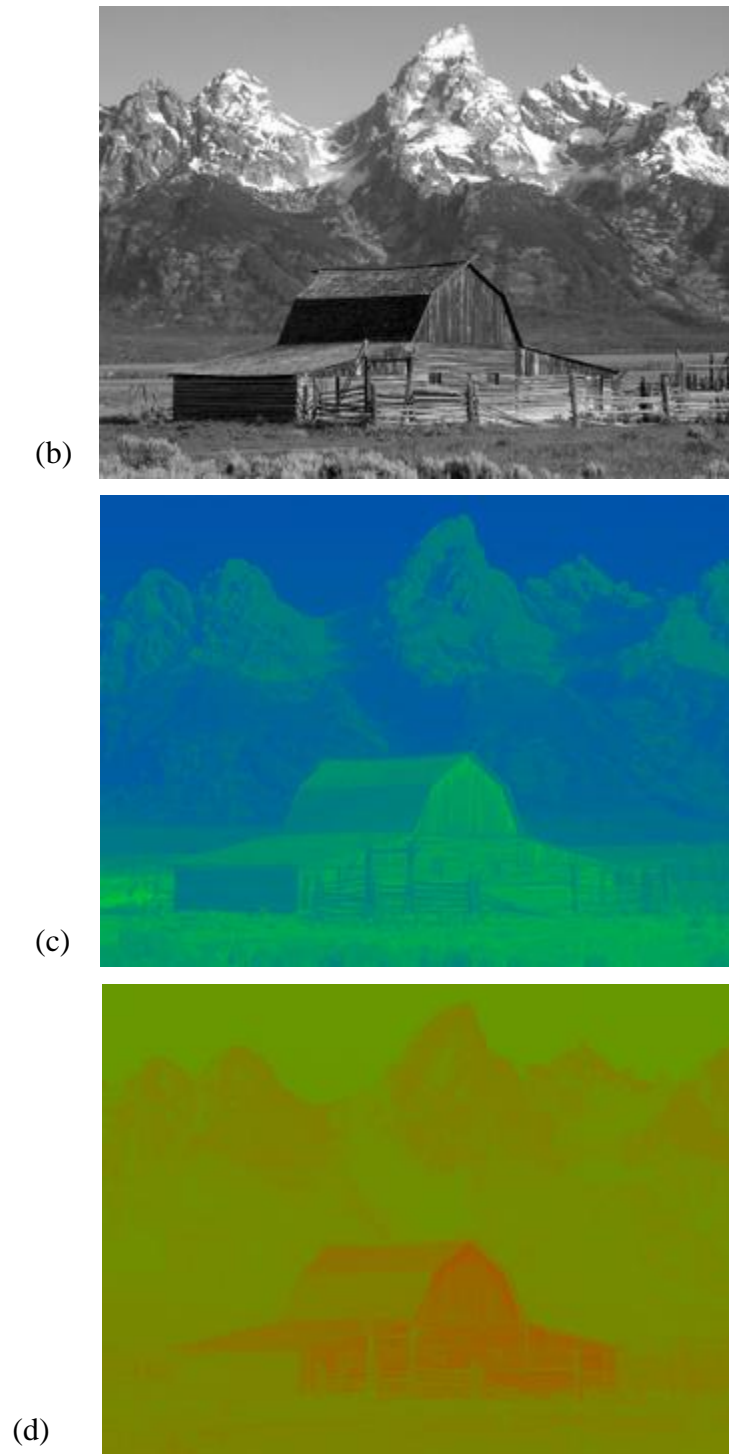
### 2.2.2 YUV Color Space

In this color space, the Y component is the brightness of the color also known as luminance, while the U and V components represents the color (Chroma). The primary advantage of this color space is that some of the information can be discarded for reducing the bandwidth. This is a common video compression standard. Recent algorithms such as MPEG-2 are dominant in use. The Chroma subsampling rate is 4:2:2. Figure 2.8 shows the YUV color space for an image.



(a)

<sup>1</sup><http://triplelift.com/2013/07/02/the-complexity-of-image-analysis-part-2-colors/>



**Figure 2.8 (a) Original RGB image, (b) Y channel, (c) U channel, (d) V channel<sup>1</sup>**

<sup>1</sup><https://en.wikipedia.org/wiki/YUV>

### 2.2.3 RGB to YUV Conversion

YUV is often used interchangeably with  $YC_bC_r$ , different equations are involved in the conversion of RGB to YUV based on the application. For digital component video,  $YC_bC_r$  color format is used. For SDTV (Standard definition TV) the following equation depicts the conversion formula.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Ranges:  
R/G/B [ 0 ... 255 ]  
Y [ 16 ... 235 ]  
Cb/Cr [ 16 ... 240 ]

RGB to  $YC_bC_r$  color conversion for SDTV

Here the range for Y is [16 to 235] and that of  $C_b$  and  $C_r$  are [16 to 240]. There is a need to have Full-range for Y,  $C_b$  and  $C_r$ . The following equations shows the conversion of RGB values to full range  $YC_bC_r$ .

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Ranges:  
R/G/B [ 0 ... 255 ]  
Y/Cb/Cr [ 0 ... 255 ]

RGB to full-range  $YC_bC_r$  color conversion

The possible range of values for luminance and chrominance reserve some footroom and headroom, which is necessary to provide some space for overshooting [27]. These [0, 128, 128] are referred as the clamping values. By omitting these values, things like negative luminance pop up, e.g. in combination with analog video equipment.

After the conversion from RGB to YUV, subsampling the Chroma channels is done to get YUV 4:2:2. A **Bi-cubic interpolation** is used to reduce the size of the U and V channels. Now the size of the Y channel will be same as that of the frame size in a video and the Chroma channels will be having half of the size of the Y channel.

## 2.2.4 YUV to RGB Conversion

This is one of the stage in Figure1.3, in this the YUV color space is converted back to RGB color space. For a standard definition TV applications (SDTV) the YUV to RGB conversion is by the following equations

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.164 & 0.000 & 1.596 \\ 1.164 & -0.392 & -0.813 \\ 1.164 & 2.017 & 0.000 \end{bmatrix} \cdot \begin{bmatrix} (Y - 16) \\ (Cb - 128) \\ (Cr - 128) \end{bmatrix}$$

Ranges:  
Y [ 16 ... 235 ]  
Cb/Cr [ 16 ... 240 ]  
R/G/B [ 0 ... 255 ]

### YCbCr to RGB color conversion for SDTV

In these equations the Y channel ranges from [16 to 235] and Cb, Cr channel ranges from [16 to 240]. To have full range for Y, C<sub>b</sub> and C<sub>r</sub>. The following equations shows the conversion of Full-range YC<sub>b</sub>C<sub>r</sub> values to RGB.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0.000 & 1.400 \\ 1.000 & -0.343 & -0.711 \\ 1.000 & 1.765 & 0.000 \end{bmatrix} \cdot \begin{bmatrix} Y \\ (Cb - 128) \\ (Cr - 128) \end{bmatrix}$$

Ranges:  
Y/Cb/Cr [ 0 ... 255 ]  
R/G/B [ 0 ... 255 ]

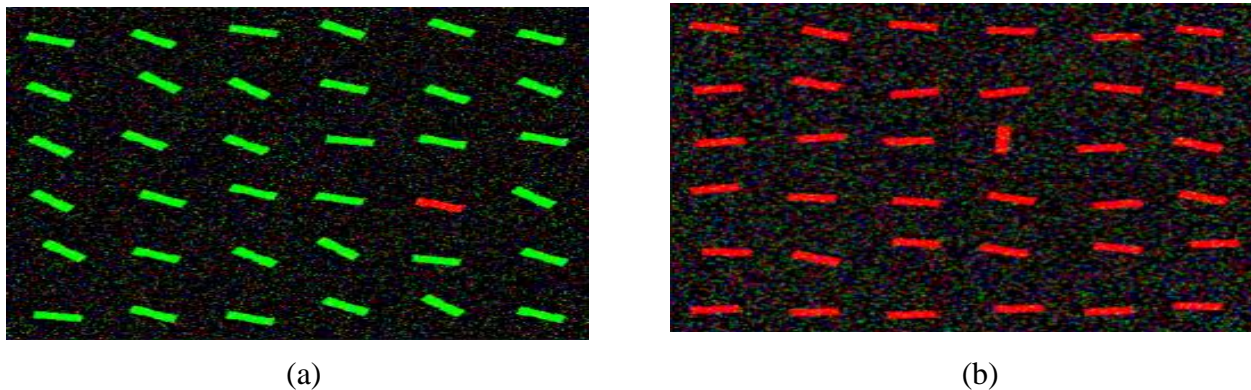
### Full-range YCbCr to RGB color conversion

Converting from YUV 4:2:2 color space to RGB color space, the first step in this algorithm is to increase the size of the Chroma channels (U and V) by 2 so that they are of the same size as the Y channel. Inorder to achieve this, **Bi-cubic Interpolation** is employed and then the conversion equations are applied onto the YUV channels.

The values in the matrix used for the conversion process are the inverse of the matrix used for RGB to YUV conversion from Sec.2.4. Typically, this full range color format is used for JPEG images. The process assigned 2 and 4 in the Figure 1.3 are formulated by employing the Full-range conversion.

## 2.3 Saliency detection

Saliency can be defined as the quality of being most prominent and easily noticeable. Visual saliency is predominant in the computer vision tasks mainly for extracting the region of interest and then to solve problems like image segmentation, object recognition and tracking. In this work, the region of interest is the salient feature to be detected and then later it was tracked by the KLT tracking algorithm along with the error detection capabilities. Figure 2.9 is an example in which one of the item appears to be highly salient than others.



**Figure 2.9 (a) Color based visual saliency, (b) Orientation based visual saliency<sup>1</sup>**

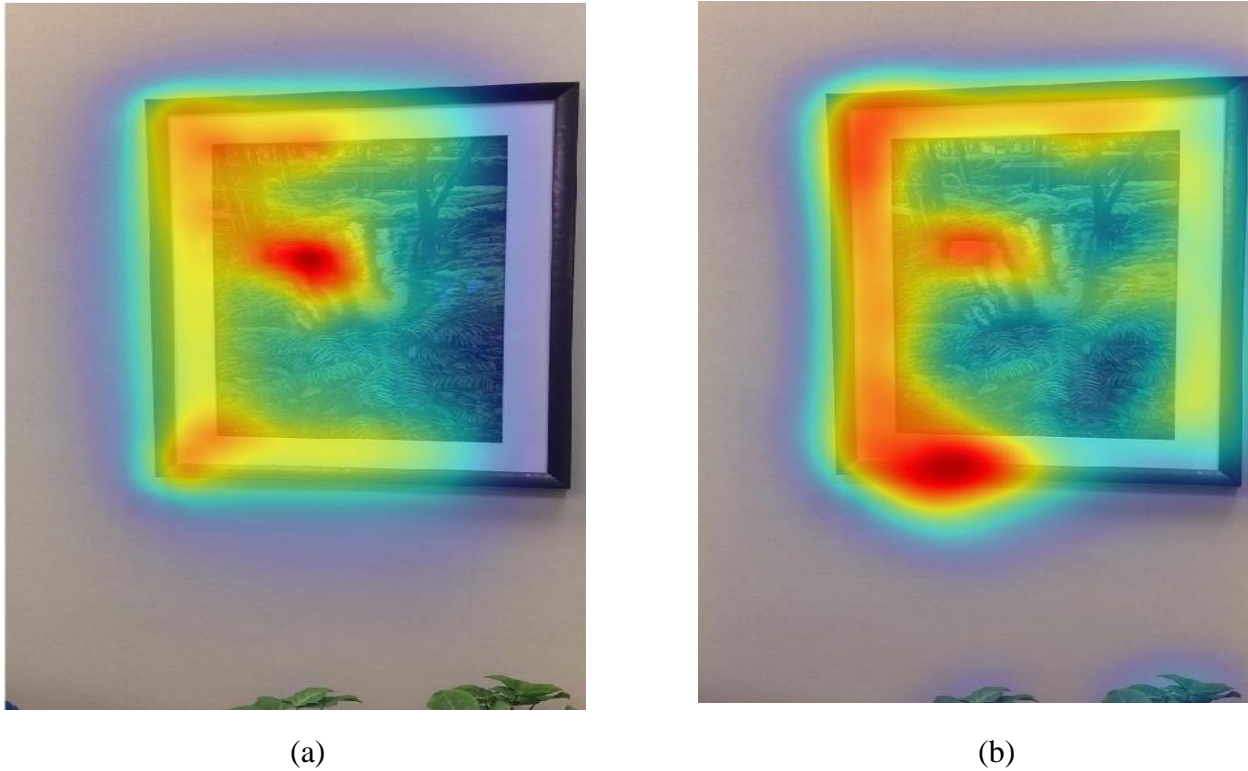
Many algorithms have been formulated to achieve saliency description in an image or video sequence. The Graph based visual saliency (GBVS) [39] and Itti, Koch, Niebur [40] algorithm determines the saliency by computing the saliency map. Saliency map is an image which has a unique representation in terms of color, intensity, and orientation.

### 2.3.1 Graph based visual saliency & Itti, Koch, Niebur Algorithm

In this the saliency map is computed by first extracting the feature vectors, forming the activation maps and normalizing the activation map followed by combining the maps to a single saliency map. Figure 2.10 (a) shows the salient region in one of the frame from the video acquired. Alternatively, Itti, Koch, Niebur algorithm is based on “feature integration” theory. In this, the input image is decomposed in a set of feature maps (color, orientation, and intensity). Then, center-surrounded differences and normalizations are computed and finally, all the maps are combined to form a master saliency map. Figure 2.10 (b), shows the salient region computed using this algorithm.

<sup>1</sup>[http://www.scholarpedia.org/article/Visual\\_salience](http://www.scholarpedia.org/article/Visual_salience)





**Figure 2.10 (a) GBVS saliency map, (b) Itti, Koch, Niebur saliency map**

Two most important factors which affect the calculated prediction precision of saliency algorithms are (1) degree of final blur (for example, the variance of a Gaussian blur kernel applied to the master map), and (2) extent of center bias. There is a difference in performance between Itti and GBVS, this was outlined in NIPS 2006 paper. As Per this paper, on a dataset of nearly 200 face images the mean ROC score for GBVS was 0.841 and 0.812 mean ROC for after optimizing for blur and center bias. The ROC score is the area under the ROC curve, which is created by plotting the true positive rate (TPR) against the false positive rate (FPR) for the eye fixation.

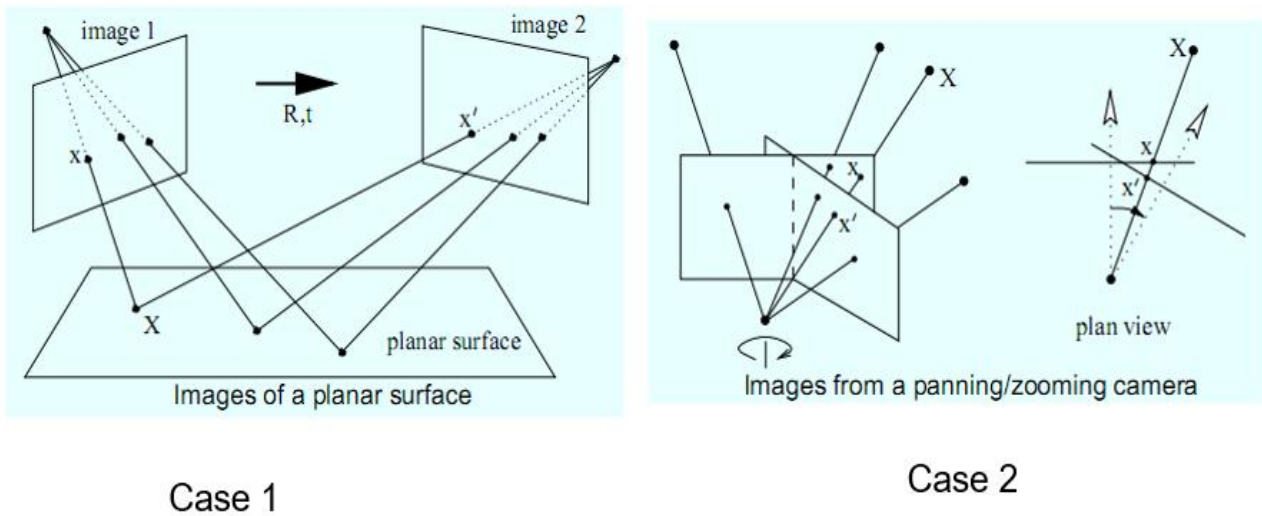
## 2.4. Planar Homography Transformation

It is a geometric transformation which has 8 degrees of freedom. There are two cases in which this transformation occurs [35].

Case 1: Images of a plane viewed under arbitrary camera motion.

Case 2: Images of an arbitrary 3D scene viewed by camera rotating and/or zooming about its optic center.

These two cases are illustrated in the Figure 2.11.



**Figure 2.11 Planar Homography transformation**

The points in a planar homography are mapped as shown in Figure 2.12. A point  $(x,y)$  is represented as  $(x,y,1)$ , its corresponding point in homogeneous coordinates is  $(x_1, x_2, x_3)$  which is the point  $(x_1/ x_3, x_2/ x_3)$

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

**Figure 2.12 Mapping of points during planar homography**

Equivalently  $X' = HX$ , where  $H$  is a  $3 \times 3$  non-singular homogeneous matrix. Where the non-homogeneous coordinates are given by Figure 2.13.

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \quad y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}$$

**Figure 2.13 Non-homogeneous coordinates**

### 2.4.1 RANSAC method for Homography estimation

In the work, we use the RANSAC method for estimating the homography matrix for all the frames. The algorithm for RANSAC method has the following steps.

Step 1. Extract features

Step 2. Compute a set of potential matches

Step 3. do

Step 3.1 select minimal sample (i.e. 7 matches)	}	(Generate hypothesis)
Step 3.2 compute solution(s) for H		
Step 3.3 determine inliers (verify hypothesis)		

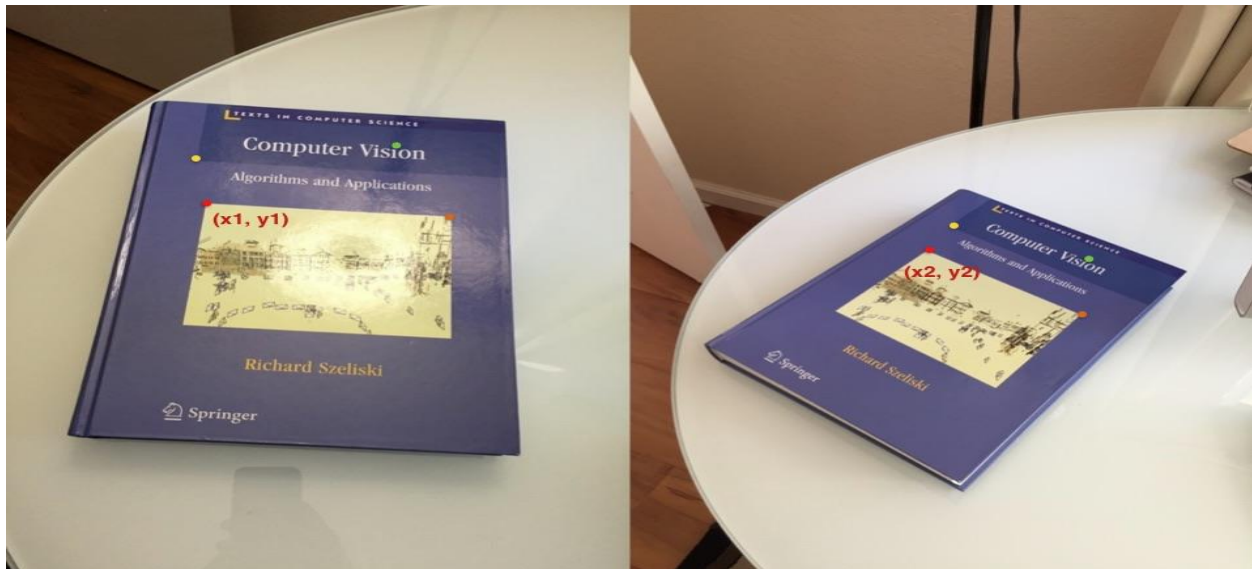
until a large enough set of the matches become inliers

Step 4. Compute H based on all inliers

Step 5. Look for additional matches

Step 6. Refine H based on all correct matches

Figure 2.14 shows an example of Projective homography transformation, where  $(x_2, y_2)$  is the transformed point for  $(x_1, y_1)$ .



**Figure 2.14 Projective transformation Example**



## 2.4.2 Limitations for Homography transformation

When warping an image to do projective transform, the smooth edges in the original images turn to jagged edges as shown in Figure 2.15.



**Figure 2.15 Jagged edges due to Projective transform**

The left side shows the original image; the right-hand side is the warped image which is prone to jagged edges. These jagged edges need to be smoothed by using appropriate filters. The best suitable filter is using a Gaussian filter. `imgaussfilt(A,sigma)` filters image A with a 2-D Gaussian smoothing kernel with standard deviation specified by sigma in MATLAB. Another method to smooth the edges is to apply an edge detector like Sobel, Canny, Prewitt, Roberts etc. using a threshold level, dilating the edges and then applying the Gaussian filtering to that image. At last the edges are replaced with the smoothed ones. The former one is fast and easy to implement in terms of complexity so, the previous one is used in this work.

### 3. Automatic ROI detection

The primary step is to automatically determine the region of interest. The region of interest is specified by using four corner coordinates in a frame. The approach is to first detect the four corner points of the region of interest in the first frame and later, using the KLT tracking algorithm to find the corner points in all the other frames.

To detect the four corner points in the first frame, Harris corner detection was applied to the binary version of the first frame.

#### 3.1 Harris Corner detection

Corners are good feature points which are more stable over the change of viewpoints. These corners can be identified by looking after the intensity values within a window. For a point to be a corner, there must be a large change in the intensity values along all the directions. Figure 3.1 shows the basic idea of a corner [38].

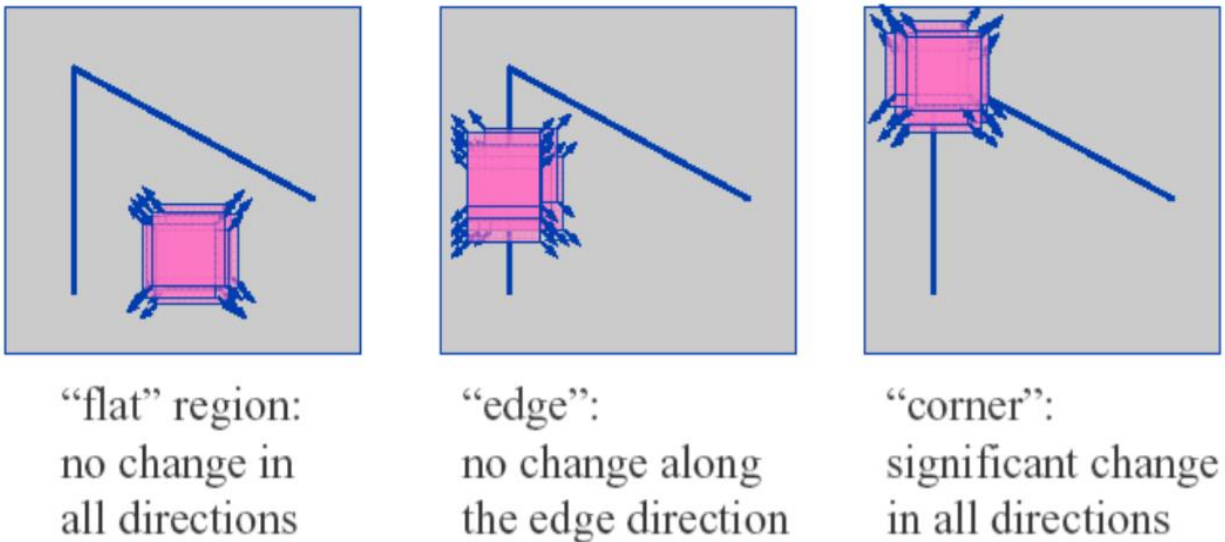


Figure 3.1 Basic idea of a “flat”, “edge” and “corner”

#### 3.1.1 Mathematical Formulation

Change of intensity for shift  $[u,v]$  is given as

$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2$$

(1)



The corner response is measured by using the equation (7).

$$R = \det(M) - k(\text{Trace } M^2)$$

$$\det(M) = \lambda_1 \lambda_2$$

$$\text{Trace } M = \lambda_1 + \lambda_2 \quad (7)$$

$k$  is an empirically determined constant;  $k=0.04-0.06$ .

Figure 3.3 is the corner response map for determining the corner point.

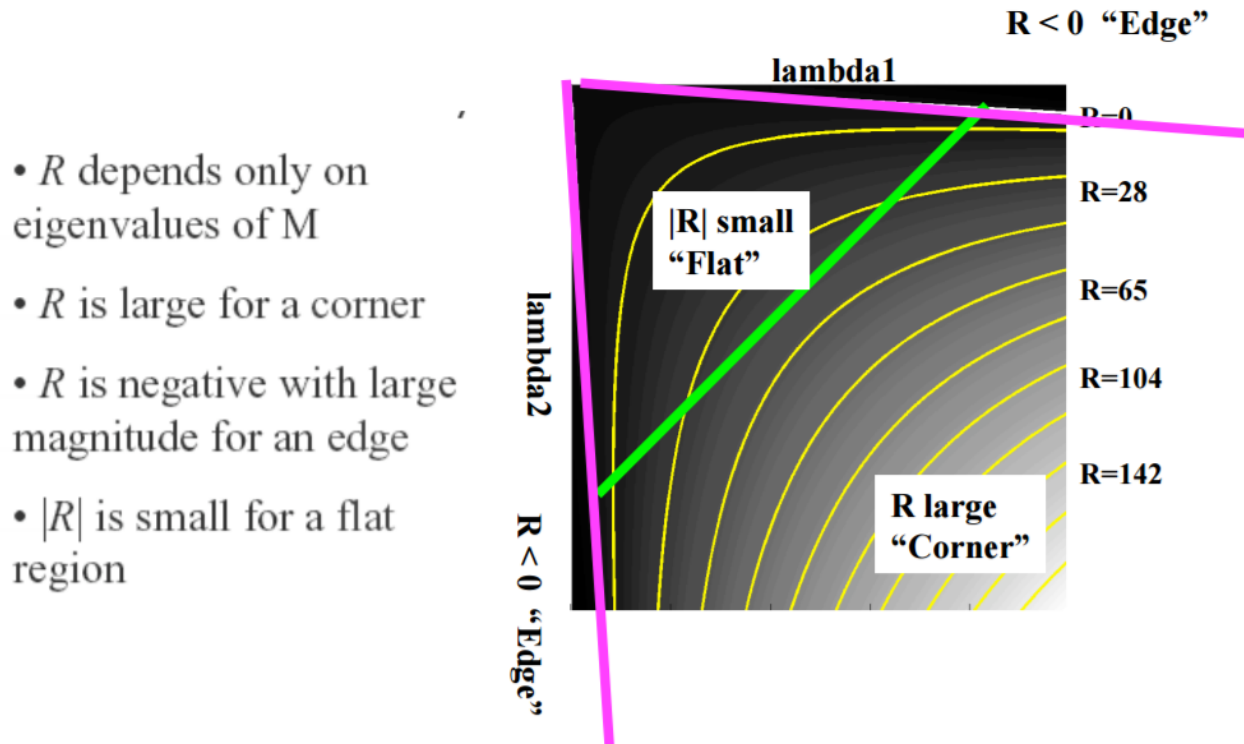


Figure 3.3 Corner Response Map

### 3.1.2 Limitations of Harris Corner Detector

Harris Corner detector detects the corners which are highly prominent. There are many erroneous corners detected. To remove the erroneous corners, there a distance constraint laid down on the detected points. For the points to be a valid corner, the distance between them is to be greater than 300. This value is determined from the fact that the size of the ROI is greater than [300, 300]. From this we get the four corners of the Region of Interest in the first frame. In order to group them in anti-clockwise order, these points are first sorted by rows and then by using the if condition, the ROI of the first frame is sorted as [(low,low); (low,high); (high,high); (high,low)].

---

**Algorithm 1:** Automatic detection of the ROI of first frame
 

---

**Input:** Video sequence

Reading the .avi file using VideoReader

Frame= Acquiring the 1<sup>st</sup> Frame from the video

I1= imbinazine(frame); % Binarize image by thresholding

I2= imcomplement(I1); % Complementing of an image

I3= Imfill (I2,'holes'); % Filling of holes, black region surrounded by white region

I4= imclearborder(k,8); % suppresses structures that are lighter than their surroundings

I5= imerode(I4, strel('diamond',1)); % eroding the image to remove the unnecessary surroundings

I6= imgaussfilt(I5); % to smoothen the Image

H= corners(I6); % Harris corner detection

N= size(H,1);

D= pdist(H,'euclidean'); %computes the Euclidean distance between pairs of objects

**while** N  $\sim$ 4 **then do**

    initialize the upper limit (ul) and lower limit (ll)

**for** i= 1: N-1

**if** distance from i<sup>th</sup> point to other points greater than 300

            valid point

            Update ll and ul

**else**

            Update ll and ul

**end**

**end**

**end**

**ROI{1,1}** = Sortrows(points); %sorts the rows of a matrix in ascending order based on the  
   % elements in the first column.

**Output:** ROI{1,1} which is the region of interest of the first frame.

---

## 3.2 Tracking using KLT Algorithm with Error Detection

The input functions required for carrying out KLT Algorithm [28,29,30,31] includes the four coordinates of region of interest for first frame, window size and the frame sequence from the input video file. The four coordinates for the region of interest in first frame are provided from the Harris Corner Detector along with the constraints. The window size should be provided in such a way that it shouldn't causes implementation issues. Smaller window size is more prone to noise and may omit larger motions while the large window size is more likely to cross the occlusion boundary. The typical values for the window size includes from 15 x 15 to 31 x 31. In our current study, window size of 15 x 15 is selected. In addition, a prior of parallelism of the two sides of the detected region of interest is used for error detection. This a valid prior since, the ROI is either in the form of rectangle or square or a parallelogram.

The video sequence has more than 250 frames, in this work only the first 250 frames are considered for computational purpose. The goal is to align the template image  $T(x)$  in the next frame to an intensity patch  $I(x)$  sub windowed by the window size in the current frame.

### 3.2.1 Template Image $T(x)$

The template image  $T(x)$  is the intensity patch sub windowed by the window size in the next frame at new tracked point corresponding to the point in the previous frame by using the three key constraints.

1. **Brightness constancy:** The point in all the frames has same brightness.
2. **Small Motion:** There exists a small motion for the points in all simultaneous frames.
3. **Spatial Coherence:** The points move like their neighbors.

Using the Brightness constancy constraint from Figure 3.4, we formulate an equation

$$I(x,y,t) = I(x+u, y+v, t+1)$$

Taking the Taylor series expansion of RHS at  $(x,y,t)$

$$I(x+u, y+v, t+1) \approx I(x,y,t) + I_x \cdot u + I_y \cdot v + I_t$$

$$I(x+u, y+v, t+1) - I(x,y,t) = I_x \cdot u + I_y \cdot v + I_t$$

$$I_x \cdot u + I_y \cdot v + I_t \approx 0$$

$$\nabla I \cdot [u' \ v']^T + I_t = 0 \tag{8}$$

This Equation (8) has two unknowns to be solved, so a window is assumed to obtain more equations for a single point. If a 5 x 5 window is selected, then there exist 25 equations for a single point as in Equation 9.

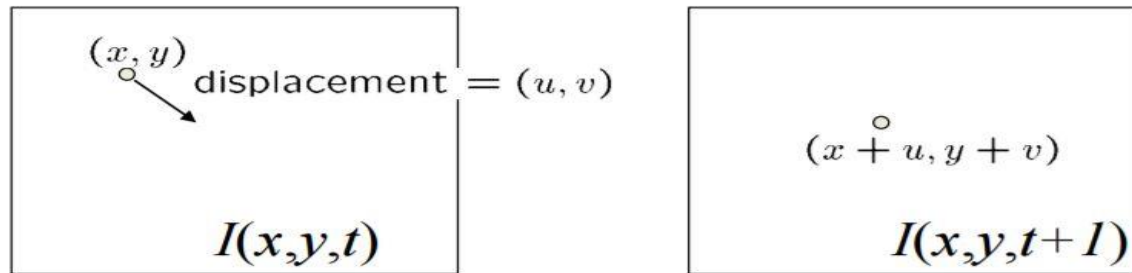


Figure 3.4 Tracking a point  $(x,y)$ <sup>1</sup>

$$\begin{pmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_1) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = - \begin{pmatrix} I_t(p_1) \\ I_t(p_2) \\ \vdots \\ I_t(p_{25}) \end{pmatrix} \quad A_{25 \times 2} \mathbf{d}_{2 \times 1} = B_{25 \times 1} \quad (9)$$

From (9) the matrix  $\mathbf{d}$  can be found by multiplying both sides with  $A^T$  to make the A matrix to be invertible.  $(A^T A) \mathbf{d} = A^T B$ . Now  $T(x)$  is the intensity patch around the point  $(x+u, y+v)$ . In order to find the exact point  $T(x)$  should be aligned with  $I(x)$ . The best alignment minimizes (10).

$$\min \sum_x [I(W(x;p)) - T(x)]^2 \quad (10)$$






Where  $W(x;p)$  is a warp function. Distinct 2-D transformations have distinct warp functions. Figure 3.5 shows distinct warp functions for different transformations.

### 3.2.2 Optimization problem and Solution Algorithm

Equation (10) is a nonlinear optimization, the warp  $W(x;p)$  has non-linear pixel values. It is assumed that some  $p$  is known and the best increment  $\Delta p$  is found. Equation (10) is modified as

$$\min \sum_x [I(W(x;p + \Delta p)) - T(x)]^2 \quad (11)$$

<sup>1</sup><https://courses.engr.illinois.edu/cs543/sp2012>

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} I & t \end{bmatrix}_{2 \times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} R & t \end{bmatrix}_{2 \times 3}$	3	lengths	
similarity	$\begin{bmatrix} sR & t \end{bmatrix}_{2 \times 3}$	4	angles	
affine	$\begin{bmatrix} A \end{bmatrix}_{2 \times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{H} \end{bmatrix}_{3 \times 3}$	8	straight lines	

$$W(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} 1 & 0 & b_1 \\ 0 & 1 & b_2 \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$W(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} c\theta & -s\theta & b_1 \\ s\theta & c\theta & b_2 \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$W(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} sco\theta & -ssi\theta & b_1 \\ ssi\theta & sco\theta & b_2 \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$W(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} a_1 & a_2 & b_1 \\ a_3 & a & b_2 \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

**Figure 3.5 Distinct Warp functions for different transformations [33]**

The Equation (11) is to be solved for automatically selecting the region of interest in each video. By performing the first order Taylor series Expansion,

$$\min \sum_x [I(W(\mathbf{x}; \mathbf{p})) + \nabla I \frac{\partial W}{\partial \mathbf{p}} \Delta \mathbf{p} - T(\mathbf{x})]^2 \quad (12)$$

Where,  $\nabla I$  is the gradient image computed at  $W(\mathbf{x}; \mathbf{p})$  and  $\frac{\partial W}{\partial \mathbf{p}}$  is the Jacobian of the Warp. To find the minimum of the equation (12), the derivative of the above equation with respect to  $\Delta \mathbf{p}$  is made equal to zero. Then the value of  $\Delta \mathbf{p}$  is given by

$$\Delta \mathbf{p} = \mathbf{H}^{-1} \sum_x [\nabla I \frac{\partial W}{\partial \mathbf{p}}]^T [T(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p}))],$$

$$\mathbf{H} = \sum_x [\nabla I \frac{\partial W}{\partial \mathbf{p}}]^T [\nabla I \frac{\partial W}{\partial \mathbf{p}}] \quad (13)$$

The algorithm for solving the equation (12) is given below,  $\mathbf{p}$  value is the ROI in next frame. As a result we get the region of interest for the 250 frames automatically. Figure 3.6 shows the Algorithm 2 in pictorial view [34].



---

**Algorithm 2:** ROI Tracking (Specularity robust)
 

---

**Input:** Four coordinates of ROI for first frame, window size  $15 \times 15$ , and the 250 frames from the input video file.

```

while not  $\|\Delta p\| \leq 0.0001$  do
  compute  $I(W(x;p))$ , by warping  $I$  with  $W(x;p)$ 
  compute  $\nabla I$  with  $W(x;p)$ 
  compute  $\frac{\partial W}{\partial p}$  at  $(x;p)$ 
  compute  $\nabla I \frac{\partial W}{\partial p}$ 
  compute  $H = \sum_x [\nabla I \frac{\partial W}{\partial p}]^T [\nabla I \frac{\partial W}{\partial p}]$ 
  compute  $\Delta p = H^{-1} \sum_x [\nabla I \frac{\partial W}{\partial p}]^T [T(x) - I(W(x;p))]$ 
  Update  $p$  to  $p + \Delta p$ 

```

**end**

finding the slope of two lines  $m1$  and  $m2$  %error detection

**if**  $m1 \sim m2$

  | sides are not parallel; stop;

**end**

**Output:**  $p$  value, which is the ROI in next frame.

---

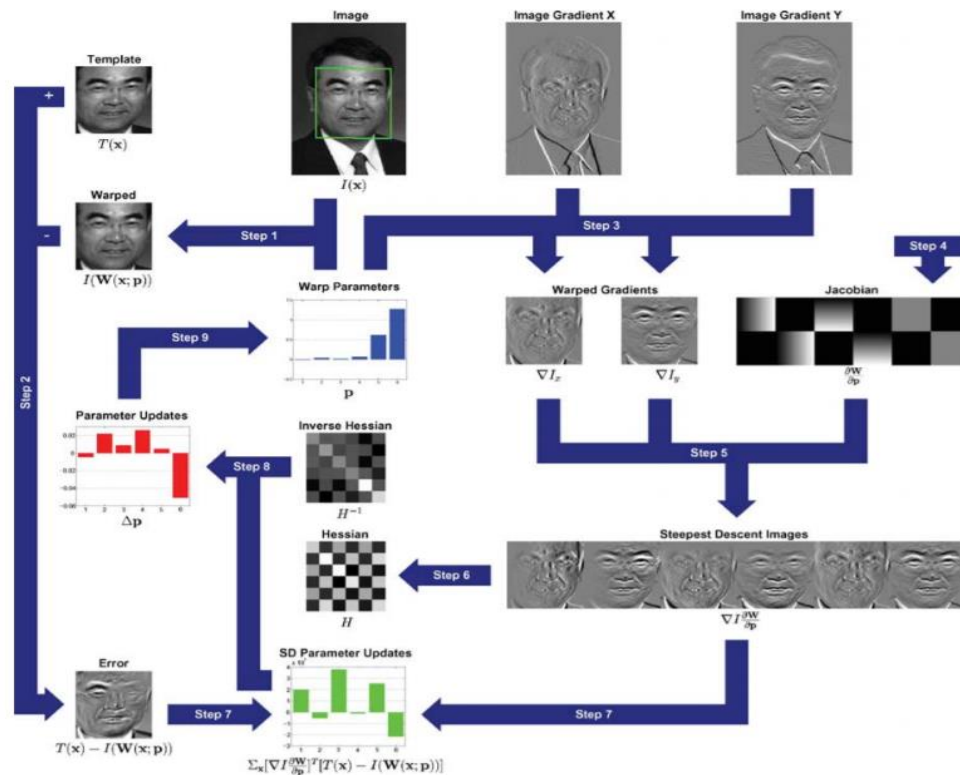


Figure 3.6 Illustration of Algorithm 2

## 4. Specularity Removal Algorithm

### 4.1. Problem Formulation

By deep analysis using various video sequences it is explored that only the Y-channel contains information about the specularity, the Chroma channels (UV) has no information regarding the specularity. So, the RGB video sequence is converted to full-range YUV 4:2:2 video. The region of interest from Algorithm (2) in the Y-channel for  $i^{\text{th}}$  frame can be expressed as the sum of transmission layer ( $T_i$ ) and error ( $E_i$ ), where  $E_i$  is due the effect of specularity which are sparse. But, the ROI in all the frames are misaligned. Then there exists 2D Homography transformation matrix  $\Gamma = \{\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_n\}$  to align the region of interest for all the frames.

$$Y \circ \Gamma = T + E + N \quad (14)$$

Where, the columns of  $Y \circ \Gamma \in \mathbb{R}^{m \times n}$  are vectorized images and  $m$  is the number of pixels,  $N$  is due to the jagged edges by homography transformation. To find the  $T$  and  $E$  for all the frames we have two unknowns in every equation which is ill posed problem [24]. To solve this problem structural priors are applied.

#### 4.1.1. First Structural Prior

The transmission layer for every frame is highly correlated except there are small changes in illumination. This suggests that  $\text{rank}(T)$  becomes as small as possible.

$$\min \text{rank}(T) \text{ s.t. } Y \circ \Gamma = T + E + N, \|E+N\|_0 \leq k \quad (15)$$

#### 4.1.2. Second Structural Prior

The gradient field of both transmitted layer and the error are sparse, there also exists independence between the above gradients. This can be expressed as

$$\begin{aligned} \|DT\|_0 &= \sum_{i=1}^n \sum_{j=1}^2 \|dj * Ti\|_0 \\ \|DE\|_0 &= \sum_{i=1}^n \sum_{j=1}^2 \|dj * Ei\|_0 \end{aligned} \quad (16)$$

Where,  $(d_1, d_2, d_3, \dots, d_j)$  are derivative filters.

Combining, the equations (15) and (16), the following objective function should be minimized.

$$\min \text{rank}(T) + \lambda_1 \|E+N\|_0 + \lambda_2 \|N\|_F^2 + \lambda_3 \|DT\|_0 + \lambda_4 \|DE\|_0 + \lambda_5 \|DT \odot DE\|_0 + \lambda_6 \| \Omega - DT - DE \|_F^2,$$

$$\text{s.t. } Y \circ \Gamma = T + E + N; T \geq 0; E \geq 0;$$

(17)

Where  $\| \cdot \|_0$  and  $\| \cdot \|_F$  denotes  $l^0$  norm and Frobenius norm respectively, the third term is for the Gaussian noise,  $\odot$  stands for element wise multiplication,  $\Omega = DY$  (gradient of  $Y$ ). The constraints are since the values in the transmission layer and the error are non-negative.

## 4.2. Optimization Problem

Equation (17) faces difficulty of non-convexity of the matrix rank,  $l^0$ -norm and the non-linearity of the constraint  $Y \circ \Gamma = T + E + N$ , due to its reliance on homography transformations  $\Gamma = \{\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_n\}$ . Through convex relaxation, the rank function and  $l^0$  norm is replaced with nuclear norm  $\| \cdot \|_*$  and  $l^1$  norm respectively and to linearize the above-mentioned constraint we modify the constraint as

$$Y \circ (\Gamma + \Delta \Gamma) \approx Y \circ \Gamma + \sum_{i=1}^n J_i \Delta \Gamma \varepsilon_i \varepsilon_i^T = T + E + N,$$

(18)

Where  $J_i = \frac{\partial}{\partial \zeta} \text{vec}(I_i \circ C) |_{C=\Gamma_i} \in \mathbb{R}^{m \times p}$  is the Jacobian of the  $i^{\text{th}}$  image with respect to  $\Gamma_i$ ,  $\Delta \Gamma \in \mathbb{R}^{p \times n}$  and  $\{\varepsilon_i\}$  denotes the standard basis for  $\mathbb{R}^n$ . By applying the above convex relaxation, Equation (17) can be modified as

$$\min \|T\|_* + \lambda_1 \|E+N\|_1 + \lambda_2 \|N\|_F^2 + \lambda_3 \|DT\|_1 + \lambda_4 \|DE\|_1 + \lambda_5 \|DT \odot DE\|_1 + \lambda_6 \| \Omega - DT - DE \|_F^2,$$

$$\text{s.t. } Y \circ \Gamma + \sum_{i=1}^n J_i \Delta \Gamma \varepsilon_i \varepsilon_i^T = T + E + N; T \geq 0; E \geq 0;$$

(19)

There are 250 frames on which we can apply the algorithms, to decrease the computational complexity and computational time there is a need to find the optimum number of frames to work upon.

### 4.3. Solution Algorithm

#### 4.3.1 Solution Algorithm for Specularity removal

The Augmented Lagrange Multiplier (ALM) with Alternating Direction Minimizing (ADM) technique [32] is used to solve the Equation (19). The auxiliary variables are introduced to make the equation (19) separable. Accordingly,  $L = T$ ,  $M = E + N$ ,  $P = DT$ ,  $Q = DE$  act as additional constraints.

The Augmented Lagrangian function for equation (19) is

$$\left\{ \begin{array}{l} \mathcal{L}_{T \geq 0; E \geq 0; (L, M, N, P, Q, T, E, \Delta \Gamma)} \\ = \|L\|_* + \lambda_1 \|M\|_1 + \lambda_2 \|N\|_F^2 + \lambda_3 \|P\|_1 + \lambda_4 \|Q\|_1 \\ \quad + \lambda_5 \|P \odot Q\|_1 + \lambda_6 \|\Omega - P - Q\|_F^2 \\ \quad + \Phi(Z_1, Y \circ \Gamma + \sum_{i=1}^n J_i \Delta \Gamma \varepsilon_i \varepsilon_i^T - T - M) \\ \quad + \Phi(Z_2, M - E - N) + \Phi(Z_3, L - T) + \Phi(Z_4, P - DT) \\ \quad \quad \quad + \Phi(Z_5, Q - DE) \end{array} \right. \quad (20)$$

Where,  $Z_1, Z_2, Z_3, Z_4, Z_5$  are the Lagrangian multipliers,  $\Phi(Z, K) = \frac{\mu}{2} \|K\|_F^2 + \langle Z, K \rangle$  (where  $\langle \cdot, \cdot \rangle$  constitutes matrix inner product,  $\mu$  is a positive scalar). The eight variables need to be solved. The solutions for them are

$$\mathbf{L}^{t+1} = \underset{L}{\operatorname{argmin}} \|L\|_* + \Phi(Z_3^t, L - T^t) = US \frac{1}{\mu^t} [\Sigma] V^T \quad (21)$$

$$\begin{aligned} \mathbf{M}^{t+1} &= \underset{M}{\operatorname{argmin}} \lambda_1 \|M\|_1 + \Phi(Z_1^t, B^t - T^t - M) + \Phi(Z_2^t, M - E^t - N^t) \\ &= \frac{S_{2\lambda_1}}{\mu^t} [B^t - T^t + E^t + N^t + \frac{Z_1^t - Z_2^t}{\mu^t}]. \end{aligned} \quad (22)$$

$$\mathbf{N}^{t+1} = \underset{N}{\operatorname{argmin}} \lambda_2 \|N\|_F^2 + \Phi(Z_2^t, M^{t+1} - E^t - N) = \frac{Z_2^t + \mu^t (M^{t+1} - E^t)}{2\lambda_2 + \mu^t} \quad (23)$$

$$\begin{aligned}
\mathbf{P}^{t+1} &= \underset{P}{\operatorname{argmin}} \lambda_3 \|\mathbf{P}\|_1 + \lambda_5 \|\mathbf{P} \odot \mathbf{Q}^t\|_1 + \lambda_6 \|\Omega - P - \mathbf{Q}^t\|_F^2 + \Phi(Z_4^t, \mathbf{P} - \mathbf{DT}^t) \\
&= \hat{S}_{\frac{\lambda_3 + \lambda_5 |\mathbf{Q}^t|}{2\lambda_6 + \mu^t}} \left[ \frac{\lambda_6 (\Omega - \mathbf{Q}^t) + \frac{\mu^t \mathbf{DT}^t - Z_4^t}{2}}{\lambda_6 + \frac{\mu^t}{2}} \right]
\end{aligned} \tag{24}$$

$$\begin{aligned}
\mathbf{Q}^{t+1} &= \underset{Q}{\operatorname{argmin}} \lambda_4 \|\mathbf{Q}\|_1 + \lambda_5 \|\mathbf{P}^{t+1} \odot \mathbf{Q}\|_1 + \lambda_6 \|\Omega - \mathbf{P}^{t+1} - \mathbf{Q}\|_F^2 + \Phi(Z_5^t, \mathbf{Q} - \mathbf{DE}^t) \\
&= \hat{S}_{\frac{\lambda_4 + \lambda_5 |\mathbf{P}^{t+1}|}{2\lambda_6 + \mu^t}} \left[ \frac{\lambda_6 (\Omega - \mathbf{P}^{t+1}) + \frac{\mu^t \mathbf{DE}^t - Z_5^t}{2}}{\lambda_6 + \frac{\mu^t}{2}} \right]
\end{aligned} \tag{25}$$

Where  $U\Sigma V^T$  is SVD of  $(\mathbf{T}^t - \frac{Z_3^t}{\mu^t})$ ,  $[\mu^t]$  is monotonically increasing positive function,  $S_\varepsilon[x] = \operatorname{sgn}(x)\max(|x| - \varepsilon, 0)$ ,  $\hat{S}_w[x]$  performs the shrinkage on the elements in  $X$  with respect to threshold entries of  $w$  and  $\mathbf{B}^t = \mathbf{Y} \circ \Gamma + \sum_{i=1}^n J_i \Delta \Gamma \varepsilon_i \varepsilon_i^T$

$$\mathbf{T}^{t+1} = \underset{T}{\operatorname{argmin}} \Phi(Z_1^t, \mathbf{B}^t - \mathbf{M}^{t+1} - \mathbf{T}) + \Phi(Z_3^t, \mathbf{L}^{t+1} - \mathbf{T}) + \Phi(Z_4^t, \mathbf{P}^{t+1} - \mathbf{DT}),$$

this can be solved by applying the circular boundary conditions and 2D FFT,

$$\mathbf{T}_i^{t+1} = \mathcal{F}^{-1} (\mathcal{F}(\mathcal{R}(O_i)) / (\overline{\mathcal{F}(\mathbf{D})} \odot \mathcal{F}(\mathbf{D}) + 2)), \text{ for } i \in [1, 2, \dots, n]$$

$$\text{and } \mathbf{O} = \mathbf{B}^t - \mathbf{M}^{t+1} + \mathbf{L}^{t+1} + \frac{Z_1^t + Z_3^t}{\mu^t} + \mathbf{D}^T (\mathbf{P}^{t+1} + \frac{Z_4^t}{\mu^t})$$

(26)

$$\mathbf{E}^{t+1} = \underset{E}{\operatorname{argmin}} \Phi(Z_2^t, \mathbf{M}^{t+1} - \mathbf{N}^{t+1} - \mathbf{E}) + \Phi(Z_5^t, \mathbf{Q}^{t+1} - \mathbf{DE}),$$

$$\mathbf{E}_i^{t+1} = \mathcal{F}^{-1} (\mathcal{F}(\mathcal{R}(G_i)) / (\overline{\mathcal{F}(\mathbf{D})} \odot \mathcal{F}(\mathbf{D}) + 1)), \text{ for } i \in [1, 2, \dots, n]$$

$$\text{and } \mathbf{G} = \mathbf{M}^{t+1} - \mathbf{N}^{t+1} + \frac{Z_2^t}{\mu^t} + \mathbf{D}^T (\mathbf{Q}^{t+1} + \frac{Z_5^{t+1}}{\mu^t})$$

(27)

where  $\mathcal{F}(\cdot)$  is 2D FFT,  $\mathcal{F}^{-1}$  is the 2D inverse FFT,  $\overline{\mathcal{F}(\cdot)}$  is complex conjugate of  $\mathcal{F}(\cdot)$ ,  $\mathcal{R}(\cdot)$  is to reshape the vectorized 2D information back to its 2D form, for  $T \geq 0$  and  $E \geq 0$  the negative values are set to zero.

---

**Algorithm 3: Specularity Removal**


---

**Input:**  $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0, \lambda_4 > 0, \lambda_5 > 0, \lambda_6 > 0$ , Y, ROI for 250 frames and transformations  $\Gamma$  using RANSAC.

Converting the Video to image sequence and choosing the first 250 frames

Selecting the 15 frames to apply the Specularity removal method (1:17:250)

**while not converged do**

$L^0 = M^0 = N^0 = T^0 = E^0 = Z_1^0 = Z_2^0 = Z_3^0 = 0 \in \mathbb{R}^{m \times n}, \Delta\Gamma^t = 0,$

$t=0, \mu^0 > 0, \rho > 1, P^0 = Q^0 = Z_4^0 = Z_5^0 = 0 \in \mathbb{R}^{2m \times n}.$

Compute the warped areas  $Y \circ \Gamma$  and their Jacobians

$J_i = \frac{\partial}{\partial \Gamma_i} (Y \circ \Gamma_i).$

**while not converged do**

Update  $L^{t+1}$  via Eq. (21);

Update  $M^{t+1}$  via Eq. (22);

Update  $N^{t+1}$  via Eq. (23);

Update  $P^{t+1}$  via Eq. (24);

Update  $Q^{t+1}$  via Eq. (25);

**for i from 1 to 15 do**

Update  $T_i^{t+1}$  via Eq. (26);

Update  $E_i^{t+1}$  via Eq. (27);

**end**

$T_i^{t+1} (T_i^{t+1} < 0) = 0; E_i^{t+1} (E_i^{t+1} < 0) = 0;$

Update  $\Delta\Gamma^{t+1}$  via Eq. (28);

Update  $Z_1^{t+1}, Z_2^{t+1}, Z_3^{t+1}, Z_4^{t+1}, Z_5^{t+1}$  via Eq. (29);

$\mu^{t+1} = \mu^t \rho; t = t+1;$

**end**

$\Gamma = \Gamma + \Delta\Gamma^t;$

**end**

Saving the 15 output images and replacing it by the 15 output images ( $I_{af}$ ).

**for i from 1 to 250 do**

masking out the region of interest from the present frame, the ROI is black with its background undisturbed ( $I_1$ ).

An image with black background and the ROI with  $I_{af}$  ( $I_2$ )

Computing  $I_3 = I_2 \circ \Gamma_i$

Final image  $I_4 = I_1 + I_3$

**end**

Combining the Y, U and V channels and converting the image sequences to RGB Video.

**Output:** Optimal solution ( $T^* = T^t, E^* = E^t$ ) and Video after Specularity removal

---

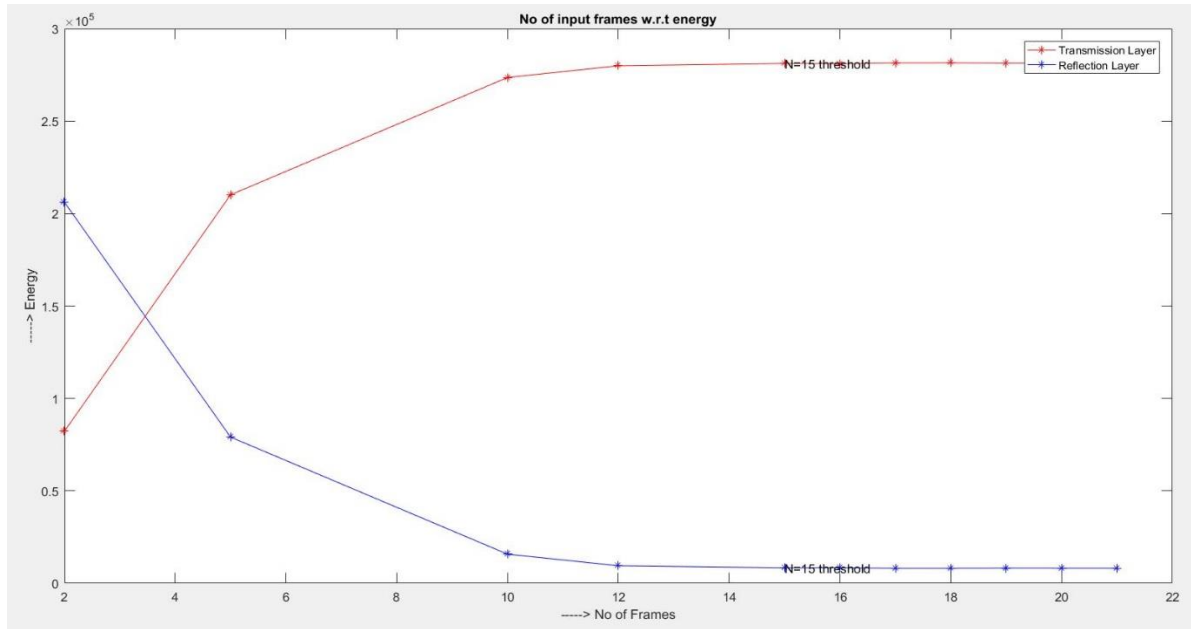
$$\begin{aligned}
\Delta\Gamma^{t+1} &= \underset{\Delta\Gamma}{\operatorname{argmin}} \Phi(Z_1^t, Y \circ \Gamma + \sum_{i=1}^n J_i \Delta\Gamma \varepsilon_i \varepsilon_i^T - T^{t+1} - M^{t+1}) \\
&= \sum_{i=1}^n J_i^\dagger (T^{t+1} + M^{t+1} - Y \circ \Gamma - \frac{Z_1^t}{\mu^t}) \varepsilon_i \varepsilon_i^T
\end{aligned} \tag{28}$$

where  $J_i^\dagger$  represents the Moore-Penrose pseudoinverse of  $J$ .

$$\begin{aligned}
Z_1^{t+1} &= Z_1^t + \mu^t (B^{t+1} - T^{t+1} - M^{t+1}) \\
Z_2^{t+1} &= Z_2^t + \mu^t (M^{t+1} - E^{t+1} - N^{t+1}) \\
Z_3^{t+1} &= Z_3^t + \mu^t (L^{t+1} - T^{t+1}) \\
Z_4^{t+1} &= Z_4^t + \mu^t (P^{t+1} - DT^{t+1}) \\
Z_5^{t+1} &= Z_5^t + \mu^t (Q^{t+1} - DE^{t+1})
\end{aligned} \tag{29}$$

The inner loop of Algorithm 3 ends when  $\|B^{t+1} - T^{t+1} - M^{t+1}\|_F \leq \delta \|Y \circ \Gamma\|_F$  with  $\delta = 10^{-6}$  or the maximum number of iterations are reached. The outer loop terminates when the change of the objective function value is less or maximum number of iterations are reached.

### 4.3.2 Determining the optimum number of Frames



**Figure 4.1 Plot of number of input frames with respect to energy in transmission layer and reflection layer**

To determine the optimum number of frames in the specular removal process, a graph between the number of input frames with respect to energy in transmission layer and reflection layer is plotted as shown in Figure 4.1. For the specular to be removed more accurately, the energy in the transmission layer should be very high and that of reflection layer should be very low. From the Figure 4.1  $N=15$  is the threshold point, the values for the energy in the transmission layer and reflection layer are nearly equal. So, 15 frames are chosen for in the specular removal process.

#### 4.4. Final image to video sequence conversion

This is the last step in the flowchart, the input to this stage are the 250 images free from specular obtained from Algorithm 3. These 250 images are converted to video sequence using Algorithm 4.

---

#### Algorithm 4: Images to Video sequence

---

**Input:** 250 frames output from Algorithm 3

Creating an object for VideoWriter and specifying its frame rate = 30

Setting the time for each image = 1 sec

Opening the VideoWriter

**for** i= 1 to 250 **do**

    convert image to frame

    write this frame to a video

**end**

close the writer object

play the video sequence

**Output:** A video sequence with the name specified in the creation step

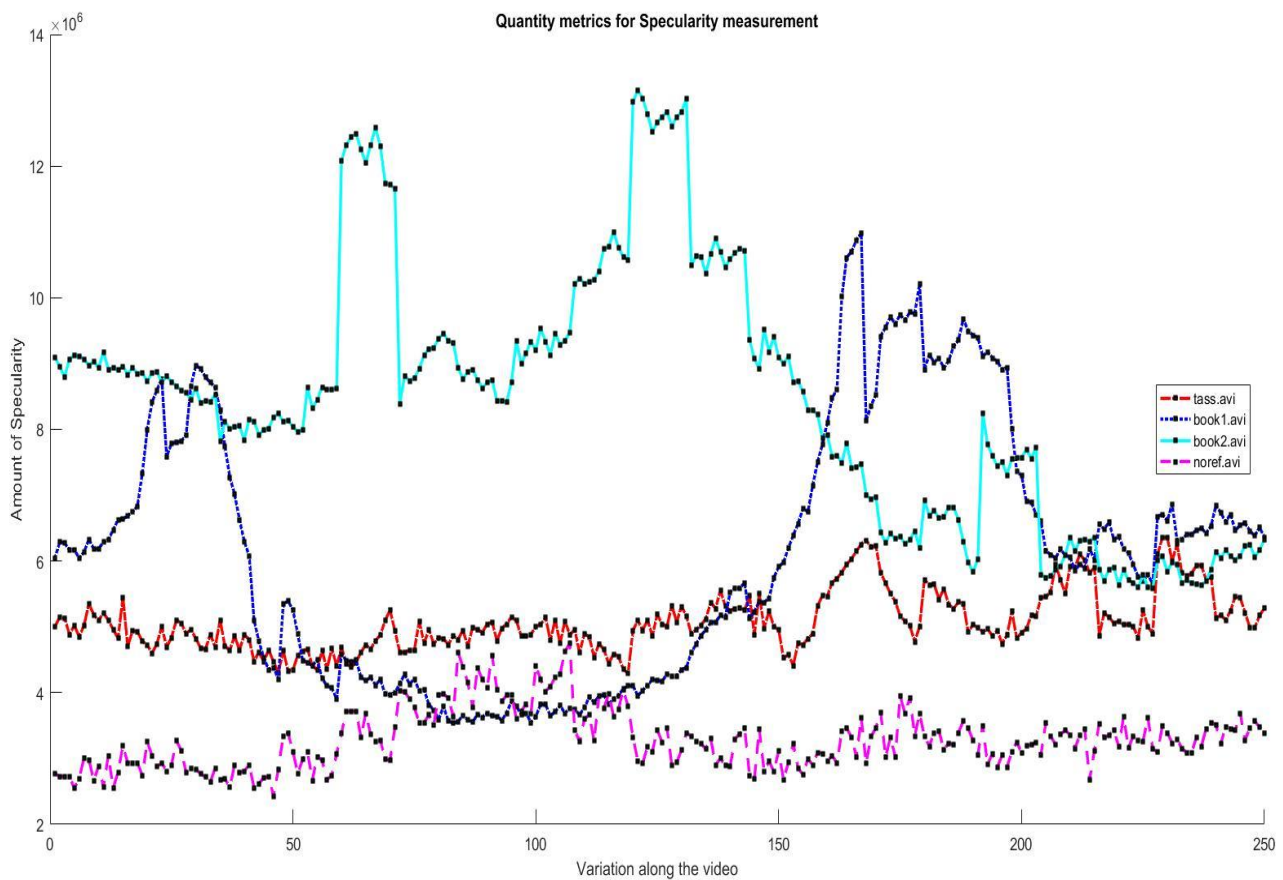
---



## 5. Quantity Metrics

The amount of specularity in a video can be determined by employing the video before specularity removal and video obtained after specularity removal. The quantity metrics is computed frame by frame basis. The difference between the frames from the two videos is taken which in turn is the specularity component in each frame. Then, the energy of the difference frame is calculated by adding the intensity value at each location.

Figure 5.1 shows the amount of specularity for all the acquired videos on frame by frame basis.



**Figure 5.1 Quantity metrics for Specularity Measurement Frame-by-Frame basis**

The mean value for all the frames is taken to have a single value for a video. Figure 5.2 shows the amount of specularity for the whole video.

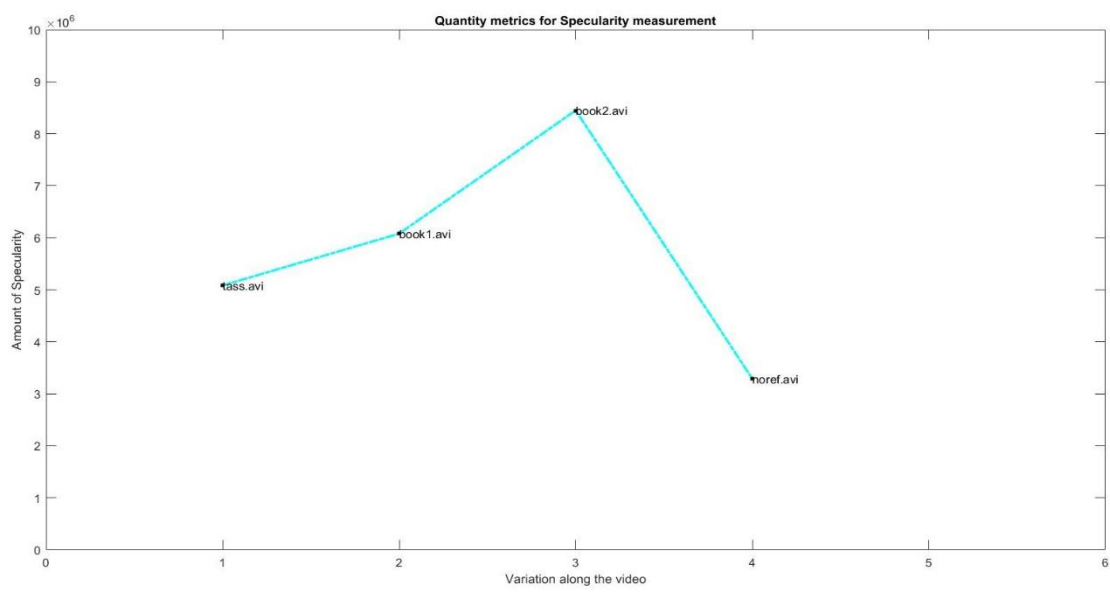


Figure 5.2 Quantity metrics for Specularity measurement

To understand deeply about the amount of specularity in each frame, a simultaneous frame and graph GUI is developed. Figure 5.3 shows the developed GUI.

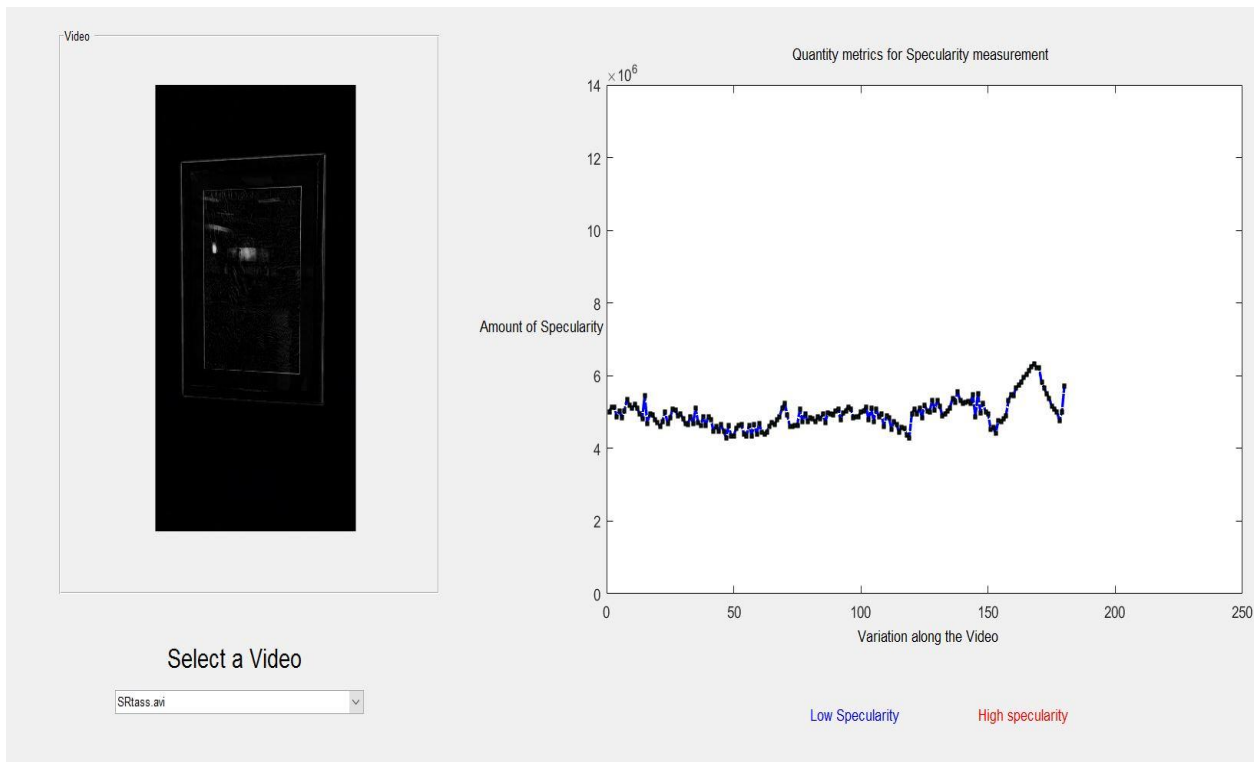
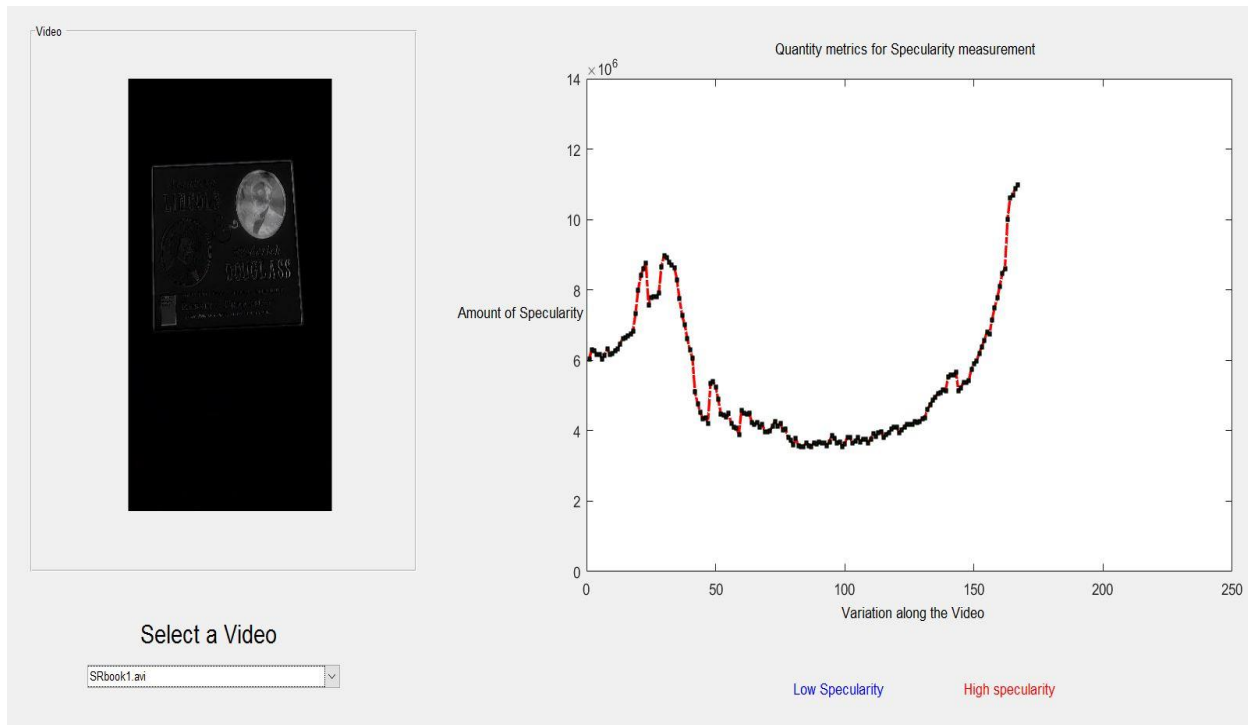


Figure 5.3 GUI for Specularity measurement (Low Specularity)

The GUI has a panel for the display of the video frames, Pop-up menu to select the video acquired from the hand-held devices and axes to display the graph representing the amount of specularity at each frame. The color of the graph depicts the amount of specularity present in the frame. The blue color of the graph represents the frames of low specularity and red color of the graph represents the frames of high specularity. Figure 5.4 shows the frame having high specularity.



**Figure 5.4 GUI for Specularity measurement (High Specularity)**

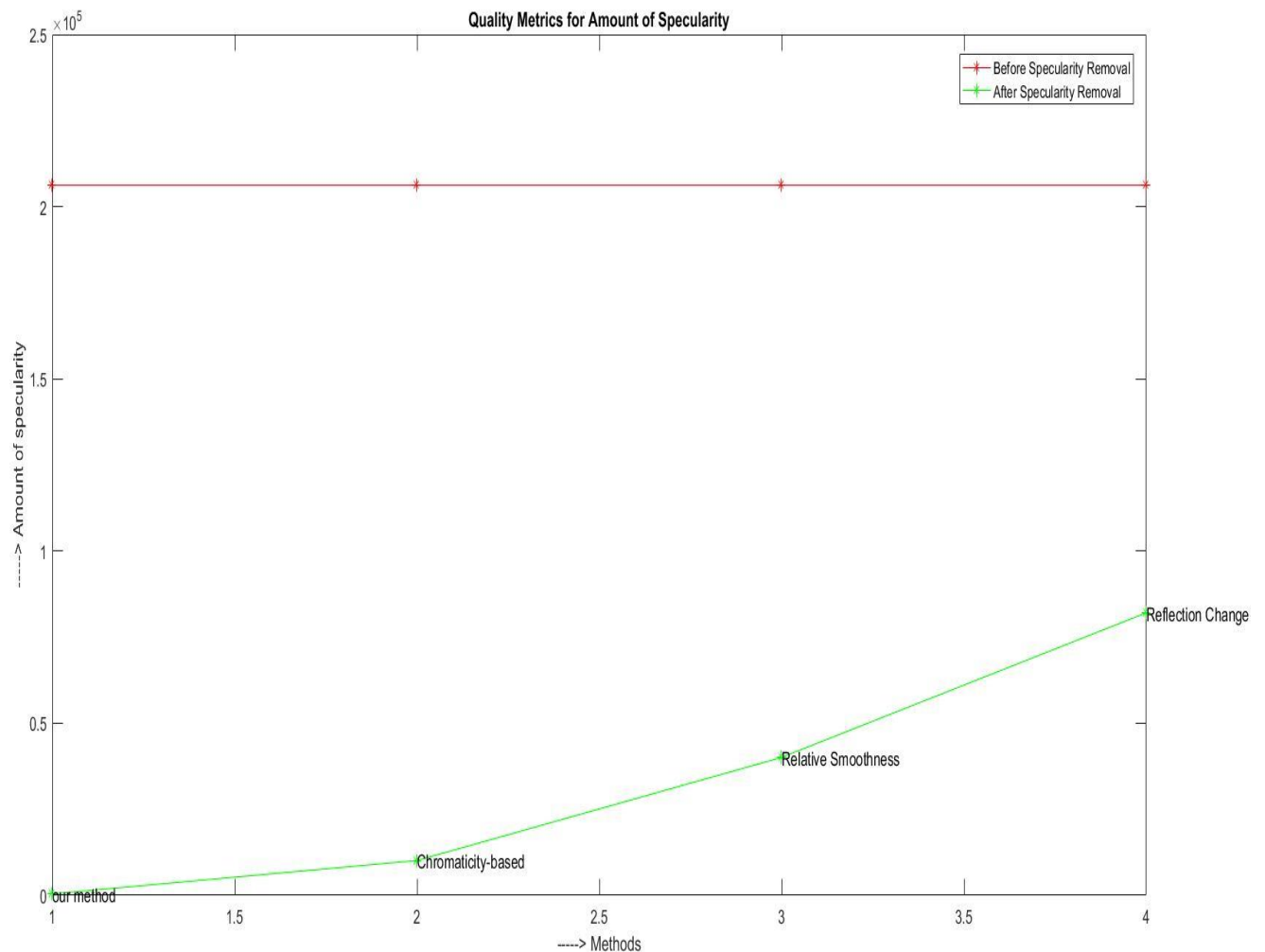
### 5.1. Quality evaluation metrics

The amount of specularity can be determined by the energy in the reflection layer, which is the sum of the intensity values at each pixel. Since, there are no papers on specularity removal from the video sequence, to compare the various algorithms, the quality evaluation is performed on the images. The color image can be decomposed into transmission layer and reflection layer. The reflection layer can be derived from the original image by subtracting the transmission layer from it.

For evaluation purpose, three different methods along with our method is compared. These methods are

1. Chromaticity-based separation of reflection components in a single image [21]
2. Single Image Layer Separation using Relative Smoothness [36]
3. Exploiting Reflection Change for Automatic Reflection Removal [37]

The amount of specularity before and after specularity removal for all the above-mentioned methods are plotted. This plot reveals that the current method is superior to other methods in terms of performance.



**Figure 5.5 Quality Metrics for Amount of Specularity**

---

**Algorithm 5:** Video Quantity Evaluation (Specularity)

---

**Input:** Original Video acquired and Video after specularity removal

n = Calculating min no. of frames from the two videos

Opening the VideoWriter

**for** i= 1 to n **do**

    convert Videos to frame (f1 and f2)

    I= f1-f2 (difference between the frames, to find the specularity)

    write this frame to a video

    s = sum (I(:)); % finding the sum of intensity levels at each pixel position

**end**

close the writer object

play the video sequence

**Output:** Specular component of the acquired video sequence, frame by frame evaluation of the amount of specularity.

---

## 6. Experimental Results

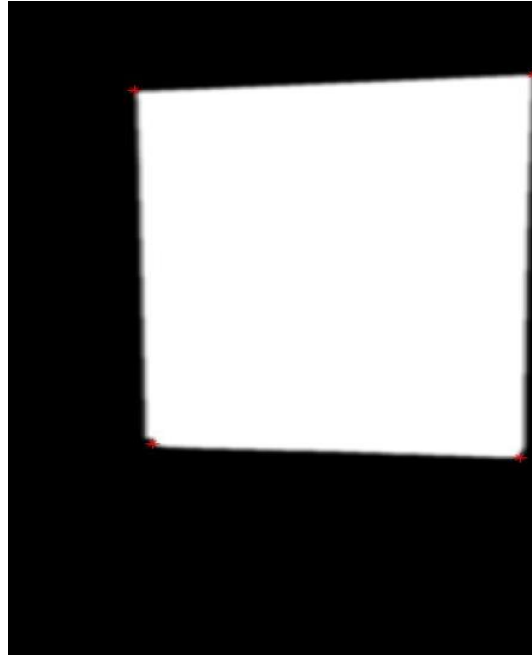
This Chapter covers the experimental results obtained from Chapter 3, Chapter 4 and implementation details of this work. This method is evaluated on high-quality videos. These videos are acquired from iPhone 6s plus and Samsung galaxy S6 edge. These videos are in .MOV and .AVI format.

### 6.1. Results of Automatic detection of ROI

The Automatic detection of the ROI was performed on the first frame. The output from this algorithm is the four corner points of the region of interest. Figure 6.1 displays output from this algorithm. The output is in the form of a mask, where the white region is the salient region (Region of interest). The four corners are marked with red “+” (+).



(a)



(b)

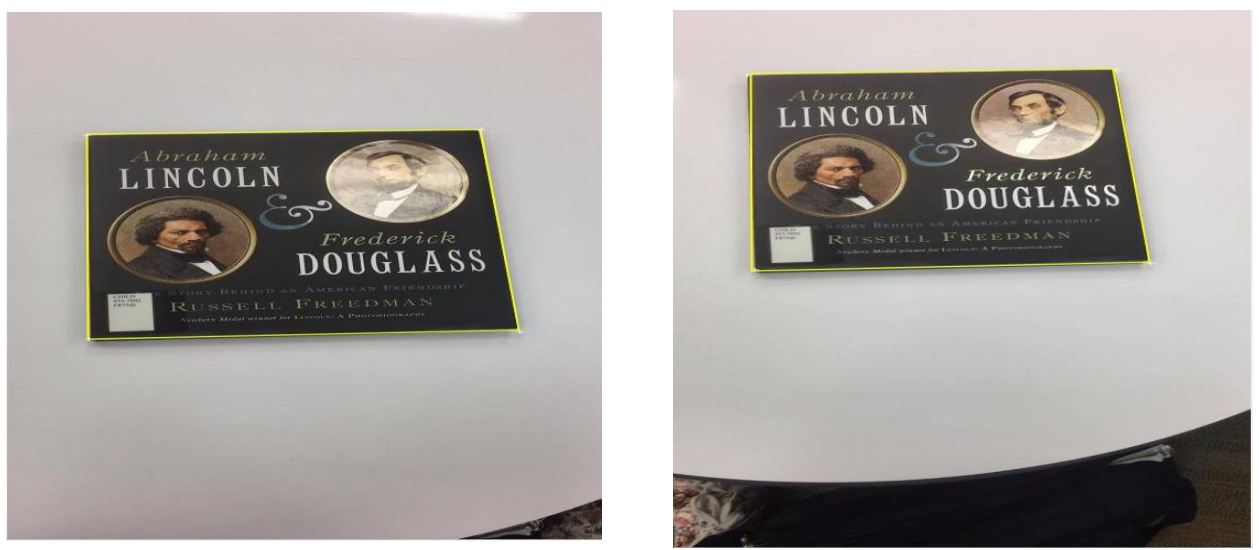


(c)

**Figure 6.1 Results from Automatic detection of the ROI in (a) Book placed on an opaque surface, (b) Photo frame, (c) Book in high illumination condition**

### 6.2. Results of Tracking using KLT algorithm

The four points corresponding to the region of interest obtained from the first frame are then tracked throughout the video. Since it is not possible to show the video sequence here, this is displayed by using two frames placed side by side. Figure 6.2 shows the result of KLT tracking. The white “+” is used to mark the tracked region of interest and the yellow region corresponds to the boundary of the region of interest.

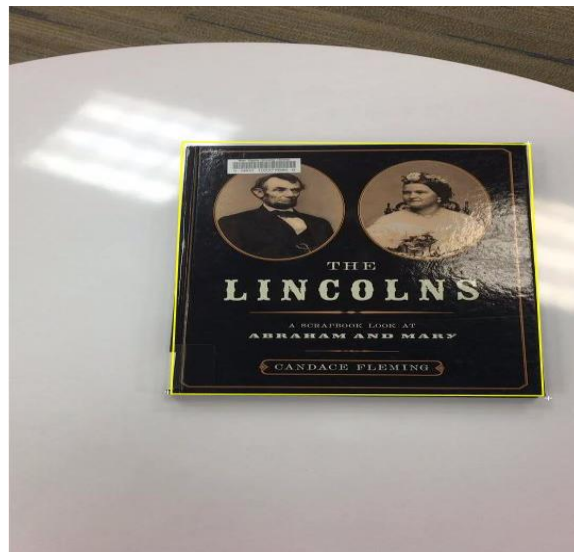
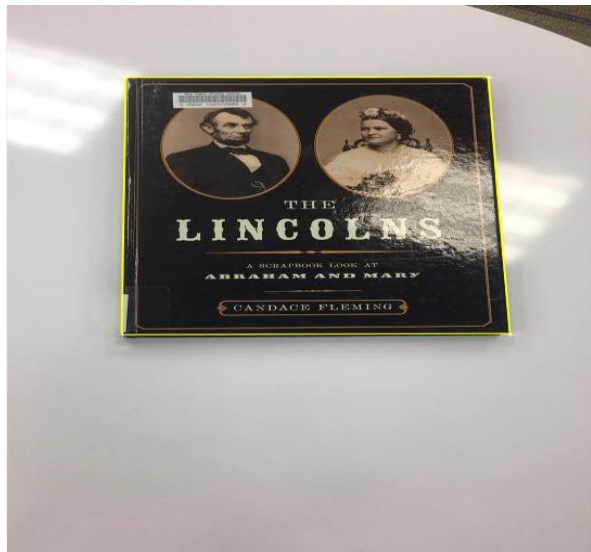


(a)



(b)



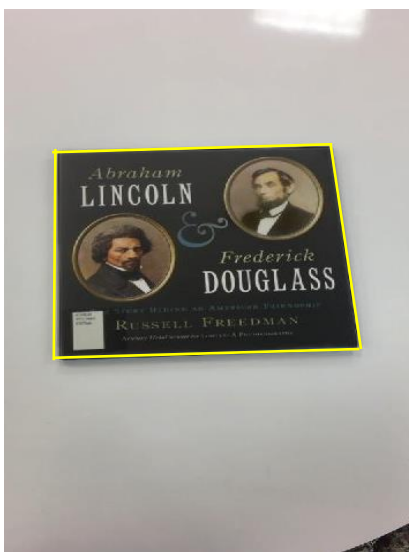
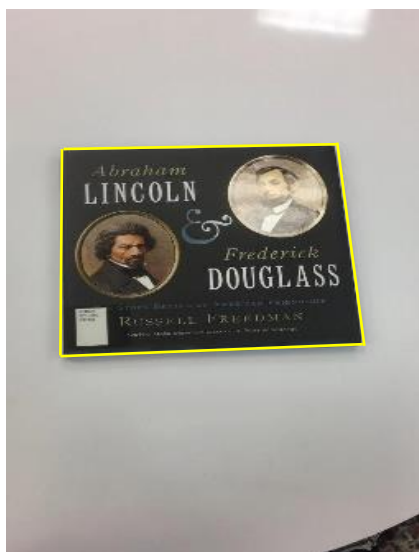


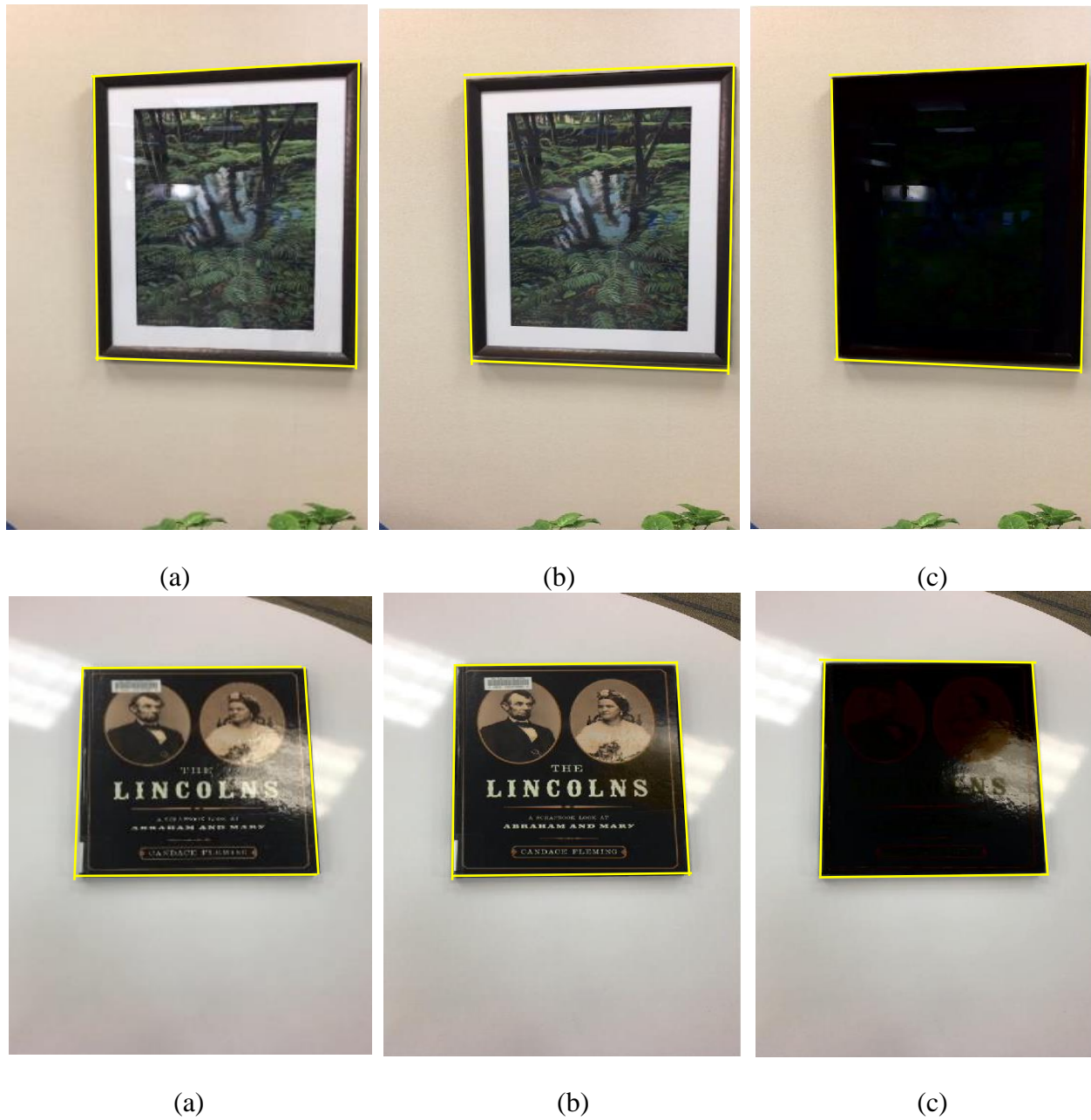
(c)

Figure 6.2 Results from ROI tracking using KLT in (a) Book placed on an opaque surface, (b) Photo frame, (c) Book in high illumination condition

### 6.3. Results of Specularity Removal

Figure 6.1. displays the results for video sequence placed on opaque surface, photo frame and a book in high illuminating condition respectively. The region of interest is bounded by the yellow boxes in all the frames.

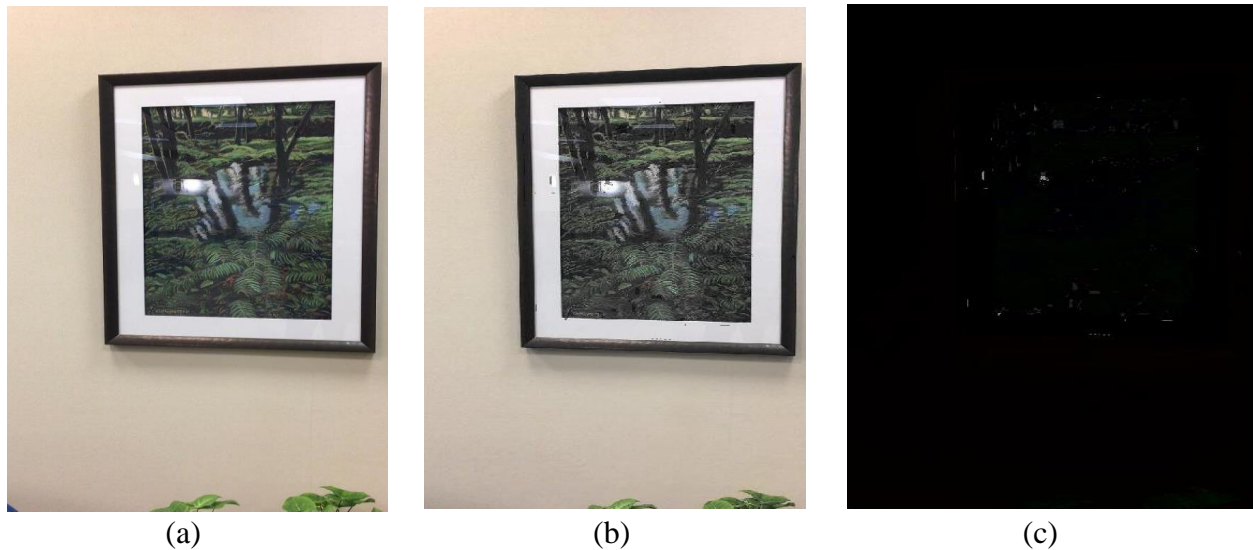




**Figure 6.3 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by our method for different video sequences**

#### 6.4. Results for Comparison with State of Art Algorithms

The quantity metrics devised in section 5 is used to compare the current formulated algorithm to the state of art algorithms. Figure 6.4 is the output obtained from the algorithm formulated in the paper Chromaticity-based separation of reflection components in a single image [21].



**Figure 6.4 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by [21]**

Figure 6.5 is the output obtained from the algorithm formulated in the paper Single Image Layer Separation using Relative Smoothness [36].



**Figure 6.5 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by [36]**

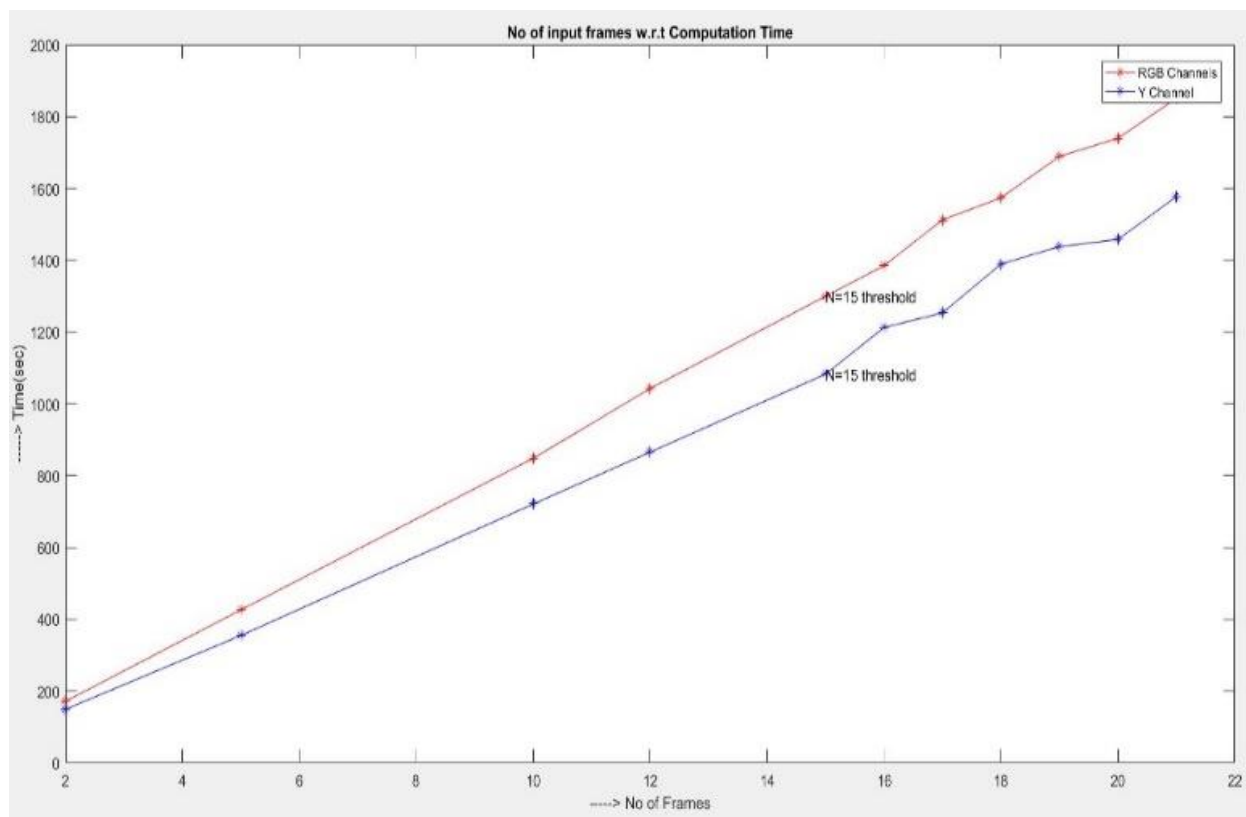
Figure 6.6 is the output obtained from the algorithm formulated in the paper Exploiting Reflection Change for Automatic Reflection Removal [37].



**Figure 6.6 Visual display of results (a) Original Frame – Y (b) Transmission layer - T and (c) Specular Reflection Layer - E recovered by [37]**

### 6.5. Implementation Details

In the Algorithm 3 the parameters  $\lambda_1 = 0.3/\sqrt{(w*h)}$ ,  $\lambda_2 = 50/\sqrt{(w*h)}$ ,  $\lambda_3 = 1/\sqrt{(w*h)}$ ,  $\lambda_4 = 5/\sqrt{(w*h)}$ ,  $\lambda_5 = 50/\sqrt{(w*h)}$  and  $\lambda_6 = 50/\sqrt{(w*h)}$ , where  $w$  is the width of the region of interest and  $h$  is the height of the region of interest. The Algorithm 4 is applied on the Y channel, after the replacement of the region of interest with the specular free image the Y, U and V channels are combined to form the original video sequence without specularity. The experiments are conducted in MATLAB on a PC running Windows 10 32bit operating with Intel Core i5 processor with 1TB hard disk and 8 GB RAM. The main advantage of the algorithm over the state of art algorithms is that it preserves the illumination change, temporal coherence and the computational complexity is minimized. Figure 6.7 compares the Algorithm presented in this paper (using Y channel) over the state of art algorithms (using RGB Channels).



**Figure 6.7 Computational time for different number of frames for Y channel and RGB channel**

## 7. Conclusion

This thesis has proposed a novel and robust method for the specular removal from hand-held video. Hand-held videos are prone to flickering and illumination changes, to deal with this a noise term is used in the problem formulation to preserve the smoothness in a video and the automaticity in selection of region of interest along with error detection for all the frames are introduced. Unlike the existing solutions for images computed on RGB channel, this paper evaluates the algorithm on Y channel. The quantity metrics is introduced to compare this novel algorithm with the existing algorithms.

The experimental results compared to the state of art algorithms manifests that the proposed algorithm has higher accuracy and speed.

### 7.1. Future Scope

The best 15 frames are to be selected for efficient and effective specular removal. The frames are selected such that there exists very little correlation between the frames. Using the best 15 frames in the above algorithm, the video obtained appears to be the same quality (to human eye and by the devise quantity metrics) as that of the video sequence obtained by using the frames (1:17:250). Every frame of the acquired video should contain the 4-sided polygon region of interest since, the correspondence between the frames is exploited to remove the highlights. In future, this algorithm and quantity metrics can aid for videos that have no correspondences. In addition, the region of interest can be varied from the 4-sided polygon to any shape. The main drawback is tracking of irregular shape region of interest.



## 8. Electron Paramagnetic Resonance Imaging

Since, I am working under Dr. Mark in his project of Electron paramagnetic resonance imaging this Chapter gives a brief explanation of the project and my contribution to this project. An Electron paramagnetic resonance (EPR) spectrometer/Imager system has been developed and locally built in West Virginia University that can operate in a wide range of frequencies, up to L-band. Its primary purpose is pre-clinical multi-functional in vivo EPR studies using small rodents, such as mice and rats. The system has been designed using the three concepts in mind

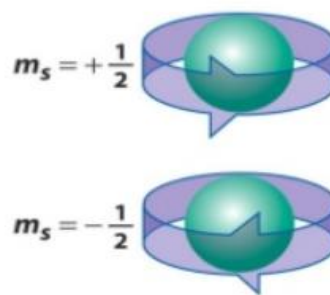
1. Low cost
2. Versatility
3. Dissemination

This has been done by using the off-shelf electronics that permit remote control from personal computers.

### 8.1 Introduction to Electron Paramagnetic Resonance

EPR stands for Electron Paramagnetic Resonance or Electron Spin Resonance (ESR). It is very like the well-known MRI technique, except from the fact that MRI measures nuclear spins and EPR does electron spins. The electron has a magnetic moment associated with its spin, intrinsic angular momentum. According to Quantum Mechanics, in the presence of external magnetic field there are two distinctive orientations along the field axis: up- and down-spin states. These two quantum states ( $m_s = -1/2$  and  $m_s = +1/2$  have different energy). The difference in energy (splitting between the two energy levels) is called Zeeman splitting.

Figure 8.1 Electron spin states



EPR phenomenon is observed in

1. Atoms having odd number of electrons.
2. Ions having partially filled inner electron shell.
3. Free radicals having unpaired electron.

## 8.2 Resonance phenomenon

Figure 8.2 shows the splitting of the energy which is due to the Zeeman Effect. The splitting between the energy levels is given by  $\Delta E = g_e \mu_b B_0$  where  $g_e$  is the g-factor of electron which is 2.0023 and  $\mu_b$  is the Bohr magneton. When the energy of photon (radiofrequency radiation) equals  $\Delta E$  resonance absorption is observed.

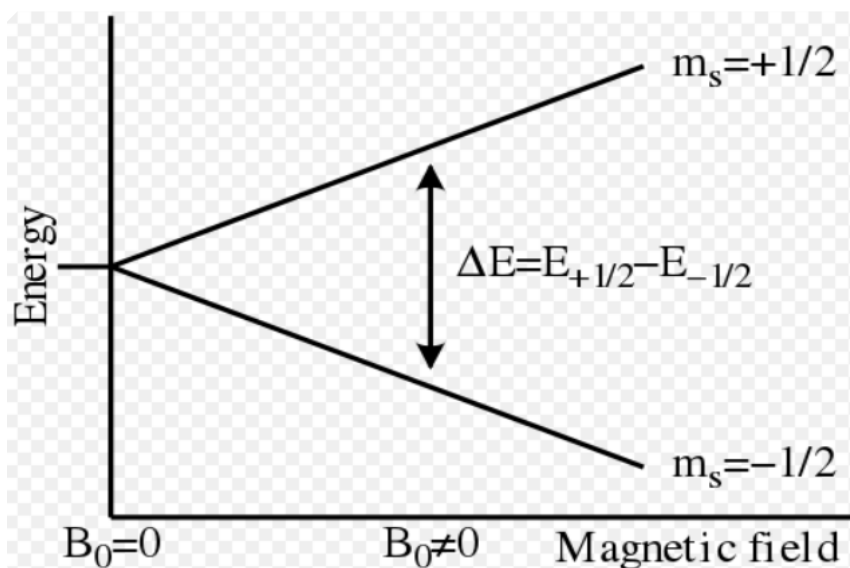


Figure 8.2 Splitting of Energy levels<sup>2</sup>

An alternative explanation electron spin resonance can be given based on the phenomenon of precession. Due to the presence of the angular momentum of the dipoles, they don't align along the magnetic field but rather precess like a mechanical top (see Figure 8.3). In EPR it is called Larmor Precession.

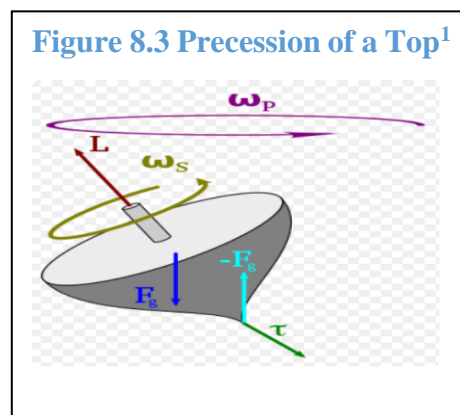


Figure 8.3 Precession of a Top<sup>1</sup>

<sup>1</sup> <https://en.wikipedia.org/wiki/Precession>

<sup>2</sup> [https://en.wikipedia.org/wiki/Electron\\_paramagnetic\\_resonance](https://en.wikipedia.org/wiki/Electron_paramagnetic_resonance)



When an Electro-Magnetic radiofrequency wave (RF) of certain frequency is applied, and the frequency of RF is equal to that of the processing spin, then the magnetization vector associated with the electrons is tilted. This also constitutes the resonance phenomenon. Component of the vector along RF magnetic field component ( $M_x$ ) is called dispersion, and  $M_y$  component is proportional to resonance energy absorption. The frequency is directly proportional to the applied magnetic field given by  $h\nu = g_e\mu_b B_0$ .

Because of the above relationship, required frequency of radiation depends on the strength of the magnetic field.

- Common field strength is 0.34 to 1.24T
- 9.5 to 35 GHz
- Microwave region
- In Vivo applications RF frequency is less than 1000MHz because of the skin effect.

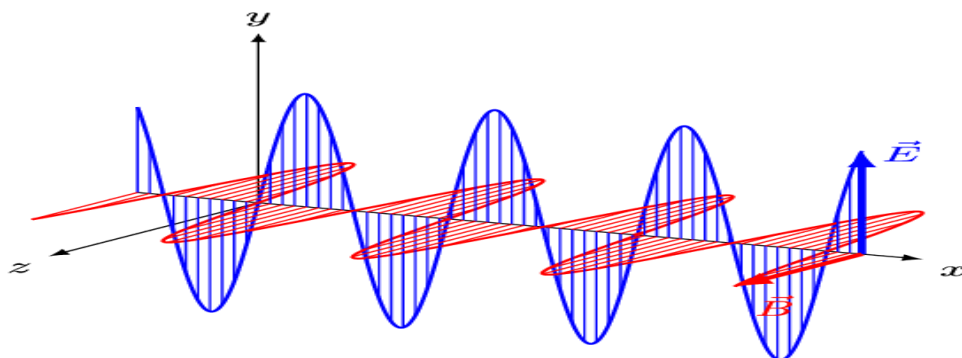


Figure 8.4 Electro-Magnetic Wave<sup>1</sup>

To measure EPR only from a certain portion of the body rather than the whole, magnetic field gradients are used, which encode spatial information into EPR spectra.

### 8.3 Bloch Equation

The Bloch equations allow computation of the absorption and dispersion components of the magnetization vector. In the equation shown below (see figure 8.5),  $T_1$  and  $T_2$  are the relaxation times., which can report local oxygen concentration, including in vivo.

<sup>1</sup><https://commons.wikimedia.org/wiki/File:EM-Wave.gif>

$$\begin{aligned}\frac{dM_x(t)}{dt} &= \gamma (M_y(t)B_z(t) - M_z(t)B_y(t)) - \frac{M_x(t)}{T_2} \\ \frac{dM_y(t)}{dt} &= \gamma (M_z(t)B_x(t) - M_x(t)B_z(t)) - \frac{M_y(t)}{T_2} \\ \frac{dM_z(t)}{dt} &= \gamma (M_x(t)B_y(t) - M_y(t)B_x(t)) - \frac{M_z(t) - M_0}{T_1}\end{aligned}$$

**Figure 8.5 Bloch Equation<sup>1</sup>**

Where,

M is magnetization vector

M<sub>x</sub> is Dispersion

M<sub>y</sub> is Absorption

B<sub>x</sub> and B<sub>y</sub> are components of continuous wave RF field along x and y axis

B<sub>z</sub> constant applied magnetic field

T<sub>1</sub> is longitudinal relaxation time

T<sub>2</sub> is transverse relaxation time, inversely proportional to O<sub>2</sub> partial pressure

Gamma is gyromagnetic ratio for electron. Longitudinal relaxation time is the time taken for the.

Figure 8.6 shows the total absorption and dispersion used to calculate the amount of oxygen.

<sup>1</sup>[https://en.wikipedia.org/wiki/Bloch\\_equations](https://en.wikipedia.org/wiki/Bloch_equations)

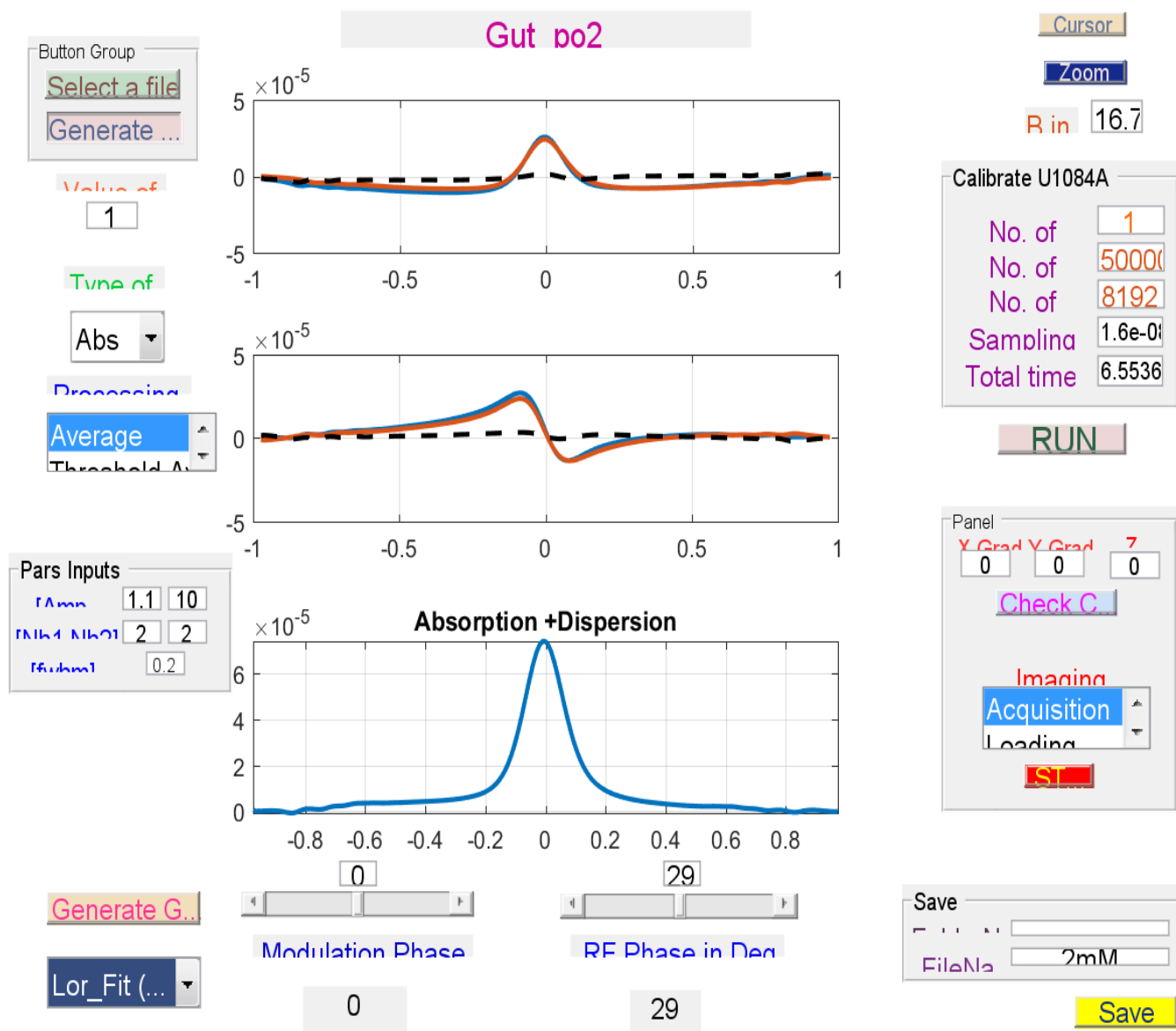


Figure 8.6 Absorption and dispersion Graph

## 8.4 Why do we Need EPR?

MRI gives the structural insight into the body, the main reason why we need EPR is to obtain functional images, such as  $O_2$  present in a tissue. This EPR is used to study cancer. The survey by National Cancer Institute depicts that, there are around 1,688,780 cancer cases and 600,920 cancer deaths. There are 1,650 cancer deaths every day. Figure 8.7 shows this.



Figure 8.7 Survey by National Cancer Institute<sup>1</sup>

The cancer is the second most leading cause of death in United states by Centers for disease control and Prevention. Figure 8.8 displays this.

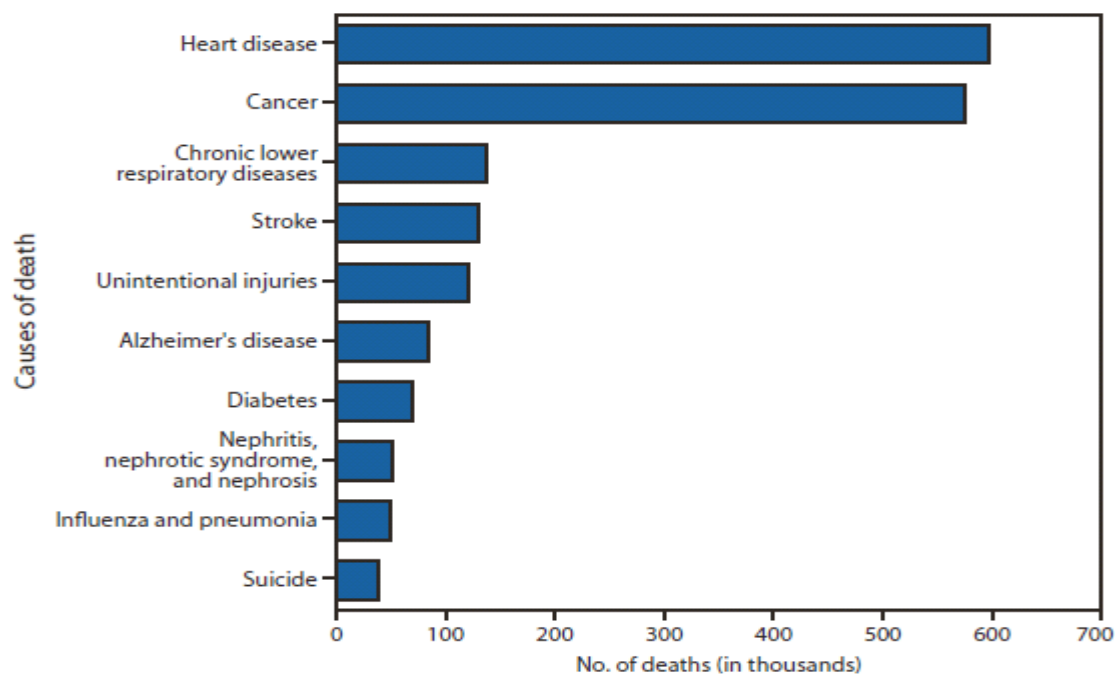


Figure 8.8 Survey by Centers for disease control and Prevention<sup>2</sup>

<sup>1</sup> <https://cancerstatisticscenter.cancer.org/#/>

<sup>2</sup> <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6208a8.htm>

The oxygen molecule is a diradical with two unpaired electrons. Whenever there is a tumor, that tissue has lower oxygen content. These tumors are treated by

- Radiation
- Chemotherapy

Both needs oxygen to be given for better results. Since, MRI gives mostly the structural information, if we want to measure something more like, the oxygen content in a tissue, EPR imaging is used.

## **8.5 Radio Frequency (RF) Electronics Design**

Since, the RF frequency in vivo applications are generally less than 1000MHz, an AWG (Arbitrary Waveform Generator) 33622A having a bandwidth of 120MHz is used. It can produce any random waveform but, only two most significant waveforms for EPR are expended. They are Frequency sweeps (for tuning) and continuous waveform (data collection). As 120MHz is very frequency for measuring oxygen in rodents, it is up-converted by mixing it with ultra-low noise, fixed frequency source. This source provides a 10MHz reference clock along with 250MHz, 750Mhz and 1GHz output signal. As a result, the maximum frequency that can be produced is 1.2GHz. For tuning purpose AWG generates a frequency sweep, and a reflected signal is digitized by the oscilloscope (DPO 5104B). The signal is read from the scope and is Fourier Transformed in real time which is read by the computer to find the frequency of resonance. After acquiring the frequency, at that frequency a continuous waveform is produced by the same AWG for collection of data using the digitizer (U1084A) to attain real time EPR signals.

## **8.6 Audio Frequency Modulation and Scans Design**

Rapid scan coils were designed and 3D printed locally using Litz wire to increase the coil efficiency. This coil produces 30G peak-to-peak sinusoidal scans. Commercially available 1.8 kW audio amplifier Cerwin-Vega CV-1800 was used to drive the coils. Since the amplifier doesn't control current, and therefore phase and amplitude of the scanning field, a digital feedback system is used to control both. This is done by using a second AWG, to produce sinusoidal scan waveforms that is amplified by CV-1800. Amplitude and phase of AWG are both computer controlled. The voltage across the coils is digitized by PicoScope 5000 USB-connected

oscilloscope. A feedback is repetitively run to guarantee the amplitude and phase stability. Because of this feedback loop, the digitized is triggered at the very beginning of each scan.

### 8.7 Magnets and Gradient System Design

An L-band electromagnet and 750 MHz permanent magnet fortified with 3D gradients are existing in the lab. To the permanent magnet that is used for imaging, a set of slow scan coils were added to permit additional  $\pm 30\text{G}$  sweep range. These coils are powered by Kepco ATE 15-5M power supply that operates in current control mode. Kepco is computer controlled using NI PCIe 6363 DAQ card and a Hall probe. The same card is used to generate voltage waveforms for magnetic field gradients. The waveforms are converted into current in the gradient coils using 262PN MRI amplifiers. Figure 8.9 is the permanent magnet with gradients inside it along with the rapid scan coils.

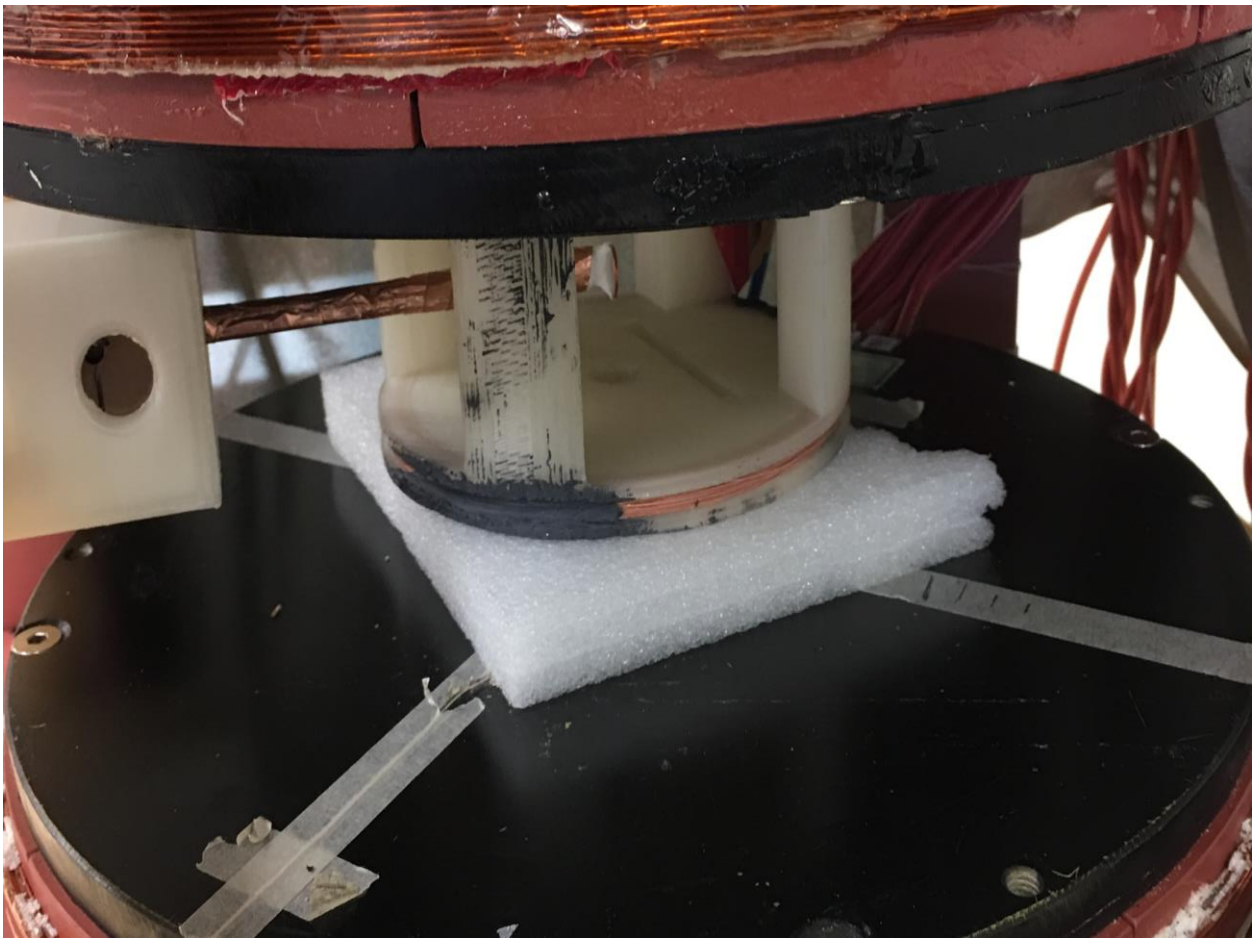


Figure 8.9 Rapid scan coil (White colored) and Permanent Magnet (black Colored)

Figure 8.10 shows the developed GUI for remote control, feedback loop, data acquisition and testing.

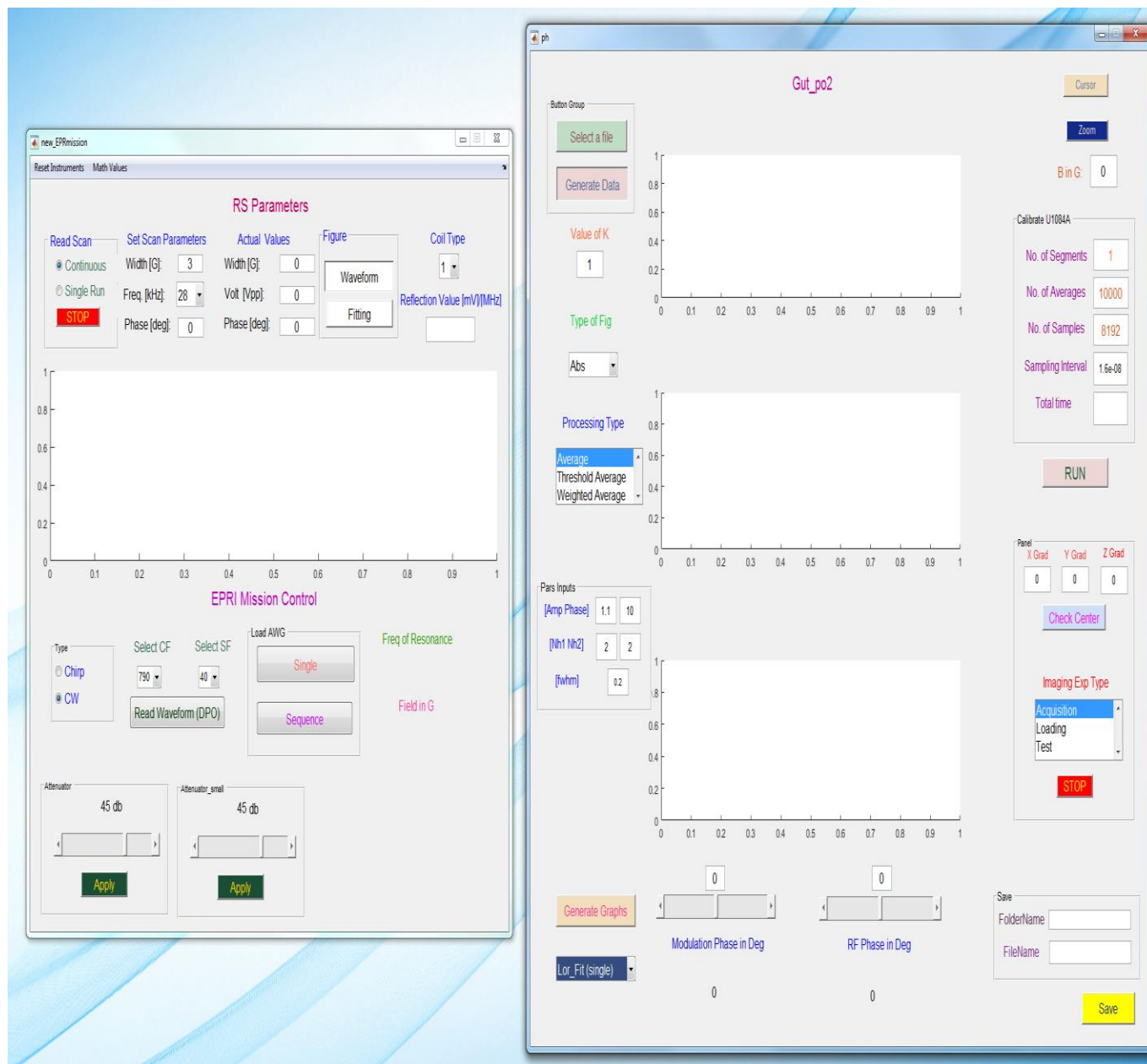


Figure 8.10 GUI for Control and Data Acquisition

## 8.8 EPR Imaging

Figure 8.11 shows the 4D image obtained by means of the rapid scan deconvolution [41].



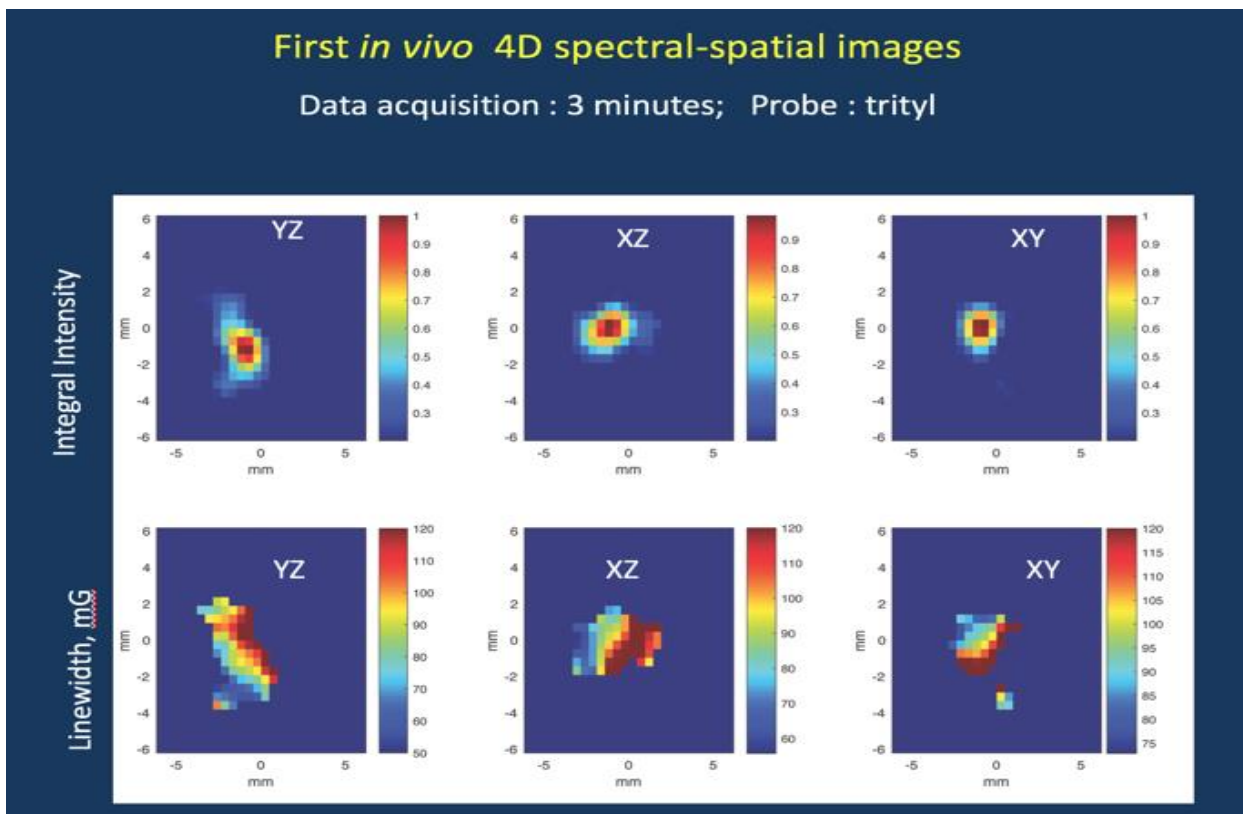


Figure 8.11 4-D spectral-spatial images in vivo

Figure 8.12 shows the data (Absorption + Dispersion), which states about the amount of oxygen in a tissue.

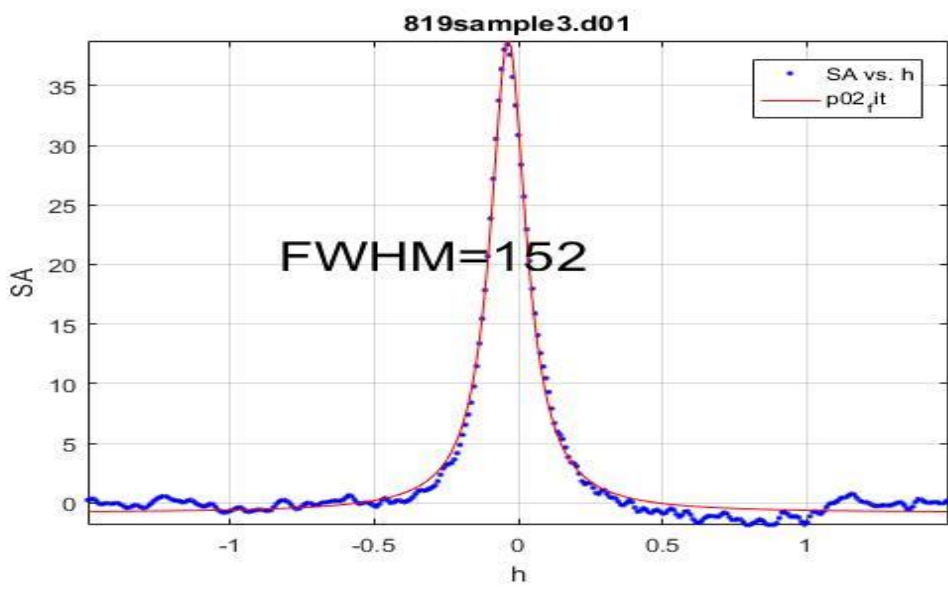


Figure 8.12 Fitted Absorption + Dispersion



If the FWHM is more (curve is broader) that means it has a high amount of oxygen at a tissue. FWHM is directly proportional to the amount of oxygen and is inversely proportional to the transverse relaxation time ( $T_2$ ).

## Bibliography

- [1] Biometrics.idealtest.org. (2017). Biometrics Ideal Test. [online] Available at: “CASIA-FaceV5, <http://biometrics.idealtest.org/>” [Accessed 9 Mar. 2017].
- [2] Jorge Bernal, F. Javier Sánchez, Cristina Rodríguez de Miguel and Gloria Fernández-Esparrach. Building up the Future of Colonoscopy – A Synergy between Clinicians and Computer Scientists.
- [3] Klinker, G.J., Shafer, S.A. and Kanade, T., 1987, June. Using a color reflection model to separate highlights from object color. In Proc. ICCV (Vol. 87, pp. 145-150).
- [4] Klinker, G.J., Shafer, S.A. and Kanade, T., 1988. The measurement of highlights in color images. *International Journal of Computer Vision*, 2(1), pp.7-32.
- [5] Schlüns, K. and Teschner, M., 1995, January. Fast separation of reflection components and its application in 3d shape recovery. In *Color and Imaging Conference* (Vol. 1995, No. 1, pp. 48-51). Society for Imaging Science and Technology.
- [6] Schluns, K. and Koschan, A., 2000, October. Global and local highlight analysis in color images. In Proc. 1st Int. Conf. Color Graphics Image Processing (pp. 300-304).
- [7] Bajcsy, R., Lee, S.W. and Leonardis, A., 1996. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *International Journal of Computer Vision*, 17(3), pp.241-272.
- [8] Tan, R.T. and Ikeuchi, K., 2005. Separating reflection components of textured surfaces using a single image. *IEEE transactions on pattern analysis and machine intelligence*, 27(2), pp.178-193.
- [9] Tan, R.T. and Ikeuchi, K., 2005. Illumination color and intrinsic surface properties physical-based color analysis from a single image. *Transactions of Information Processing Society of Japan* 46 (2005), 17–40.
- [10] Yoon, K.J., Choi, Y. and Kweon, I.S., 2006, October. Fast separation of reflection components using a specularly-invariant image representation. In *Image Processing, 2006 IEEE International Conference on* (pp. 973-976). IEEE.

- [11] Mallick, S.P., Zickler, T., Belhumeur, P.N. and Kriegman, D.J., 2006, May. Specularity removal in images and videos: A PDE approach. In European Conference on Computer Vision (pp. 550-563). Springer Berlin Heidelberg.
- [12] Mallick, S.P., Zickler, T., Belhumeur, P. and Kriegman, D., 2006, July. Dichromatic separation: specularity removal and editing. In ACM SIGGRAPH 2006 Sketches (p. 166). ACM.
- [13] Quan, L. and Shum, H.Y., 2003, October. Highlight removal by illumination-constrained inpainting. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on (pp. 164-169). IEEE.
- [14] Tappen, M.F., Freeman, W.T. and Adelson, E.H., 2005. Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(9), pp.1459-1472.
- [15] Angelopoulou, E., 2007, October. Specular highlight detection based on the Fresnel reflection coefficient. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (pp. 1-8). IEEE.
- [16] Lin, S., Li, Y., Kang, S.B., Tong, X. and Shum, H.Y., 2002, May. Diffuse-specular separation and depth recovery from image sequences. In European conference on computer vision (pp. 210-224). Springer Berlin Heidelberg.
- [17] Feris, R., Raskar, R., Tan, K.H. and Turk, M., 2004, October. Specular reflection reduction with multi-flash imaging. In Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on (pp. 316-321). IEEE.
- [18] Agrawal, A., Raskar, R., Nayar, S.K. and Li, Y., 2005. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics (TOG)*, 24(3), pp.828-835.
- [19] Shen, H.L. and Zheng, Z.H., 2013. Real-time highlight removal using intensity ratio. *Applied optics*, 52(19), pp.4483-4493.
- [20] Shen, H.L. and Cai, Q.Y., 2009. Simple and efficient method for specularity removal in an image. *Applied optics*, 48(14), pp.2711-2719.

- [21] Shen, H.L., Zhang, H.G., Shao, S.J. and Xin, J.H., 2008. Chromaticity-based separation of reflection components in a single image. *Pattern Recognition*, 41(8), pp.2461-2469.
- [22] Gai, K., Shi, Z. and Zhang, C., 2012. Blind separation of superimposed moving images using image statistics. *IEEE transactions on pattern analysis and machine intelligence*, 34(1), pp.19-32.
- [23] Peng, Y., Ganesh, A., Wright, J., Xu, W. and Ma, Y., 2012. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), pp.2233-2246.
- [24] Guo, X., Cao, X. and Ma, Y., 2014. Robust separation of reflection from multiple images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2187-2194).
- [25] Color space. (2017, March 7). In Wikipedia, The Free Encyclopedia. Retrieved 06:00, March 9, 2017, from [https://en.wikipedia.org/w/index.php?title=Color\\_space&oldid=769021088](https://en.wikipedia.org/w/index.php?title=Color_space&oldid=769021088).
- [26] Gonzalez, Rafael C and Richard E Woods. *Digital Image Processing*. 3rd ed. Reading, Mass.: Addison-Wesley, 1998. Print.
- [27] "Color Conversion - Equasys GmbH". *Equasys.de*. N.p., 2017. Web. 9 Mar. 2017, from <http://www.equasys.de/colorconversion.html>.
- [28] Lucas, B.D. and Kanade, T., 1981. An iterative image registration technique with an application to stereo vision.
- [29] Carlo Tomasi and Takeo Kanade. *Detection and Tracking of Point Features*. Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.
- [30] Jianbo Shi and Carlo Tomasi. *Good Features to Track*. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593-600, 1994.
- [31] Stan Birchfield. *Derivation of Kanade-Lucas-Tomasi Tracking Equation*. Unpublished, January 1997.
- [32] Lin, Z., Chen, M., Wu, L. and Ma, Y., 2009. The Augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report UILU\_ENG-09-2215, UIUC, Technical Report.

- [33] Szeliski.org. (2017). Computer Vision: Algorithms and Applications. [online] Available at: <http://szeliski.org/Book/> [Accessed 9 Mar. 2017].
- [34] Baker, S. and Matthews, I., 2004. Lucas-kanade 20 years on: A unifying framework. *International journal of computer vision*, 56(3), pp.221-255.
- [35] Capel, D. and Zisserman, A., 2003. Computer vision applied to super resolution. *IEEE Signal Processing Magazine*, 20(3), pp.75-86.
- [36] Li, Y. and Brown, M.S., 2014. Single image layer separation using relative smoothness. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2752-2759).
- [37] Li, Y. and Brown, M.S., 2013. Exploiting reflection change for automatic reflection removal. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2432-2439).
- [38] "Matching with Invariant Features", Darya Frolova, Denis Simakov, The Weizmann Institute of Science, March 2004.
- [39] Harel, J., Koch, C. and Perona, P., 2006, December. Graph-based visual saliency. In *NIPS* (Vol. 1, No. 2, p. 5).
- [40] Itti, L., Koch, C. and Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11), pp.1254-1259.
- [41] Tseytlin, M. (2017). Full cycle rapid scan EPR deconvolution algorithm. *Journal of Magnetic Resonance*, 281, pp.272-278.