

# **DYNAMIC FACIAL APPEARANCE CAPTURE USING SIX PRIMARIES**

by

Anika Mahmud

B. Sc., The Shahjalal University of Science and Technology, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2012

© Anika Mahmud, 2012

# Abstract

Facial appearance changes as people talk, move and, change expression. The key indication of this change is the skin color. In current appearance models skin color is reconstructed by considering the scattering and absorption of light within skin layers caused by melanin and hemoglobin in a three dimensional color space. Capturing dynamic facial appearance requires capturing both the temporal deformation of facial geometry and the temporal change of skin color by measuring the hemoglobin concentration change. Existing passive multi-stereo capture systems can capture high resolution temporal facial geometry. Integrating skin appearance capture with the existing passive capture system is challenging because to capture facial appearance the system requires precise color calibration and cross polarization, which are inheritably not required in a facial capture system. Integrating them in the existing system requires expensive hardware and constrained capture setup. In this thesis we present a novel method that can capture dynamic facial appearance without changing the existing passive capture system. We use a six dimensional color space to reconstruct the skin color, which can be adapted easily to any existing capture pipeline.

# Preface

This thesis was done with the collaboration of Dr. Tim Weyrich, Associate Professor at University College London and Timothy Scully. All the hemoglobin maps shown (Hemoglobin maps of Figure 3.6 and 3.7) as part of this research were generated by Tim Weyrich's group using their skin model.

The multispectral camera used in this research was partially built by Alex Gukov at University of British Columbia. The multispectral skin data (Figure 6.3) used in this research was captured by him.

# Table of Contents

<b>Abstract.</b>	ii
<b>Preface.</b>	iii
<b>Table of Contents.</b>	iv
<b>List of Figures.</b>	vii
<b>Acknowledgements.</b>	ix
<b>1 Introduction.</b>	1
<b>2 Related Work.</b>	5
<b>2.1 Background.</b>	5
2.1.1 Human Skin Coloring.	5
<b>2.2 Facial Appearance Model.</b>	12
2.2.1 Image Based Skin Color Analysis and Synthesis.	12
2.2.2 Reflectance Based Model.	13
2.2.3 Bidirectional Texture Function.	14
2.2.4 Physiology Measurement Based Models.	14
2.2.5 Emotional Appearance Models.	17
<b>2.3 Radiometric Calibration of the Capture System.</b>	18
<b>2.4 Facial Performance Capture.</b>	20
2.4.1 Fitting Face Models to Images.	21
2.4.2 Markers and Face Paint.	21
2.4.3 Active Light.	22
2.4.4 Passive Capture.	23
2.4.5 Commercial Systems.	24

<b>3</b>	<b>Skin Coloring Model from Six Primaries.</b>	<b>26</b>
3.1	Skin Color Plane from Three Primaries.	27
3.2	Hardware Setup for Capturing Skin Color from Three Primaries.	27
3.2.1	Positional Constraint of Camera and Light System.	28
3.2.2	Energy Reduction by Cross-polarization.	29
3.3	Reflection Model.	31
3.4	Skin Color Plane from Six Primaries.	32
<b>4</b>	<b>Radiometric Calibration of Capture Device.</b>	<b>37</b>
4.1	Method Overview.	38
4.2	Data Acquisition.	40
4.2.1	Multispectral Camera.	40
4.2.2	Off-the-shelf Camcorder.	41
4.3	Estimating Spectral Response.	41
4.4	Validation.	42
4.5	Result.	43
<b>5</b>	<b>Facial Performance Capture.</b>	<b>46</b>
5.1	Acquisition Setup.	46
5.2	Camera Parameter Setting.	47
5.3	Multi-Camera Synchronization.	48
5.4	Multi-Camera Calibration.	49
5.5	Radiometric Camera Calibration.	50
5.6	Multi-View Reconstruction.	53

<b>5.7 Geometry and Texture Tracking.</b>	55
5.7.1 Reference Mesh.	55
5.7.2 Frame Propagation.	56
5.7.3 Computing 2D Texture.	57
5.7.4 Smoothing.	57
<b>5.8 Output Capture Sequence.</b>	59
 <b>6 Results.</b>	 61
 <b>7 Conclusion and Future Work.</b>	 68
 <b>References.</b>	 70

# List of Figures

Figure 1.1	Digitally created old face. . . . .	1
Figure 1.2	The organization of this thesis. . . . .	4
Figure 2.1	Different layers and parts of human skin. . . . .	6
Figure 2.2	A schematic diagram of the optical pathways in skin. . . . .	8
Figure 2.3	Model of human skin. . . . .	10
Figure 2.4	Skin color surface in a 3D color space and 2D Skin color look-up surface. . . . .	11
Figure 2.5	Melanin and hemoglobin maps are extracted from a single image. . . . .	13
Figure 2.6	Illustration of Bidirectional Texture Function. . . . .	14
Figure 2.7	Example of noncontact point measure chromophore maps. . . . .	15
Figure 2.8	Facial appearance for anger from extracting hemoglobin concentration. . . . .	16
Figure 2.9	Spectral Response Curve of Canon 40D. . . . .	18
Figure 2.10	Color Chart. . . . .	19
Figure 2.11	Face Markers are used to capture the facial performance of an actor. . . . .	22
Figure 2.12	Active light methods project structured light on the actor. . . . .	23
Figure 2.13	Mova's capture system. . . . .	24
Figure 3.1	Specular reflection and diffuse surface after cross polarization. . . . .	28
Figure 3.2	Angular effect of using linear and circular polarizer. . . . .	29
Figure 3.3	Cross-polarization absorbs the specular light energy. . . . .	30
Figure 3.4	Of-the-shelf polarizer has a very low transmittance. . . . .	30
Figure 3.5	Skin data capture using six primaries. . . . .	33
Figure 3.6	Hemoglobin map from six primaries using JPEG format image. . . . .	34
Figure 3.7	Hemoglobin map from six primaries using AVCHD format data. . . . .	35
Figure 3.8	Three RGB, $R_fG_fB_f$ texture pairs capturing different parts of the face. . . . .	36

Figure 4.1	A generic Image Acquisition Pipeline. . . . .	39
Figure 4.2	Captured and reconstructed RGB values of the color chart. . . . .	43
Figure 4.3	Spectral response generated for a camera without filter. . . . .	44
Figure 4.4	Spectral response generated for a camera with filter. . . . .	45
Figure 5.1	Summing the frames in linear intensity space the final image is created. . . . .	48
Figure 5.2	CalTag calibration grid is captured for calibration. . . . .	49
Figure 5.3	SG and DC Macbeth ColorChecker charts. . . . .	51
Figure 5.4	The ColorChecker is captured during the face capture session. . . . .	52
Figure 5.5	Point clouds are created from each pair of stereo camera. . . . .	54
Figure 5.6	A single point cloud is generated by merging four point clouds. . . . .	55
Figure 5.7	Computing vertex positions for the next frame. . . . .	56
Figure 5.8	2D texture and color coded contribution map. . . . .	58
Figure 5.9	Mesh smoothing. . . . .	59
Figure 5.10	Output capture sequences. . . . .	60
Figure 6.1	Six basic expressions. . . . .	62
Figure 6.2	Results from a sequence capture of the expression “Joy” . . . . .	63
Figure 6.3	Multi-spectral facial skin capture of different subjects. . . . .	64
Figure 6.4	Results of the expression “Anger” capture. . . . .	65
Figure 6.5	Results of the expression “Sadness” capture. . . . .	66
Figure 6.6	Results of the expression “Fear” capture. . . . .	67



# Acknowledgements

First I would like to thank my supervisor, Dr. Wolfgang Heidrich, for his guidance throughout my Masters degree, and for helping me with different research ideas. I have learned so much from him during the time I have worked with him which has prepared me for my future endeavors.

Second, I would like to thank Dr. Derek Bradley and Brad Atcheson for their support during the initial phase of my research. Your patient answers to my endless questions have helped to better understand and take deeper looks in solving my research problem. I would also like to thank Cheryl Lau for just being there.

Finally, I would like to thank my mother for her lifelong support for my education.

# Chapter 1

## Introduction



Figure 1.1: Digitally created old face in the movie The Curious Case of Benjamin Button. This old face is created from capturing actor Brad Pitt's face. Recent advancement in capture system made this movie possible (Image courtesy of CBS).

In recent years computer generated faces have gained huge popularity both in motion picture (Figure 1.1) and gaming industry. The main reason behind it is the recent advancement in capture systems. These systems are capable of producing detail facial geometry with facial deformation over time from captured videos of an actor. Previously this was done by highly skilled artists with tremendous amounts of manual work. Recent research on capturing real objects aims to automate this work by capturing and modeling real features and decrease the demand on artists who create 3D models and textures. Still, capturing the facial features let alone the dynamic

facial features remains a great challenge. The face and its appearance change over time both physically and physiologically as people talk, show emotion, do physical exercise or even when they drink alcohol. Therefore, typical facial appearance changes that artists desire include the time varying geometry of the face itself, texture of the face, surface reflectance and their motion in skin color domain.

Recent research in facial performance capture has focused mostly on the geometry and its motion capture [Bradley et al. 2010b; Beeler et al. 2011]. Jimenez et al. [2010] generated a dynamic skin appearance model from skin chromophore concentration and its change over time. To our knowledge, no existing system can capture both temporal facial geometry and facial skin appearance changes simultaneously. The reason behind this is capturing skin appearance and reproducing it, is itself a hard problem. Skin is the outermost part of the human body. As a result, people are very aware of, and very sensitive to, the appearance of skin. Again skin has a complicated layered structure [Cotton and Claridge 1996]. The final appearance of skin results from complex optical interactions of many different skin components with light. This is hard to model exactly because the physiology of skin varies over race, age and gender. It also depends on the individual subjects' skin type. To render a dynamic fictional human character realistically in movies and video games through a facial capture system, it needs to capture all the subtleties of actual human skin.

The key challenge in skin appearance capture is to be able to reproduce the skin color realistically. Skin color is determined by scattering and absorption of light within the skin layers, caused mostly by concentrations of two chromophores, melanin and hemoglobin [Anderson et al. 1980. Cotton and Claridge [1996] showed that using these approximations all possible colors occurring within normal human skin have to lie on a simple curved surface patch within a three-dimensional (Red, Green, Blue) color space bounded by two physiologically meaningful axes, one corresponding to the amount of melanin and the other to the amount of hemoglobin. This observation gives an extremely useful transformation which allows for a dimension reduction from a three dimensional color space to a two dimensional surface. If we can describe this surface then it is possible to provide a mapping from a certain position on the surface to hemoglobin and melanin measurements. In other words while given a skin image in

RGB color space, it is possible to separate the hemoglobin from the melanin part. Park et al. [2002] show the time frame of change for melanin concentration is hours to weeks. So we can assume hemoglobin is the only varying factor for short period facial appearance change and measuring the changes in hemoglobin concentration map is enough to understand the dynamic facial appearance change.

In this thesis we propose a novel facial appearance capture method that can combine existing capture system with facial appearance capture very easily without adding any costly hardware or setup constraints. Our method uses six primaries instead of three to reproduce the skin data without requiring a polarizer or explicit diffuse surface constraint. Usually any digital image is presented using three primary colors. The most common primaries used in digital cameras are Red, Green and, Blue (RGB). We use a color filter to capture the same image which gives us three other primaries which are shifted by the filter response from the RGB primaries. We define our problem in these six dimensional domains of RGB and shifted RGB. While this proposed capture system is our contribution, the actual melanin and hemoglobin map separation is done by our collaborators. Since we use off-the-shelf camcorders as our capture device, which cannot save RAW data, we also present a method to calibrate the capture device without using RAW data.

In the rest of this thesis we present, related work, our method of using six primaries to reproduce the hemoglobin concentration from captured skin data, a radiometric calibration method that produces the spectral response curve of our capture devices (camcorders), our facial performance capture system and our capture results. An overview of the organization of this thesis is shown in Figure 1.2.

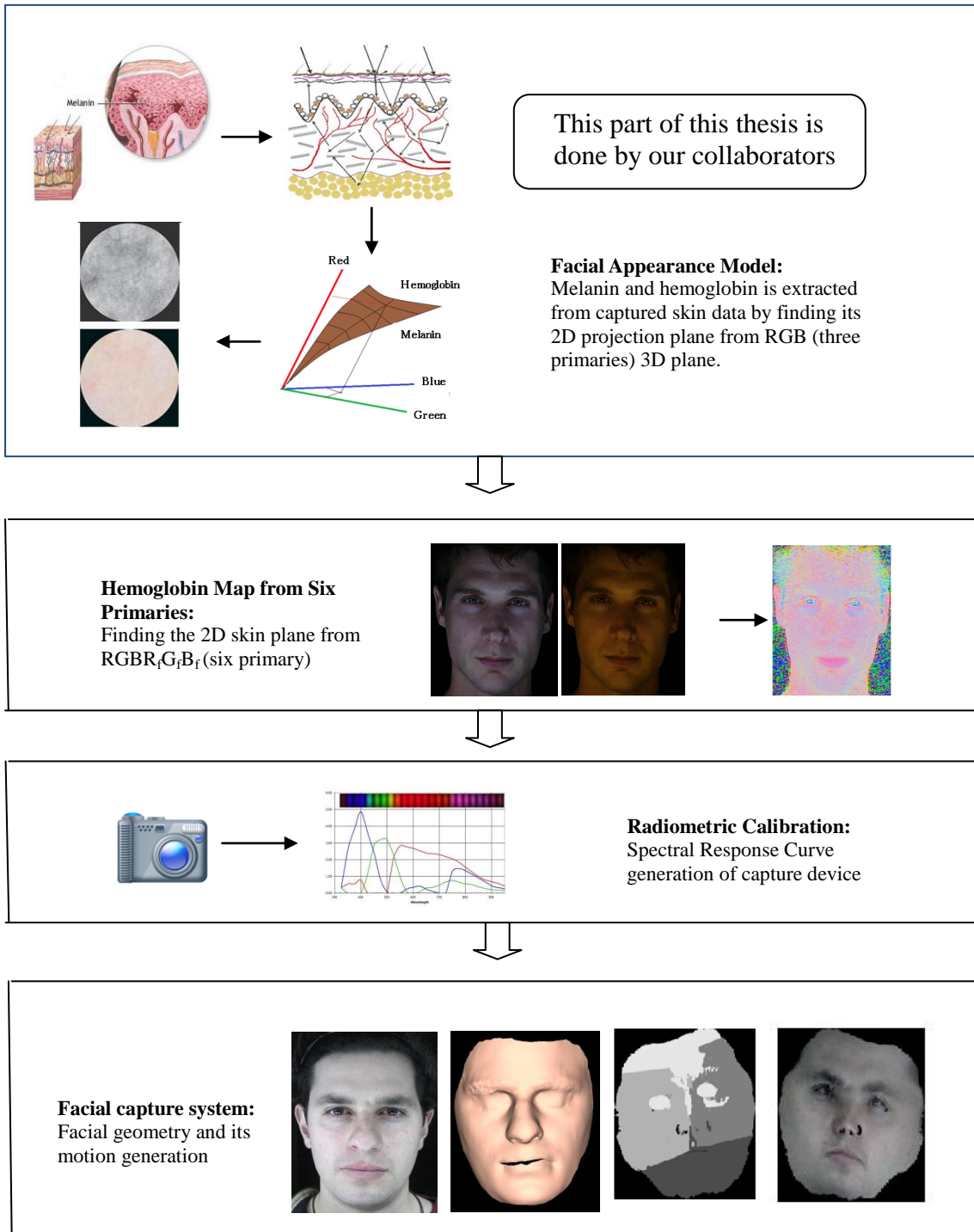


Figure 1.2: The organization of this thesis: skin color model from three primaries, skin color model from six primaries, radiometric calibration of capture device, our capture system for facial appearance capture and results.

# Chapter 2

## Related Work

This chapter outlines previous work related to the different areas of this thesis. We start with techniques related to hemoglobin and melanin concentration extraction from skin data and physical appearance model generation from melanin and hemoglobin concentration data, followed by radiometric calibration of the capture devices, and end with facial performance capture systems.

### 2.1 Background

In this section we present human skin coloring model. This model uses a common approximation of skin as a two layered translucent material and the color of the skin is defined by the distribution of two chromophores: hemoglobin and melanin.

#### 2.1.1 Human Skin Coloring

Human skin coloration is dependent almost exclusively on the concentration and spatial distribution of the skin chromophores - melanin and hemoglobin [Anderson et al. 1981]. Cotton and Claridge [1996] has proposed a model of color formation within human skin which can reproduce the skin color by considering the skin to be a layered construction of stratum corneum, epidermis, papillary dermis and reticular dermis (Figure 2.1), and by utilizing the optical properties at the interface between these layers.

To understand the development of coloring within the skin we need to study the optics of the skin. The skin is a complex structure which can essentially be thought of as two separate and structurally different layers [Cotton and Claridge 1996], the epidermis and dermis. Each of these layers has very different absorption, scattering and refracting properties.

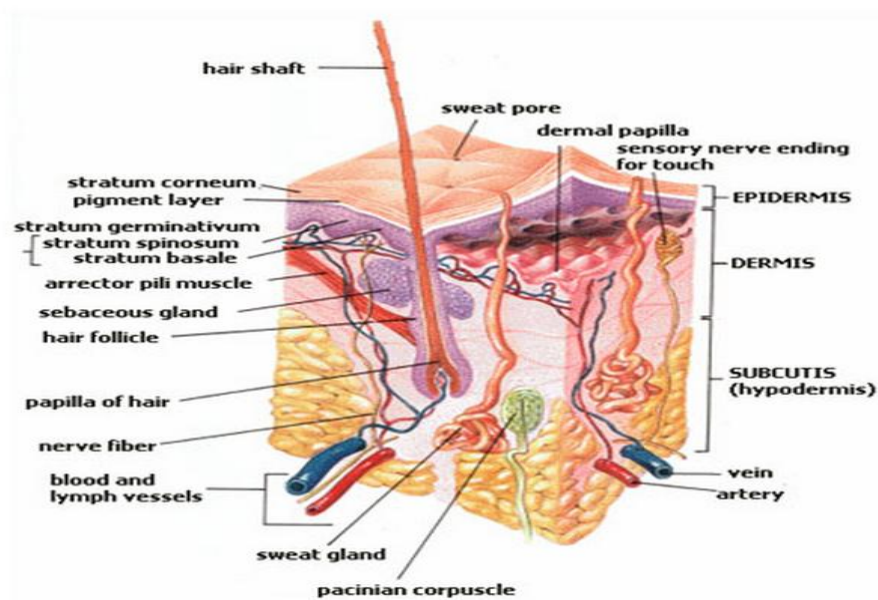


Figure 2.1: Different layers and parts of human skin.

### Interaction of light with the epidermis

The epidermis is a layer of epithelium cells which contain varying amounts of melanin with an outer layer, the stratum corneum, of keratinised epithelium cells. The stratum corneum varies considerably in thickness. It is chiefly the thickness of the stratum corneum that is responsible for the differences in thickness between the epidermis of thick and thin skin [Ross and Romrell 1989]. As the surface of the stratum corneum is not smooth and planar,

normal skin lacks specular reflectance but rather causes the incident radiation to become diffuse. Within the body of the epidermis there are many absorption peaks [Anderson and Parrish 1981]. In the visible portion of the spectrum, melanin is the only pigment affecting the transmittance of normal human epidermis, giving rise to the wide range of skin colors from ‘black’ to ‘white’ [Anderson et al. 1981]. As the epidermis contains no blood vessels, blood vessels are confined to the dermis.

### **Interaction of light with the dermis**

The dermis contrasts strongly in structure to that of the epidermis being highly vascular, containing many sensory receptors and being made largely from *collagen* fibres. Between the epidermis and dermis the junction presents an extremely uneven boundary that consists of finger like dermal protrusions called *dermal papillae*. The dermis can be split into two further histologically distinct layers the *papillary dermis* and the *reticular dermis* within which the structure of the collagen fibres differs significantly [Ross and Romrell 1989]. The first of these layers is situated directly below the epidermis within which the collagen exists as thin fibers [Lever and Schaumburg-Lever 1990]. This contrasts with the reticular dermis where the collagen fibres are aggregated into thick bundles. These then give rise to vertical branches whose caliber diminishes as they extend into the dermis. The dermis, being constructed from a densely fibrous collection of collagen fibres and blood vessels, has distinctly different optical properties to the epidermis. The absorption coefficient for bloodless dermis is far smaller than the scattering coefficient [Cotton and Claridge 1996]. The blood borne pigments are the major absorbers. Within the dermis and epidermis the chromophores responsible for scattering appear to be different from those that cause absorption [Anderson and Parrish 1981]. Within the dermis these separate into collagen fibres and blood born pigments. These collagen fibres provide scattering and blood born pigments provide absorption within the dermis. This helps the skin modeling because, in normal skin, the scattering component is therefore fixed while the absorption depending on the amount and nature of the blood present.



## Overall effect of epidermis and dermis on incident light

The light that enters the epidermis is rendered diffuse by the non-planar surface of the stratum corneum. Within the epidermis scattering is negligible, and in the visible spectrum absorption is due solely on melanin. Melanin absorbs short wavelength more strongly than long wavelength. The resulting light entering the dermis loses a blue component, depending on the amount of melanin present. Within the dermis this “brown” light gets scattered and the shorter wavelength light gets more scattered than the long wavelength. This largely determines the depth of penetration of the incoming light and controls the amount and nature of blood vessels in various parts of the spectrum encounter. These blood vessels contain pigments which absorb blue light more strongly than red, hence giving the red color of blood [Cotton and Claridge1996].

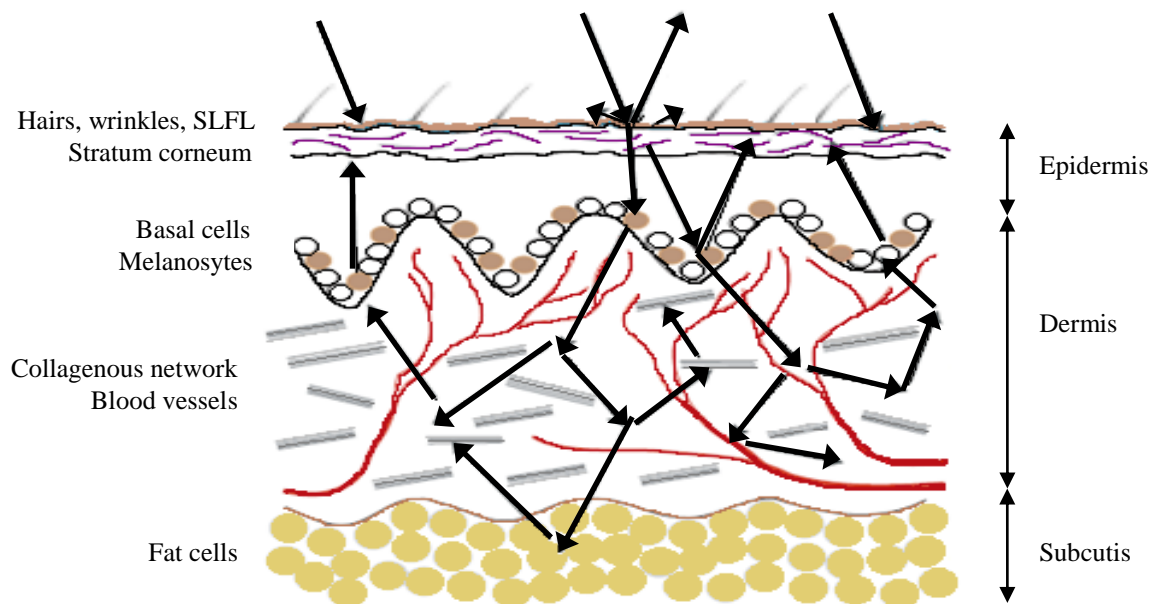


Figure 2.2: A schematic diagram of the optical pathways in skin. Part of the incident light is reflected at the surface of skin. The rest of the light penetrates into the skin layers. In the epidermal layer, the light is absorbed by melanin. In dermal layer, the light is scattered multiple times by collagen fibers and absorbed by hemoglobin.

Most of the light incoming to the dermis gets absorbed and remitted. There is zero transmittance for wavelengths less than 600nm [Anderson and Parrish 1981]. This light then passes back through the epidermis where further melanin absorption takes place before being remitted through the stratum corneum. In normal skin “brown” color due largely to melanin absorption, and red hues being due to absorption within the vascular dermis. That’s why in their skin model Cotton and Claridge [1996] consider both the epidermal and dermal layers as translucent materials where the light is scattered and absorbed by various materials embedded within these layers. These materials, which include melanin, collagen fibres, blood vessels and lead to a description of the skin as a translucent inhomogeneous material [Cotton and Claridge 1996].

### Model of entire skin

Considering the interaction of light at each layer of the skin Cotton and Claridge [1996] proposed a model for skin (Figure 2.3) which can be represented in the RGB color space as:

$$R(\rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}, d_m) = \int_0^\infty T_{total}(\lambda, \rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}) S(\lambda) S_R(\lambda) d\lambda,$$

$$G(\rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}, d_m) = \int_0^\infty T_{total}(\lambda, \rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}) S(\lambda) S_G(\lambda) d\lambda,$$

$$B(\rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}, d_m) = \int_0^\infty T_{total}(\lambda, \rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}) S(\lambda) S_B(\lambda) d\lambda,$$

where,  $\lambda$  is the wavelength,  $d_m$  is melanin absorption path length,  $S_R(\lambda)$ ,  $S_G(\lambda)$ ,  $S_B(\lambda)$ , are the device response curves,  $S(\lambda)$  represents the incident light,  $\rho_{ud}$  and  $\rho_{ld}$  are the amount of blood born pigments present in the upper and lower papillary dermis respectively and,  $d_{ud}$  and  $d_{ld}$  are their thickness respectively.

The dermal thickness can be considered to be constant and if it is assumed that the division between upper and lower papillary dermis is also constant, then,  $R$ ,  $G$  and  $B$  values are left solely

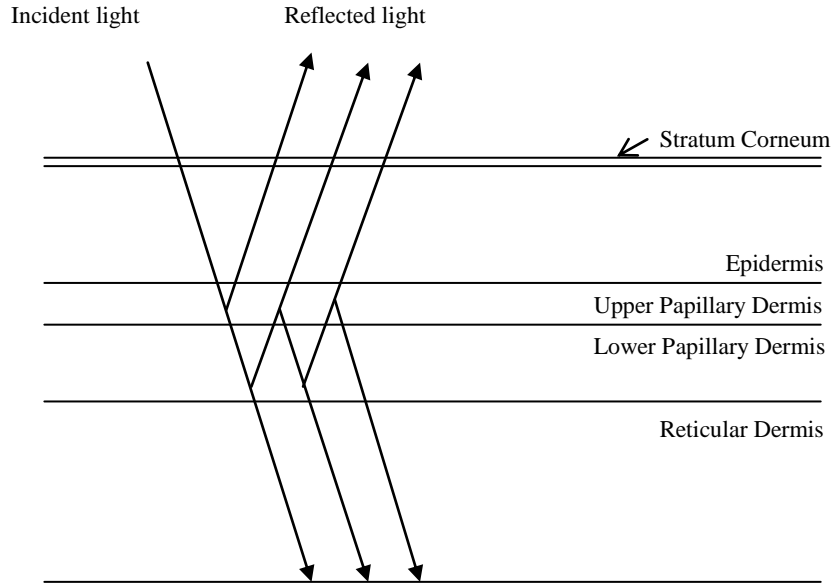


Figure 2.3: Model of human skin

depending on  $\rho_{ud}$  and  $\rho_{ld}$ , amount of blood born pigments. Cotton and Claridge [1996] assumed the ratio of the amount of blood born pigments in the upper papillary dermis to that in the lower papillary dermis does not vary the above equations and thus define a single trajectory within RGB space. The combined effect of the absorption processes within the dermis and epidermis result in the range of colors in human skin. This range of colors can be found by identifying the locus of points generated by a vector  $\mathbf{P}(\rho_{ud}, \rho_{ld}, d_{ud}, d_{ld}, d_m)$  for all possible melanin absorption path lengths, dermal absorption path lengths and the values of  $\rho$ . The light finally emitted from the skin is a product of that incident light transmitted by the epidermis and remitted by the dermis. As both of these produce defined trajectories the product define a fixed surface within RGB space. To generate this locus of points it is necessary to calculate the absorption coefficients for melanin and blood and for different path lengths. These values have been measured and published [Anderson et al. 1981]. By combining these with the RGB spectral response curves of the capture device and integrating over all possible values of  $d_m$  and  $\rho$  it is then possible to obtain the surface within RGB space representing all skin colors (Figure 2.4).

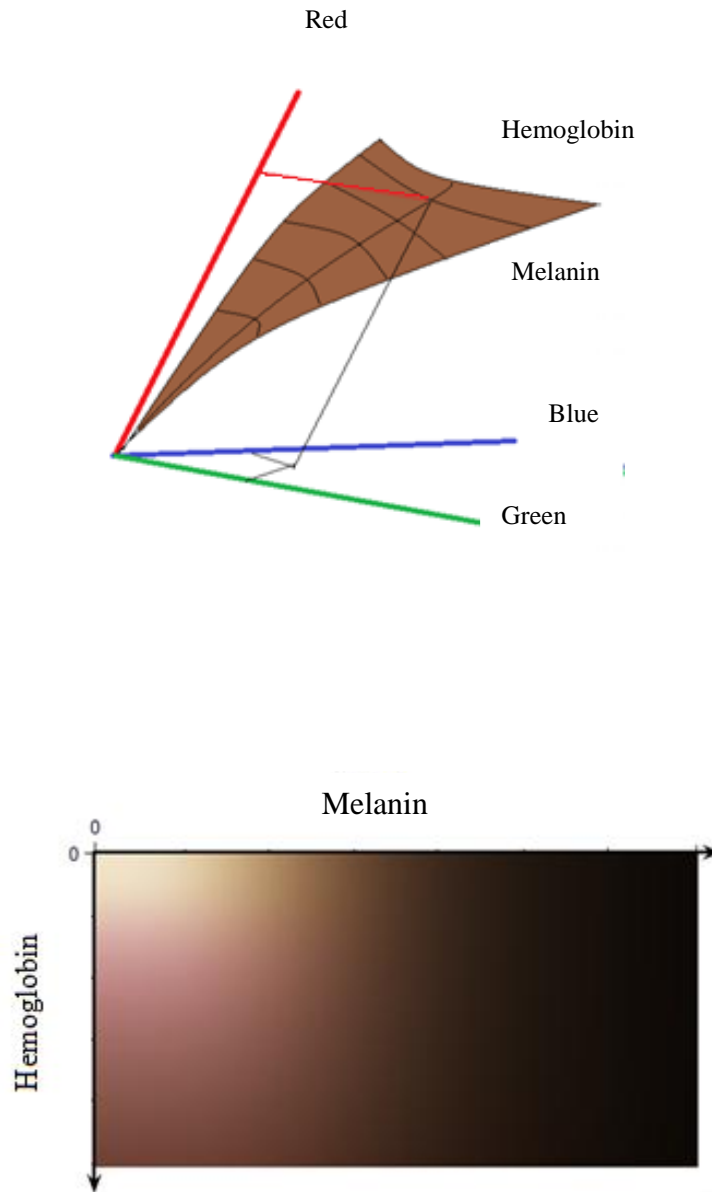


Figure 2.4: (Top) Skin color surface in a 3D (RGB) color space. (Bottom) 2D Skin color look-up surface. The 2D parameters are hemoglobin and melanin concentration.

## **2.2 Facial Appearance Model**

In recent years facial appearance models have gained popularity in the gaming and the movie industry. Facial appearance changes constantly while talking, doing exercise and while going through emotions. Realistically reproducing facial appearance is challenging because people are very aware of, and very sensitive to, the appearance of skin. Most of the facial color appearance models approximate skin as a two layered translucent structure and color appearance is defined by the distribution of melanin and hemoglobin. This approximation has proven to well describe a wide range of skin appearances [Donner and Jensen 2006; Donner et al. 2008].

### **2.2.1 Image Based Skin Color Analysis and Synthesis**

Image based skin color analysis techniques depend on captured skin data to reproduce skin color. Being a purely image based technique none of these is tied to any facial deformation or dynamic physical appearance change of the skin. Using independent component analysis, Tsumura et al. [1999; 2003] extract hemoglobin and melanin pigmentation from a single skin image (Figure 2.5). Tsumura et al. [2003] added a pyramid-based texture analysis/synthesis technique to create digital cosmetics.

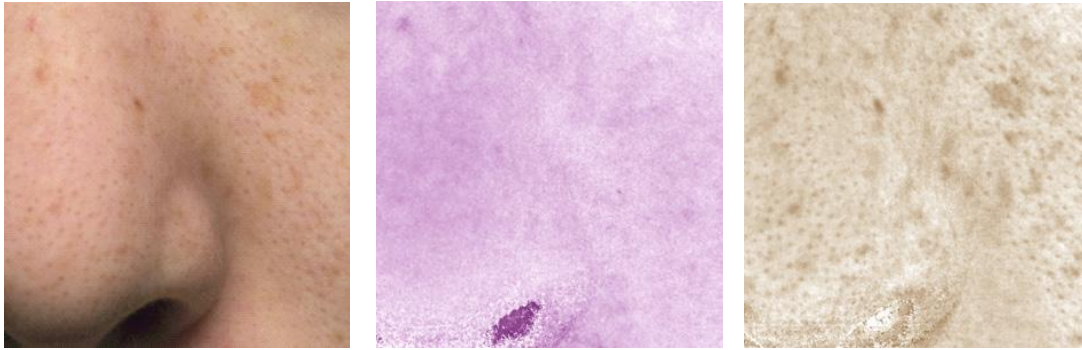


Figure 2.5: Melanin and hemoglobin maps are extracted from a single image using independent component analysis. (Image courtesy of Norimichi Tsumura; reprinted Figure 7 partially with permission [Tsumura et al. 2003]; © 2003 ACM Inc)

### 2.2.2 Reflectance Based Model

The appearance of skin varies greatly depending on the illumination and viewing conditions. This results from the directional variation of reflection of skin. Weyrich et al. [2006] analyze variations in the reflectance of facial skin under varying external conditions of 149 subjects, and compute the parameter values of a skin reflectance model that models the spatially varying subsurface scattering, texture, and specular reflection. Donner et al. [2008] simulate skin reflectance by taking into account inter-scattering of light between skin layers. They use known chromophore spectra to derive spatial chromophore distributions from multi-spectral photographs of skin. Ghosh et al. [2008] use structured light and polarization to determine skin layer properties using a multi-layered scattering model.

### 2.2.3 Bidirectional Texture Function

The appearance of skin is uneven in fine scale as skin has various structural details such as pores and wrinkles and color variations such as spots and freckles). A model of skin appearance must therefore include such features. It's often referred to as skin texture. The Bidirectional Texture Function is a useful concept for modeling the appearance of texture presented by Cula et al. [2005] and takes into account appearance variations caused by changes in illumination and viewing direction (Figure 2.6). Cula et al. [2004] use captured images of skin while changing the viewing and illumination directions. As skin texture tends to vary according to body region, age, and gender, the images they obtained were from different body regions. Thus they created a unique Skin Texture Database that includes high resolution images of various types of skin.

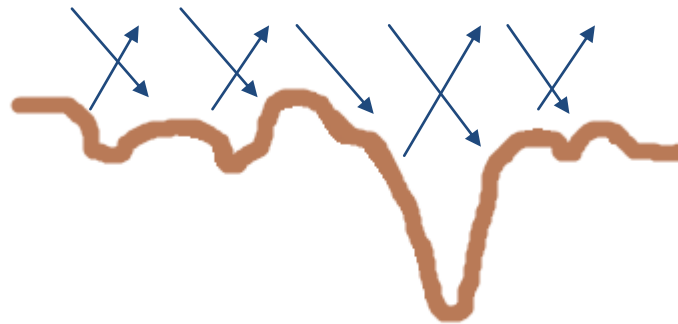


Figure 2.6: Illustration of Bidirectional Texture Function.

### 2.2.4 Physiology Measurement Based Models

In dermatology physiology measurement based methods has been used such as, ex vivo histological examination, or in vivo non-invasive point measures of blood hemoglobin [Matts 2007] (Figure 2.7). Recently Jimenez et al. [2010] used a similar non-invasive method for in-vivo mapping of hemoglobin concentration and distribution across large areas of the skin.

They correlate the change of hemoglobin with dynamic facial expression (Figure 2.8). For reconstructing the skin color Jimenez et al. [2010] uses a two layer skin model [Cotton and Claridge 1996].

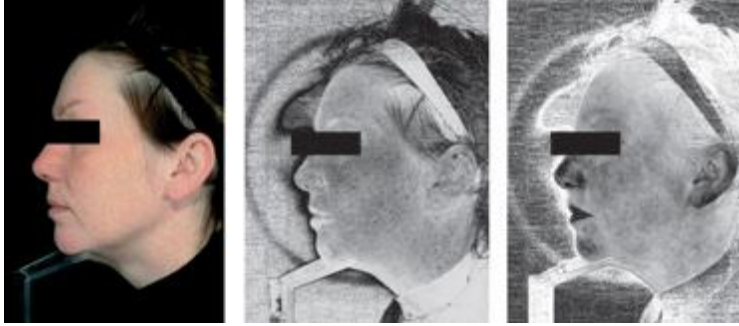


Figure 2.7: Example of noncontact point measure chromophore maps [Matts et al. 2007]. Left: original digital photo. Middle: Melanin concentration map. Right: Hemoglobin concentration map. (Image courtesy of P. J. Matts.)

The current state of the art technique to capture accurate reconstructions of the hemoglobin and melanin distribution within the facial skin is called non-contact SIAscope™ system [Cotton et al. 1999]. Non-contact SIAscope™ is implemented using finely calibrated yet conventional digital camera and lighting system. The light system needs to be completely cross-polarized to eliminate specular reflection or in other words no subsurface information. This system treats the camera as a three band spectrometer by reading the RAW data (RGB Bayer pattern). For every pixel of the RAW image, using a light transportation model of skin this system calculates melanin and hemoglobin concentrations. Using this system with current state of the art facial capture system poses some major problems. Most capture systems rely on high-resolution off-the-shelf camcorders which cannot save RAW format data. Therefore, the linearity required for calibrating the capture device through RAW data does not exist while capturing with camcorders. This problem can only be resolved if we can calibrate the camcorders without RAW data. The second problem of using this approach in a capture system is that it requires cross-polarization. Cross-polarization can be achieved by using linear polarizer. While using a



polarizer the position of the light source and the camera become crucial for getting complete diffuse surface. This positional constraint makes it hard to position the camera and light source freely to cover the majority of the face and get best possible temporal geometry. Another major drawback of using a polarizer is that, depending on the transmittance of the polarizer it can eliminate a large amount of light. A regular polarizer can have only 30% transmittance. If, the light source is not bright enough, using polarizer can result in failed stereo reconstruction.



Figure 2.8: Using non contact point measurement Jimenez et al. [2010] extract hemoglobin concentration and its change over time. They use this measure to model different expressions. In this image facial appearance for anger has been shown. (Image courtesy of Jorge Jimenez; reprinted Figure 12 partially with permission [Jimenez et al. 2010]; © 2010 ACM Inc)

### 2.2.5 Emotional Appearance Models

There has been some work on emotional appearance where skin has been simulated for different facial expressions. These methods are mostly user guided and without real correlation to actual hemoglobin or melanin maps. Kalra and Magnenat-Thalmann [1994] simulate skin changes from blushing and pallor. Yamada and Watanabe [2007] measure changes in facial skin temperature and color due to anger and dislike. Jung et al. [2009] parameterize skin changes via a set of fourteen emotional states [Plutchik 1980]. Melo and Gratch [2009] simulate blushing by directly applying a user-defined color change in different areas of the face.

For this research our collaborators use a non contact point measure method similar to Matts et al. [2007] for measuring hemoglobin and melanin concentration. As we use camcorders for our capture and off the shelf camcorders cannot store RAW data, we use AVCHD format data for calibration purpose. All non contact chromophore measurement methods use cross-polarization to suppress subsurface information. Instead of using a polarizer we present a novel method that uses six primaries to reconstruct hemoglobin and melanin concentration for facial appearance capture.

## 2.3 Radiometric Calibration of the Capture System

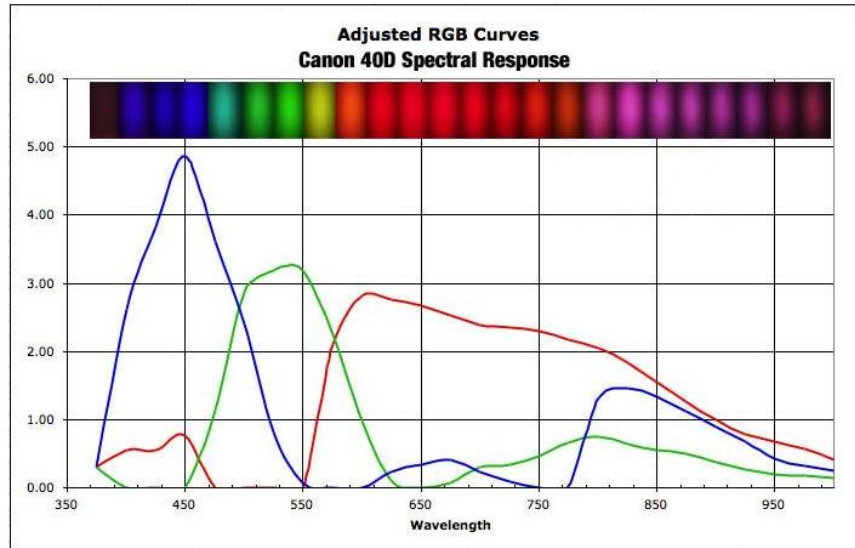


Figure 2.9: Spectral Response Curve of Canon 40D.  
(Image courtesy of [www.MaxMax.com](http://www.MaxMax.com).)

When we capture a scene with a camera, we get an image which is a two dimensional array of “brightness” values. These values are rarely true measurements of relative radiance in the scene. Instead, there is a nonlinear mapping that determines how radiance in the scene becomes pixel values in the image. This nonlinear mapping is the composite of several nonlinear mappings that occur in the photographic process. To obtain scene radiance information from images we need to transform the non-linear mapping into a linear one by calibrating the radiometric response of the camera system or otherwise finding the spectral response curves of the camera system. Figure 2.9 is showing a sample spectral curve of the Canon 40D.

The most common method to measure the spectral response of a camera, which is also EMVA Standard<sup>1</sup>, is to capture the response of the camera under tunable monochromatic illumination [Vora et al. 1997; Farrell et al. 2008]. But this technique requires costly hardware and a considerable amount of manual effort to capture each wavelength. To reduce the effort of such technique other methods have been proposed. These methods use color chart (Figure 2.10) images to recover the spectral response. Hubel et al. [1994] capture the Macbeth color chart (Figure 2.10) using a light source filtered by 16 narrowband and 8 broadband filters. First they recover coarse shape of the spectral response curve by the narrowband illumination then they fine tune it with the broadband illumination.



Figure 2.10: Color Chart (Macbeth SG color chart).

Some spectral response recovering methods exploit the fact that recovering the spectral response of a camera with known spectral reflectance is the same as recovering spectral reflectance from camera responses with known camera characteristics. So, they use least-squares pseudo-inverse methods to estimate spectral reflectance. Vrhel and Trussel [1999], Cheung et al. [2005], Solli et al. [2005] used such methods. Shen and Xin [2004a], [2004b], [2006] performed spectral

---

<sup>1</sup> EMVA, 2010. EMVA Standard 1288, <http://www.emva.org/cms/index.php?idcat=26>.

characterization of scanners using color charts. Ebner [2007] used similar approach but solved for an energy minimization problem.

All these previous methods mentioned here require controlled lighting conditions. Only recently Rump et al. [2011] proposed a method that does not require controlled lighting. But they assume that a spectral characterization of the dominant illuminant is available. The method we use for our calibration is similar to Rump et al. [2011]. We capture the color chart with our camcorders and also with a multispectral camera. The camcorders capture the image in RGB color space while the multispectral camera captures the spectral response of the same color chart over 400 to 720 nm generating 33 wavelength bands. We find the spectral response of our light source by referencing to a known light source. By solving a least square regularization problem we estimate the spectral reflectance of our camcorders.

## **2.4 Facial Performance Capture**

Data driven facial animation was first introduced over two decades ago [Williams 1990; Guenter et al. 1998]. Data driven facial animation involves facial performance capture, where the shape and motion of an actor's face is reconstructed to create realistic facial animations. The performance is usually captured using multiple video cameras. Then frames are extracted from the videos and per frame texture map and geometry is generated, combined and rendered to get the final facial animation.

Current state of the art methods for facial performance capture can largely be divided into the following categories: methods that fit face models to images, methods that use marker or special paints to aid the reconstruction, methods that use structured light, and finally methods that use passive facial performance capture.

### **2.4.1 Fitting Face Models to Images**

One method for capturing the face is to start with a deformable template face and then determine the parameters that best fit the template to images or frames of a video sequence of the actor [Li et al. 1993; Essa et al. 1996; DeCarlo and Metaxas 1996; Pighin et al. 1999; Blanz et al. 2003]. The resulting reconstructed face is very low resolution and lacks details. The template fitting method reconstructs the approximate shape of the face which often does not match with the captured actor.

### **2.4.2 Markers and Face Paint**

A very common method for performance capture is to put marker points (Figure 2.11) on the face of the performer and track the markers or paint over time using video camera. Marker-based face capture was first introduced by Williams [1990]. A large number of black dots or fluorescent colors were used in some later works [Guenter et al. 1998; Lin and Ouhyoung 2005]. Bickel et al. [2007] use markers, face paint, and an initial scan. Furukawa and Ponce [2009] presented a face capture technique that deforms a laser-scanned model to match a highly painted face. These techniques can provide robust tracking and can be used under variety of lighting conditions. But this technique requires manual placement and occasional digital removal of the markers. The resolution of the marker is limited for natural reason and also limits the pore level detail capture.

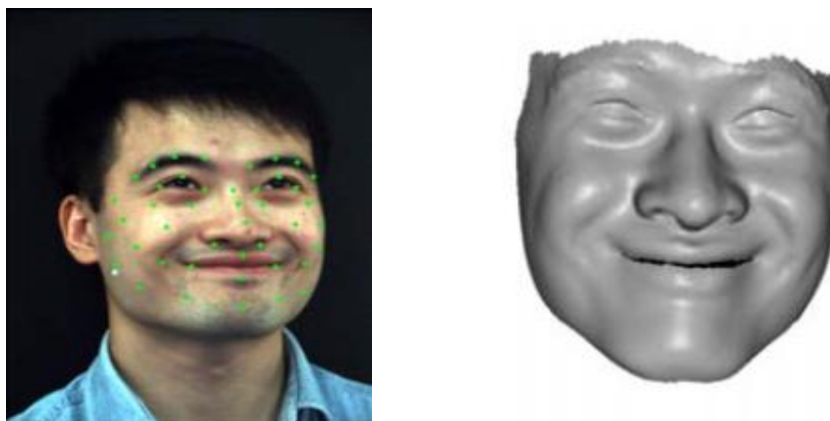


Figure 2.11: Face Markers are used to capture the facial performance of an actor where these markers are treated over time to find the facial deformation (Image courtesy of Li Zhang; reprinted Figure 5 partially with permission [Zhang et al. 2004]; © 2004 ACM Inc)

### 2.4.3 Active Light

Another approach alternative to putting markers or paint on face is to project active light (Figure 2.12) on the actor using one or more projectors in order to provide dense surface texture. Zhang et al. [Zhan 2004] use structured light with space-time stereo to reconstruct maps of a face. Wang et al. [2004] reconstruct 3D shape in real-time by projecting phase-shifted color-fringe patterns onto the face. This approach requires less manual setup unlike the marker based system. This system can be invasive to the actor and not suitable for capturing facial color information and sacrifices temporal resolution.

Hernandez and Vogiatzis [2010] proposed an active light technique using tri-color illumination along with photometric and multi-view stereo to get the facial geometry in real time. By combining active light and markers with a light stage [Ma et al. 2008 ; Alexander et al. 2009] were able to produce high-resolution facial reconstructions.



Figure 2.12: Active light methods project structured light on the actor using projectors to provide dense surface texture (Image courtesy of Christian Benderoth).

Wilson et al. [2010] used changing light conditions of spherical gradient illumination where they combine stereo with photometric normal maps to generate facial geometry. A comprehensive facial performance capture system was recently proposed by Fyffe et al. [2011]. Their proposed system can produce facial geometry along with detailed reflectance information using gradient illumination.

#### **2.4.4 Passive Capture**

A recent focus of facial performance capture research has been on passive reconstruction, which does not require markers or active light. The passive performance capture method used by Bradley et al. [2010b] creates high resolution facial geometry with automatic temporal alignment. The capture method used by Beeler et al. [2010] can reconstruct pore-scale facial geometry for static frames. The reconstructed skin geometry is synthetic and does not rely on facial reflectance. In their recent research [Beeler et al. 2011] presented an anchor based capture



system that produce high-resolution facial geometry along with previously mentioned synthetic detail skin geometry with temporal alignment. The anchor frames which represent similar frames in the capture sequence aids tracker drift, occlusion, and motion blur by tracking similar pixel from reference frame to anchor frame.

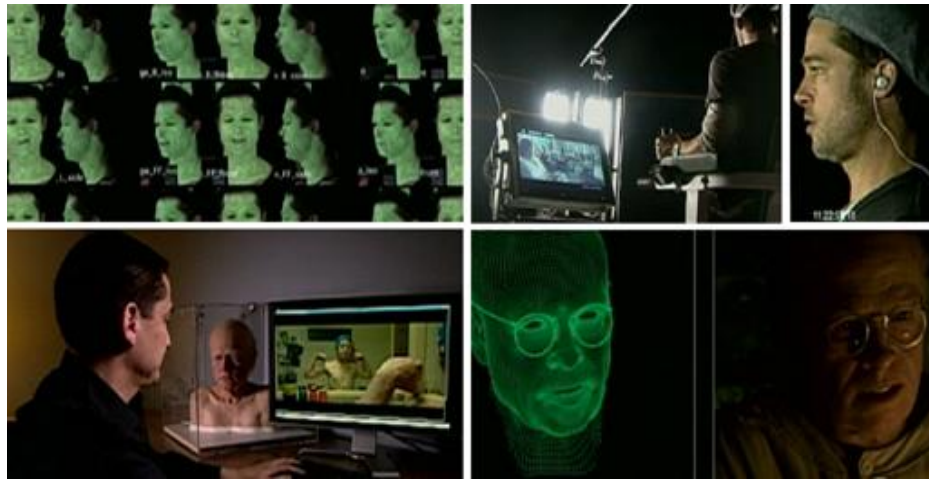


Figure 2.13: Mova's capture system was used in the movie The Curious Case of Benjamin Button. This system uses fluorescent makeup for capture (Photo courtesy of CBS).

## 2.4.5 Commercial Systems

There exist quite a few commercial face capture systems. Mova's CONTOUR Reality Capture<sup>2</sup> is one of the most leading systems available. Mova uses fluorescent makeup. Their system has been used in recent movies like - The Avengers, The Curious Case of Benjamin Button

---

<sup>2</sup> [www.mova.com](http://www.mova.com)

(Figure 2.13), and Harry Potter and the Deathly Hallows. Other commercially available capture systems include Dimension Imaging 3D<sup>3</sup> and Vicon<sup>4</sup> which use a marker based capture system.

For this research we have used a passive facial capture system similar to Bradley et al. [2010b]. We reconstruct the geometry and dense texture of the performance along with the motion. For facial appearance capture we put a filter in front of one of the camera from each stereo pair. We calculate two separate color values, one with and one without filter, for each pixel of the face from the 3D stereo reconstruction data. This modification allows us to capture facial appearance along with facial geometry without any hardware overhead.

---

<sup>3</sup> [www.di3d.com](http://www.di3d.com)

<sup>4</sup> [www.vicon.com](http://www.vicon.com)

## Chapter 3

# Skin Coloring Model from Six Primaries

Human skin coloration is dependent almost exclusively on the concentration and spatial distribution of the skin chromophores melanin and hemoglobin [Anderson et al. 1981]. Research has shown that the facial melanin concentration and distribution is static over hours to weeks [Park et al. 2002]. This leaves hemoglobin as the only varying factor affecting the skin color. This assumption is only justified if the captured skin image is diffuse and does not contain any specular reflection. In practical case, depending on the quality of the skin the specularity of the skin varies from person to person. Some people have oily skin which is highly specular while others have dry skin which is diffuse in property. To assume skin color as a combination of two chromophores all existing hemoglobin extraction methods use cross polarization to capture skin data. This requirement adds two major overheads in the existing capture system. Any cross polarization adds positional constraint on the positions of camera and light source. Again with two layers of polarizer, used for cross polarization, the final incident light needs to be bright enough to capture the detailed surface texture, which is essential for stereo reconstruction. In this work we present a novel method to capture the skin color with existing capture system without the explicit requirement of cross polarization. Our proposed system uses a six primary lookup for skin color reconstruction.

### **3.1 Skin Color Plane from Three Primaries**

As light propagates through skin it is both scattered and absorbed. Scattering mainly occurs in the tissue and the absorption occurs in the tissue pigments. As discussed in chapter 2, skin is considered as a two-layered structure consists of epidermis and dermis. The incoming light passes through the epidermis. The melanin present in this layer absorbs some fraction of the incoming light. The light then passes into the dermis and gets scattered by the collagen and absorbed by the hemoglobin. If the scattering coefficient of collagen and absorption coefficients of melanin and hemoglobin are known, for a specific wavelength remitted light can be predicted.

As shown by Cotton and Claridge [1996] skin color can be reproduced from the concentration information of melanin and hemoglobin and the thickness of the dermal layer. By considering the dermal thickness as constant, the skin color can be defined in a single trajectory within the RGB space. For our current work we have used the skin model [Jimenez et al. 2010] of our collaborators which is based on three primaries.

### **3.2 Hardware Setup for Capturing Skin Color from Three Primaries**

The standard system for capturing skin as a diffuse surface uses essentially conventional camera (capable of storing RAW data), finely calibrated digital camera and lighting system. The light system needs to be fully cross-polarized to eliminate specular reflection (Figure 3.1), and by nature no subsurface information. In this system the camera is treated as a three-waveband spectrometer, using the RGB Bayer filter mosaic over the camera sensors. The spectral distribution of the light source and the raw response of the sensors are determined accurately over the visible range (400–700 nm). For every pixel of the original RAW image concentrations of melanin and hemoglobin are calculated. We faced certain challenges while trying to use the standard appearance capture hardware setup with our passive facial performance capture setup.

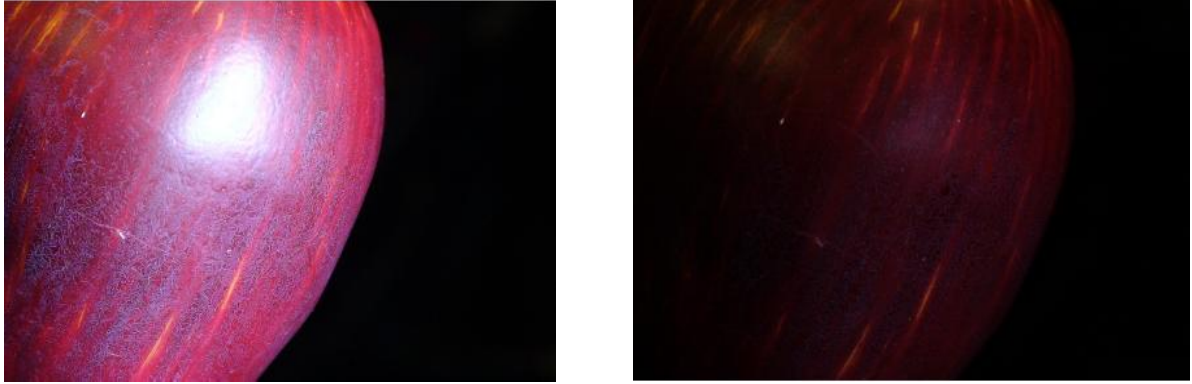


Figure 3.1: Left: Specular reflection. Right: Diffuse surface after cross polarization.

### 3.2.1 Positional Constraint of Camera and Light System

Integrating the cross polarized lighting into the existing capture system requires the camera and the lights lay on the same plane. As passive face capture systems depend on high resolution texture capture and face is not a planer object, to capture the high resolution details we need to put the camera and light to a flexible orientation that covers the whole face and gives uniform illumination all over the face. Directional light also increases the shading effect on the captured image. One way to get rid of specularly without the plane constraint is to use circular polarizer which performs best at the Brewster's angel. But our experiments show that depending on the incident angel the reflected light actually varies color (Figure 3.2). This is not suitable in our case as we need a calibrated lighting system to reproduce the skin color.

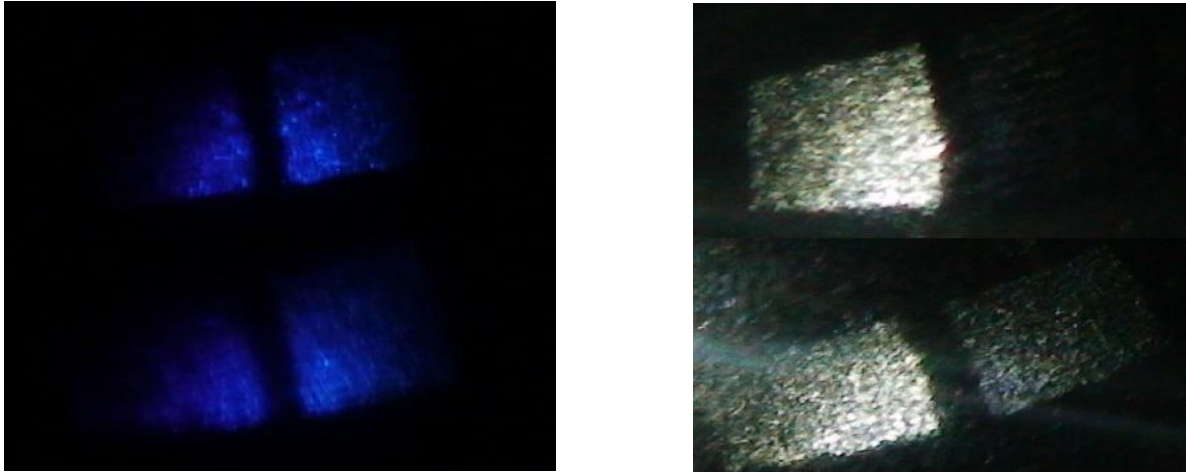


Figure 3.2: Left: Angular effect of using circular polarizer: Depending on the angel of the incident light the reflected light varies color. Right: Angular effect of using linear polarizer: If the camera and the light source are not on the same plane then we have a partially diffuse surface.

### 3.2.2 Energy Reduction by Cross-polarization

Cross polarization requires one set of horizontal polarizer on the light source so the incident light is vertically polarized. This polarized light then falls on a surface and gets reflected. Another set of horizontal polarizer is set on the capture device so the reflected vertical light cannot pass and the final captured surface is completely diffuse (Figure 3.3). This requirement eliminates most of the light from the capture system. Figure 3.4 shows the light source used in our system after vertical polarization. To reproduce the facial geometry we need to capture the surface texture in high resolution and without any added noise or gain and this requires adequate amount of illumination. For this conflict we needed a capture system that does not require cross-polarization but still can capture skin appearance.

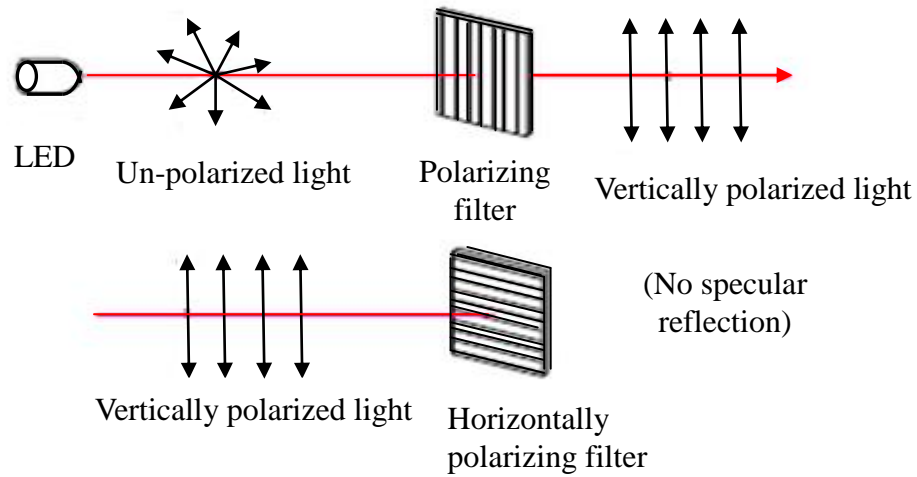


Figure 3.3 Cross-polarization absorbs the specular light energy. At the same time only the diffuse part of the light gets reflected from the surface which is very low in energy.

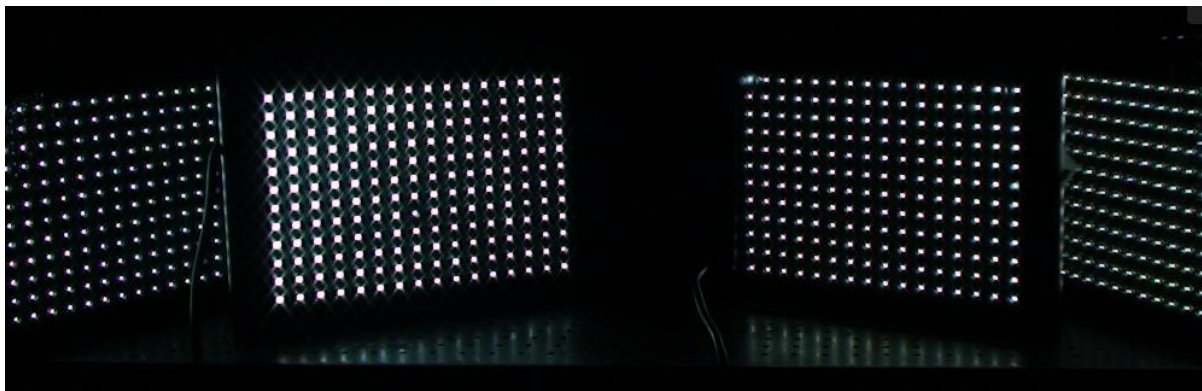


Figure 3.4: Of-the-shelf polarizer has a very low transmittance, usually around 30% which blocks most of the light. Our system uses LED panel as light system. Using polarizer reduces most of the light. As a result the captured surface texture does not contain enough detail for stereo reconstruction.

### 3.3 Reflection Model

With the dichromatic reflection model the image intensity can be expressed as the linear sum of diffuse and specular reflection component [Shafer 1985]. By ignoring camera gain and noise, and assuming the neutral interface reflection, the image intensity can be written as,

$$I(x, \lambda) = w_d(x) S_d(x, \lambda) E(x, \lambda) + w_s(x) S_s(x, \lambda) E(x, \lambda)$$

According to dichromatic reflection model the reflected light  $I(x, \lambda)$  of wavelength  $\lambda$  and at an image position  $x$  is the linear combination of two independent parts, the diffuse and the specular reflections. They depend on geometric parameters,  $w_d(x)$  and  $w_s(x)$ , the diffuse reflectance function  $S_d(x, \lambda)$ , the specular reflectance function  $S_s(x, \lambda)$ , and the light source  $E(x, \lambda)$ . According to the dichromatic reflection model,  $S_d(x, \lambda)$  and  $S_s(x, \lambda)$  depends on the wavelength at a certain position, but are independent of geometry. The geometrical parameters  $w_d(x)$  and  $w_s(x)$  depends only on geometry at a certain position but independent of wavelength. Assuming the color of the illuminant is constant across the scene, the color of the specular reflection can be approximated by the color of the incident light also known as the neutral interface reflection assumption. So we can write the previous equation as,

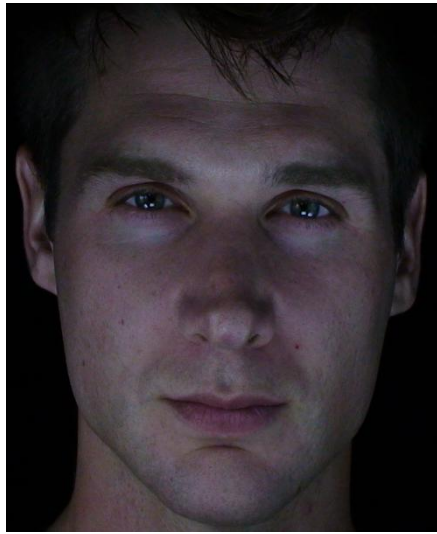
$$I(x, \lambda) = w_d(x) S_d(x, \lambda) E(\lambda) + w_s(x) E(\lambda)$$

The existing method for skin appearance capture does not contain any specular co-efficient. From the dichromatic reflection model it is obvious that the new appearance capture system needs to take into account this fourth degree of freedom, specularity into account to find the skin surface from RGB space.

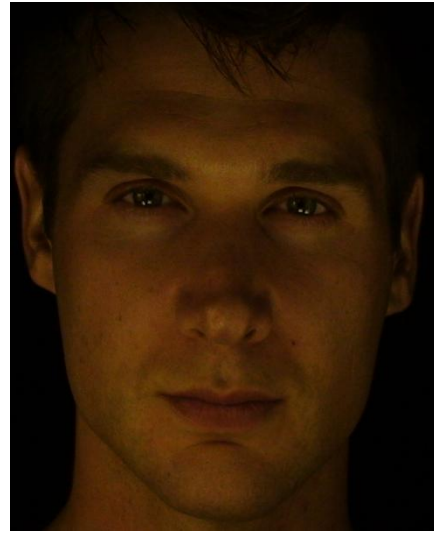


### 3.4 Skin Color Plane from Six Primaries

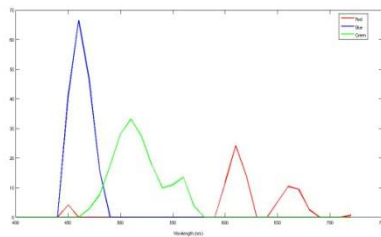
To include specularities in current appearance capture method we propose a novel method that uses six primaries instead of three. With the regular RGB primaries we use a filter to capture the same image, which gives us three shifted primaries  $R_f G_f B_f$ . As our cameras are set as a pair of binocular stereo, this can be done very easily with existing capture pipeline. We add a filter on one of the cameras in each stereo pair. Binocular stereo gives us a location in the 3D along with six primary colors for the same pixel as this point has been seen by two cameras at the same time. Now we find the light source vector by integrating the combined light and camera spectrum for both with and without filter primaries. We then produce two spectral skin textures at the light/camera wavelengths and integrate this with the light and camera spectrum to produce the RGB and  $R_f G_f B_f$  lookup for skin in the camera's color space (Figure 3.5). This allows us to find a mapping between the two skin lookups to Hemoglobin/Melanin values (Figure 3.6). While we used our novel method to capture the skin appearance data, the hemoglobin maps (Figure 3.6) are generated by our collaborators using their skin model [Jimenez et al. 2010]. Because, the camcorders used in our system cannot store RAW data, all the still data are saved as JPEG and all the videos are saved as AVCHD format. The compression effect of the captured images can be seen in the hemoglobin map of figure 3.6. But the effect is reduced in the frames extracted from the capture videos because of the high resolution texture (Figure 3.7).



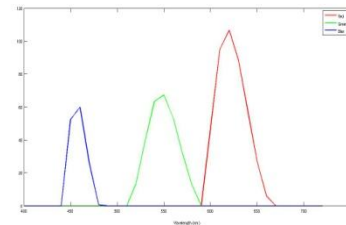
RGB space



$R_fG_fB_f$  space



RGB responses of the camera



$R_fG_fB_f$  responses of the camera

Figure 3.5: Skin data capture using six primaries. Top left: Captured skin data in RGB space. Top right: Captured skin data in  $R_fG_fB_f$  space. Bottom left: RGB spectral response of the camera. Bottom right:  $R_fG_fB_f$  spectral response of the camera.

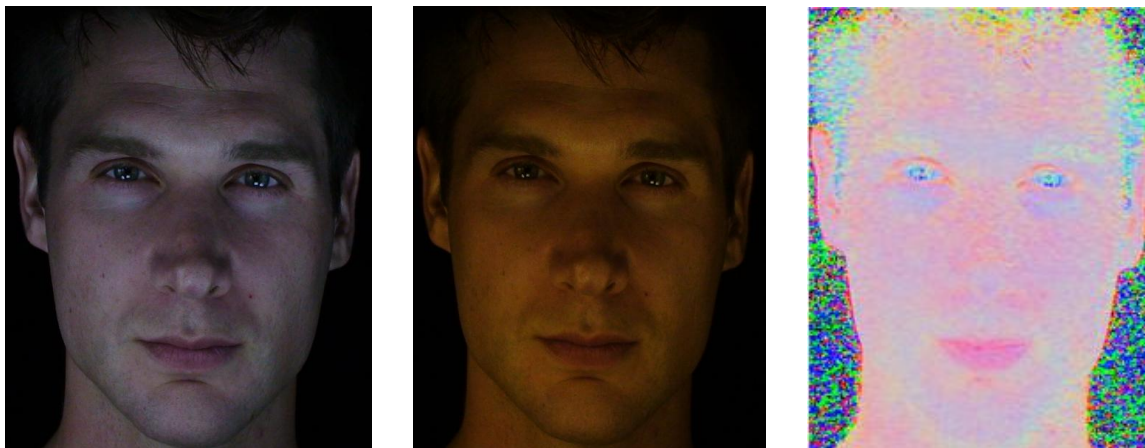


Figure 3.6: Hemoglobin map generated from six primaries using JPEG format image. Left: Skin captured in RGB space, Middle: Skin captured in  $R_fG_fB_f$  space, Right: Hemoglobin map extracted from six primaries.

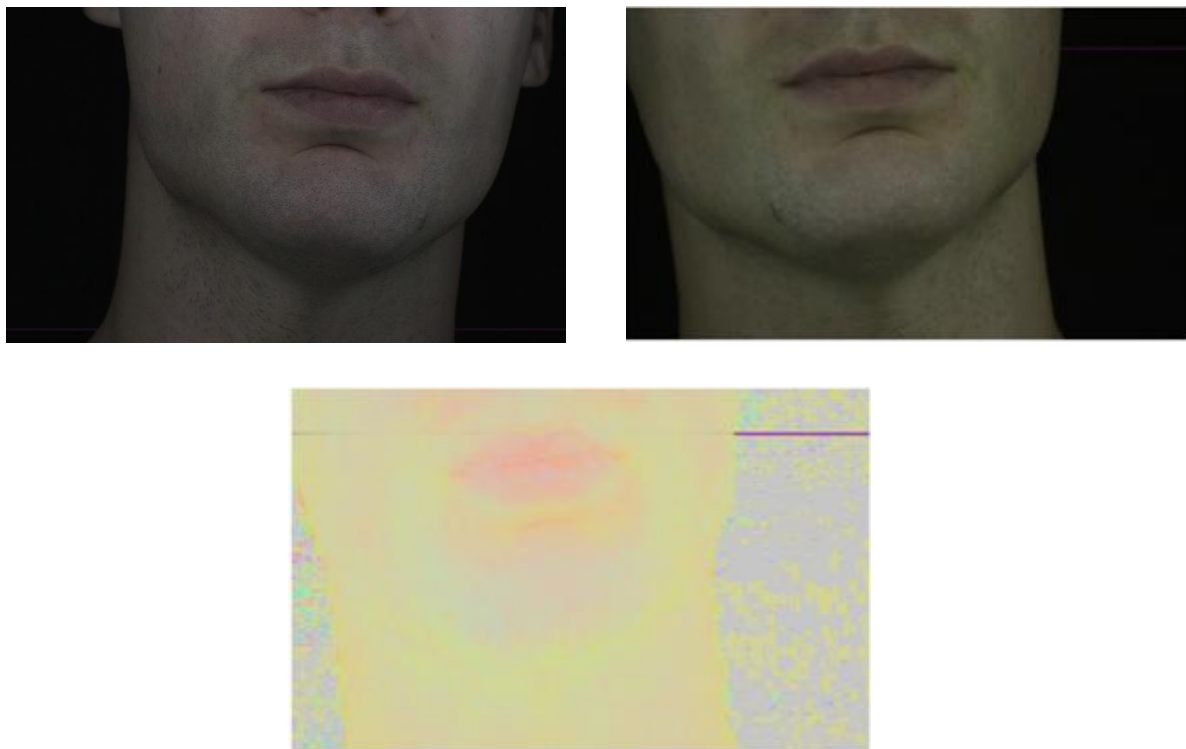


Figure 3.7: Hemoglobin map generated from six primaries using extracted frames of an AVCHD format video. (Top Left) Skin captured in RGB space, (Top Right) Skin captured in  $R_fG_fB_f$  space, (Bottom) Hemoglobin map extracted from six primaries.

Our calibration (Chapter 4) shows that each of the cameras has its unique response curve. So we need to consider the textures generated by each camera pair separately rather than using a common lookup for all the RGB images and all the  $R_fG_fB_f$  images.



Camera pair one



Camera pair two



Camera pair three

Figure 3.8: Three RGB,  $R_f G_f B_f$  texture pairs capturing different parts of the face.

## Chapter 4

# Radiometric Calibration of Capture Device

Facial appearance capture mainly depends on realistic reproduction of skin color. For reproducing the captured skin color we need a device independent representation of captured image. To subtract the effect of a capture device we need to characterize how the device transforms its input. Accurate estimation of spectral response of the capture device is most important in reproducing the color of skin.

Digital cameras or camcorders use charge coupled device (CCD) or complementary metal oxide semiconductor (CMOS) arrays to image the scene. Though the charge collected by a CCD/CMOS is proportional to its irradiance, most digital cameras apply a nonlinear mapping to the CCD/CMOS outputs before storing them. This nonlinear mapping is the composition of several nonlinear mappings that occur in the imaging process. Figure 4.1 shows a generic image acquisition pipeline for a digital camera. The image processing flow can vary between different cameras.

## 4.1 Method Overview

Different factors influence the imaging process of a digital camera and produce an image which has a pixel value  $P(x_i)$  of a pixel  $x_i$ . In an ideal system the spectral response can be written in terms of  $P(x)$  as:

$$P(x) \propto \int_{N_x} C(x, x') \int_{\Lambda} E(\lambda) T_o(\lambda) L_{scene}(\lambda, x') d\lambda dx', \quad (1)$$

where  $\lambda$  represents wavelength,  $\Lambda$  represents respective range of integration,  $L_{scene}$  represent the scene radiance imaged at  $x'$ ,  $T_o$  is the transmittance of the optical system,  $E$  is the quantum efficiency of the sensor and  $C$  the crosstalk between pixels  $x$  and  $x'$  in the neighborhood  $N_x$  of  $x$ . But in reality many of factors such as transmittance of the optical system, quantum efficiency of the sensor are not known unless they are provided by the manufacturer. So instead of measuring individual factors we can calculate the total effect of all unknown factor  $R$  as,

$$P(x) \propto \int_{\Lambda} R(\lambda) L_{scene}(\lambda, x) d\lambda. \quad (2)$$

Here we ignore cross talk which can be easily removed by averaging the neighboring pixels.

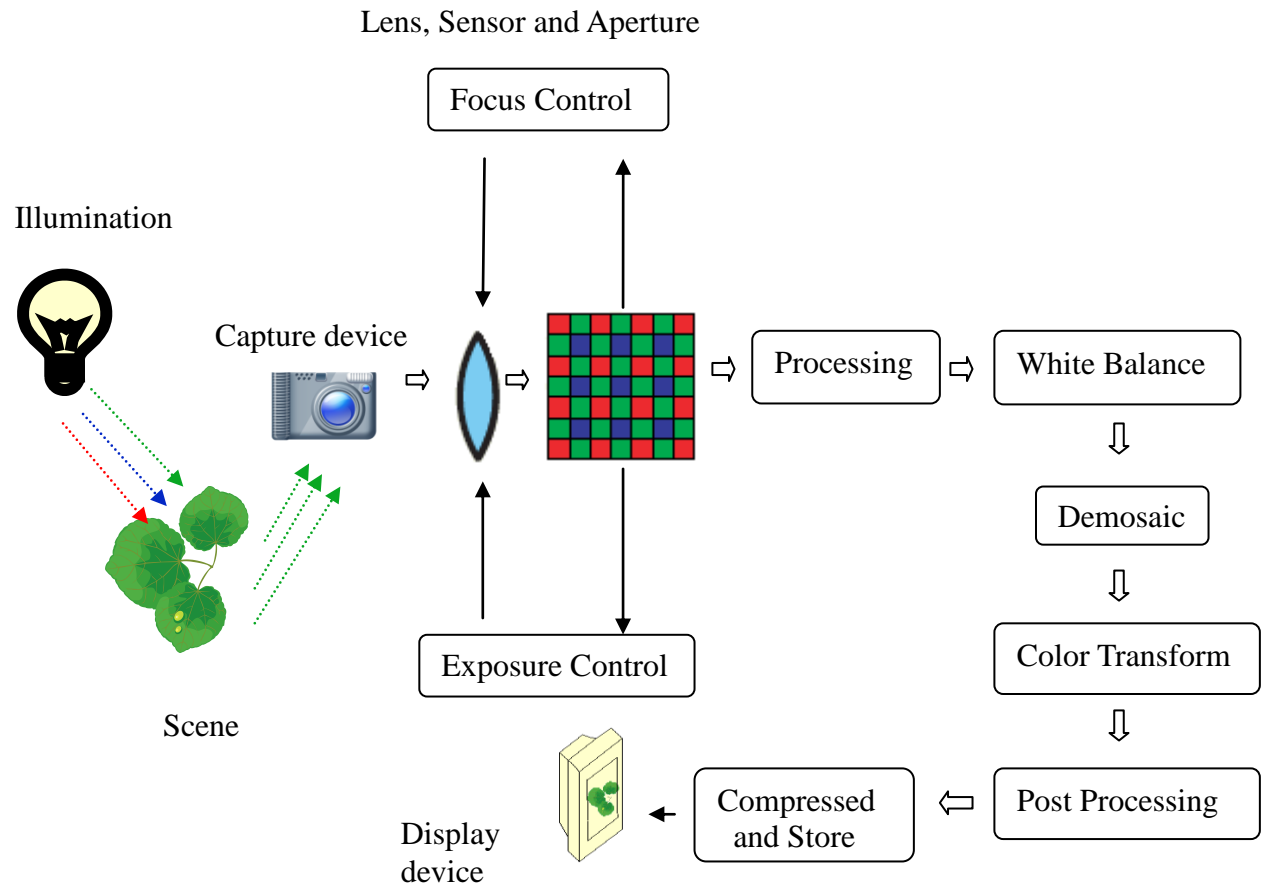


Figure 4.1 : A generic Image Acquisition Pipeline.



In this work we propose a radiometric calibration method for our capture devices which are off-the-shelf camcorders. Using the basic idea from Equation (2) we capture a color chart using a multi spectral camera and a camcorder. The general assumption is, the main non-linear mapping applied by the image processing pipeline is gamma. So after linearizing the digital camera image for gamma we assume a standard mapping exists from spectral to RGB domain. Using captured RGB and spectral color chart data we solve for a linear least-square.

## 4.2 Data Acquisition

To estimate the spectral response of the capture, we need photographs of the color chart as well as the relative spectral power distribution of the illumination. The spectral power distribution is acquired by capturing multispectral data of the color chart. We used an existing multispectral camera for this purpose. For our six primary capture setup, we obtain two sets of calibration data for the multispectral camera, one with a filter and another without filter. In our four pair stereo camera setup four of the cameras have filter in front of them and rest of the four does not have any filter. We captured the color chart using each of these eight cameras.

### 4.2.1 Multispectral Camera

The main components of the multispectral camera sensor used in our system are, a monochrome camera, a Liquid Crystal Tunable Filter (LCTF), an objective lens and an afocal field of view extending optics. It is a Point Grey Chameleon. To capture high dynamic range multispectral images this particular camera can vary both the exposure and wavelength of the camera. The centre wavelength of the pass band of its LCTF can be configured electronically within 400-720 nm range. The transmission band has Gaussian shape and is 10 nm in width. The spectral response of this complete multispectral camera system was calibrated using a reference light source and a factory calibrated spectroradiometer. This device can capture spectral response of the target scene over 400 to 720 nm in 1 nm steps with 1 nm pass band generating 33

wavelength bands. The camera can remove dark frame noise from the raw capture. Taking three low dynamic range images for each channel this particular camera can assemble them into a single high dynamic range image [Robertson et al. 1999].

### 4.2.2 Off-the-shelf Camcorder

As our camcorders cannot save RAW data we use frames of the color chart extracted from the video. To find the affect of not having RAW data we applied the same calibration method using RAW data and found that not having RAW data didn't have much effect on the quality of the reproduced response curve. The camcorders used in our setup comes with minimum user controllable parameters. It does not provide any apparatus control. Exposure and gain is provided on a single radial dial. We always try to keep this value to a point where exposure is maximum and gain is zero with manual inspection. Any auto mode is turned off.

## 4.3 Estimating Spectral Response

The estimation process is to optimize for spectral response curves  $R$  that explain our measurements in best possible manner. In other words the difference between the RGB colors of the color fields captured by the camera and the known reflected spectral radiance calculated from the multi spectral camera and projected on RGB color space using  $R$  is minimized:

$$E(R) = \sum_{c=0}^3 E_d(R_c) + \alpha E_{sm}(R_c)$$

with

$$E_d(R_c) = \sum_{j=0}^J \left\| \sum_{i=0}^k [S_{i,j} L_i R_{i,c}] - D_{j,c} \right\|^2,$$

$$E_{sm}(R_c) = \sum_{i=0}^{k-1} \left\| \sum_{i=0}^k R_{i,c} - R_{i+1,c} \right\|^2.$$

Where,  $R_{i,c}$  are the unknown effective spectral responses for color channel c,

$\lambda_i$  is a discrete wavelength ,

k is the number of spectral bands, which is 33 in our case,

j is the field on the color checker,

J is the total number of color fields which is 152 for us ignoring dark fields,

L is the spectral power distribution of the illuminant,

S is the known reflectance spectra of the color fields

and D is the average camera response to the respective color field in the photograph.

## 4.4 Validation

In our calibration method we consider the whole image acquisition pipe as a black box as we don't have direct control over exposure, aperture and gain of the capture device. To validate our method we have captured our calibration grid with a camera capable of storing RAW data. We generated the response curve of this camera with the RAW data and reconstructed the colors of the color chart. Then we calculated the reprojection error by comparing the difference between the captured data and the reconstructed data (Figure 4.2). This difference gives us the reprojection error. We compared the reconstruction errors of the color grid data for the RAW data capturing camera and our camcorders to validate our calibration on consumer camcorders.

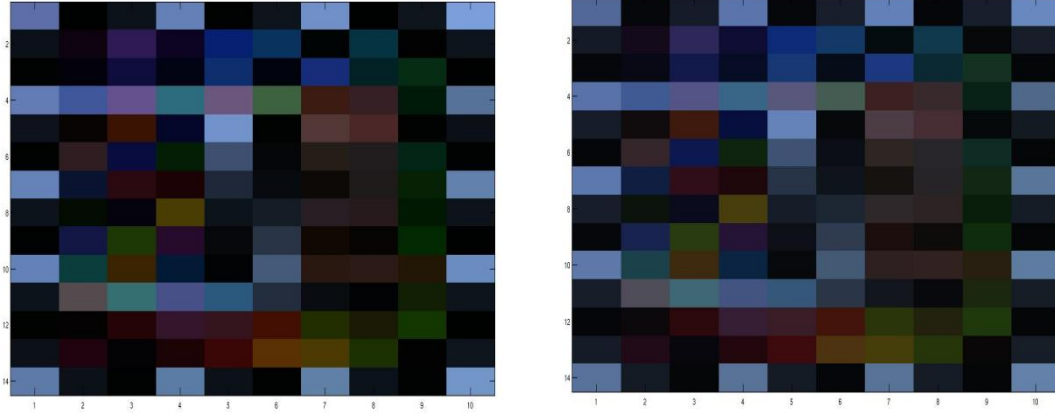


Figure 4.2: Captured RGB values of the color chart (Left).  
Reconstructed RGB values of the color chart (Right).

## 4.5 Result

Using the proposed method we generate spectral response for each camera. The cameras with filter in the capture setup are calibrated with the filters on. Figure 4.3 shows the spectral responses generated for camera number one and three without filter and Figure 4.4 shows the spectral responses generated for camera number two and four with filter.

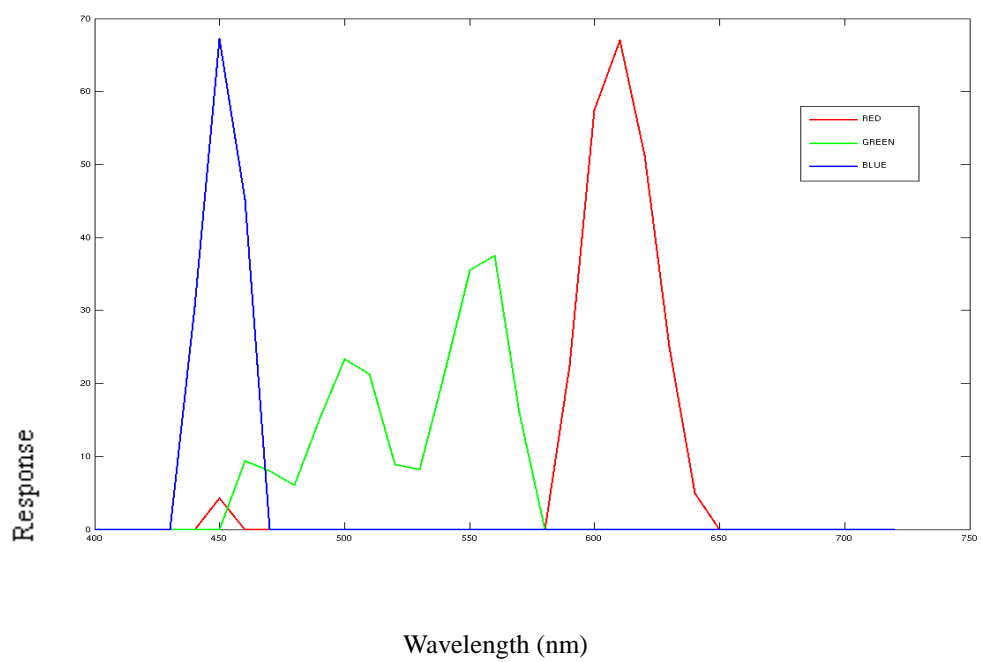
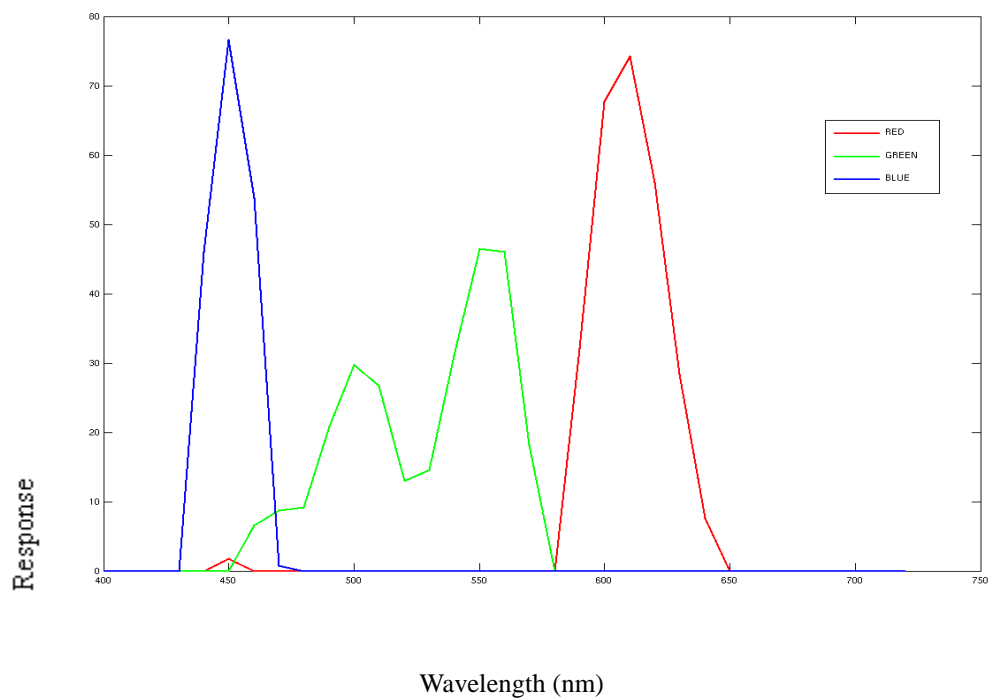


Figure 4.3: Spectral response generated for a camera without filter.  
 Top: Camera one. Bottom: Camera three.

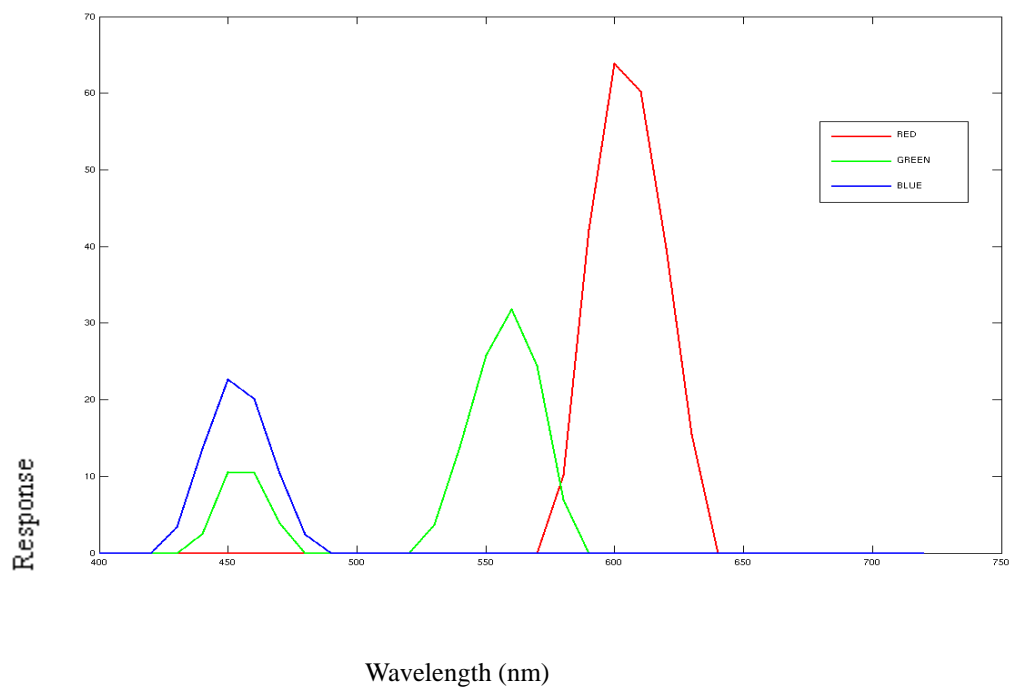
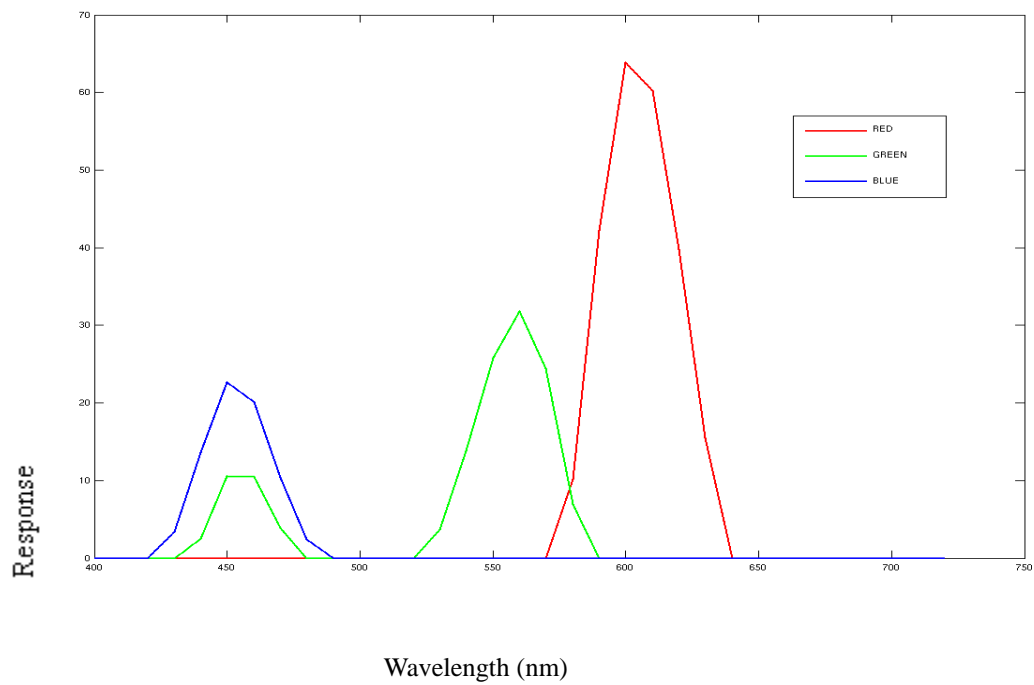


Figure 4.4: Spectral response generated for a camera with filter.  
Top: Camera two. Bottom: Camera four.

## Chapter 5

# Facial Performance Capture

Facial performance capture has evolved into a major tool for creating realistic data driven animations in both movie and game industries. A facial performance capture system can generate sequences of detailed meshes and its geometric deformations over time. The triangulation of the mesh and their mapping between frames generated by the capture system is compatible over time. Facial performance capture enables not only to extract geometry of the actor and deformation of the face over time it can also be used to capture the dynamic facial appearance. As people undergo different emotional feelings like anger, joy, disgust etc. the color of the skin changes. While a person is moving, talking, even drinking alcohol or working out the color of the face changes over time. To realistically reproduce these appearances a capture system requires capturing the color information truthfully. Some capture system does it by generating dynamic texture from the capture sequences. The texture and the mesh can be both edited later by an artist. For our current research we have used a passive marker less capture system [Bradley et al. 2010b]. Our novel method can obtain a high-resolution sequence of compatibly triangulated mesh as well as a high-resolution sequence of facial appearance map without adding any hardware overhead to the existing system.

### 5.1 Acquisition Setup

Our acquisition setup consists of 8 high definition Sony HDR-SR7 camcorders. These camcorders are arranged in an array of four stereo pairs and zoomed to capture the fine-scale

facial details for reconstruction. One camera from each stereo pair holds a color filter. As light sources we use 9 white LED panels arranged in a way that provide uniform illumination on the performer's face. The reason for choosing led panels is that they can be programmed for strobe lighting. We use strobe lighting for camera synchronization which is discussed later in this chapter. The cameras and the lights are controlled using Arduino technology<sup>5</sup> and controlling software. It allows all the cameras in the array to be controlled simultaneously. The control includes basic camera operations. The light panels are connected in a daisy chained fashion and also can be controlled. The basic control for the light panels include on, off and strobe. Our actors do not wear any makeup. We only make sure the skin is clean (no excessive oil is present in the skin) to avoid specularity.

## 5.2 Camera Parameter Setting

The camcorders used in our setup are really cheap and does not provide much control over camera parameters such as aperture, exposure, gain and white balance. These camcorders do not provide any control over the aperture at all. The exposure and gain parameters are linked to a single rotating dial. Once the maximum exposure time is surpassed the gain parameter starts. This parameter is important for stereo reconstruction. Too much added gain reduces the natural skin detail in the captured image. The user cannot see the numeric values of these parameters. The exposure/gain parameters of our cameras are set every time during a capture to maximize the exposure and minimize the gain by manual inspection. The auto white balance is turned off for every camera.

---

<sup>5</sup> [www.arduino.cc](http://www.arduino.cc)



## 5.3 Multi-Camera Synchronization

The cameras need to be temporally synchronized in order to capture time varying objects where individual frames from multiple camera needs to be temporally aligned. Usually machine vision cameras are used for this purpose. But these cameras are expensive and do not have any internal storage. As a result they require an array of computers with expensive hardware to stream the video. Consumer camcorders are cheap, have internal hard drive but they typically don't support hardware synchronization and have rolling shutter effect. Bradley et al. [2009] used strobe lights to synchronize consumer camcorders. They also corrected the rolling shutter distortion. Strobe lights create simultaneous exposures for all cameras. The simultaneous strobe light also removes the rolling shutter effect but the scanlines for a single flash are distributed across two adjacent frames. These two partially exposed frames are used to form a single frame for each camera (Figure 5.1). We have used a similar technique and existing system to synchronize our camcorders.

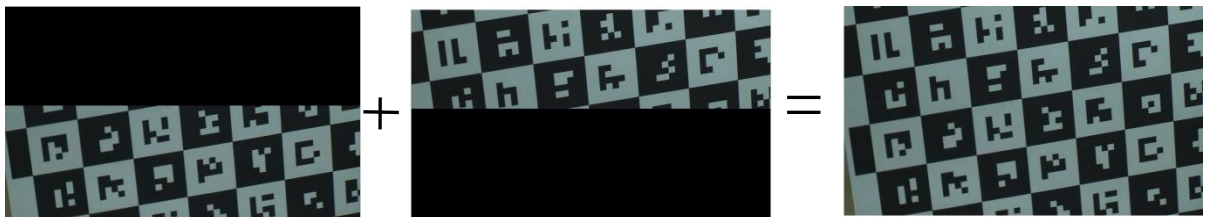


Figure 5.1: The exposed scanlines overlap with a ramp up at the beginning and ramp down at the end. Summing the frames in linear intensity space creates the final image.

## 5.4 Multi-Camera Calibration

Camera calibration or geometric camera calibration is the process to determine the intrinsic and extrinsic parameters of a camera. This is usually accomplished using a calibration pattern or checkerboard. For calibrating our camera we use CALTag pattern [Atcheson et al. 2010] and Zhang's calibration [Zhang 2000] method. We shot the CALTag calibration grid (Figure 5.2) during our capture. The calibration grid is rotated at random in front of the camera. But we make sure to cover rotation and translation of the grid and also covering the entire space. Once the capture is done and the frames are extracted, using these multi calibration frames we extract the projection of known grid points and then solve for the extrinsic and intrinsic camera parameters.

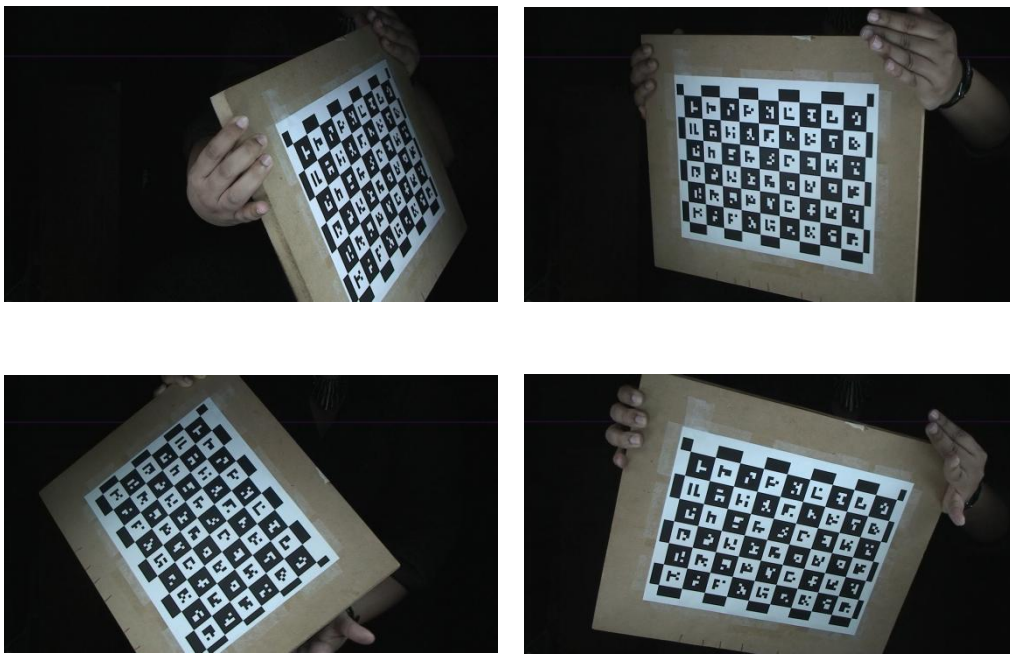


Figure 5.2: CalTag calibration grid is captured such that it covers the entire space.

But this results into one set of intrinsic parameters for the entire sequence, and one set of extrinsic parameters for every single frame. For a single camera calibration usually the extrinsic parameters that minimize the reprojection error of the detected points are used.

Bradley et al. [2010a] show that in a binocular camera setup all checker board locations in all frames can be used to evaluate each set of extrinsic parameters and more accurately determine the calibration for a pair of camera by using reprojection error for the entire volume spanned by the calibration grid over the whole sequence of frames. In order to be a binocular pair while calibrating, a camera pair needs to observe some common points of the calibration grid in the same frame. We calibrate each of our camera pairs using this method and perform binocular stereo reconstruction in pair. Later we combine and align each of these reconstructed surfaces using ICP [Besl and Mckay 1992].

## 5.5 Radiometric Camera Calibration

The ColorChecker<sup>6</sup> chart is used for calibrating the capture devices (Figure 5.3). The SG chart contains 140 and DC chart contains 24 color patches formulated to imitate common natural colors such as skin, sky in addition to additive and subtractive primaries, and a gray scale. We use Both SG and DC chart for more sample points and to compensate for the semi glossiness of the SG chart. Prior to performance capture we capture these two SG and DC charts with our camcorders and as well as with a multispectral camera. We also capture the charts during the performance capture. As we can't control the exact level of exposure for each camera, these captured frames are used as a reference for exposure level or any special variation in the capture setup (Figure 5.4). Using these data we calibrate our camcorders. As the camcorders we use in our setup cannot store RAW data we use AVCHD format data for our calibration. The detail of the radiometric camera calibration method has been discussed in Chapter 4.

---

<sup>6</sup> ColorChecker product No. 50105 (Standard) or product No. 50111 (Mini), manufactured by the Munsell Color services laboratory of Gretag Macbeth.

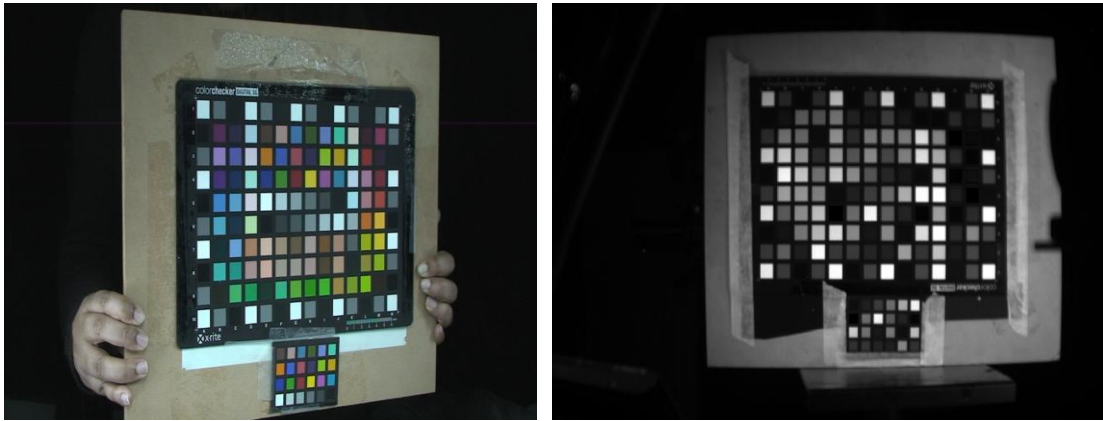
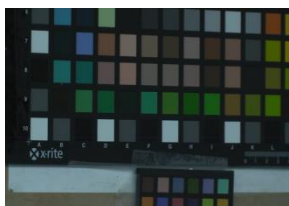


Figure 5.3: Both SG and DC Macbeth ColorChecker charts are captured for radiometric calibration using our camcorders (left) and a multi-spectral camera (right).

Stereo pair: 1



without filter

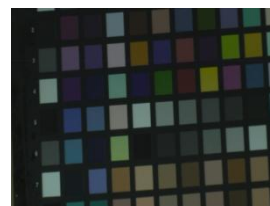


with filter

Stereo pair: 2



without filter

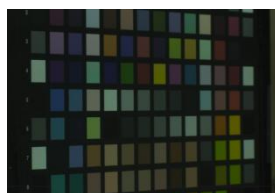


with filter

Stereo pair: 3



without filter



with filter

Stereo pair: 4



without filter



with filter

Figure 5.4: The ColorChecker is captured during the face capture session for calibrating any spectral variation from the original calibration setup.

## 5.6 Multi-View Reconstruction

Our eight cameras capture four different parts of the face and work as four binocular pair. Using binocular stereo for each of the four facial parts we generate four dense point clouds (Figure 5.5). Each of these point clouds contains two separate colors one from the left camera (without filter) and one from the right camera (with filter). These four point clouds are then merged into a single dense point cloud again containing two separate colors and converted to a triangle mesh (Figure 5.6) using an existing multi-view reconstruction system [Bradley et al. 2008a; Bradley et al. 2010b]. As the stereo is run in gradient domain, adding a color filter to one of the cameras in a stereo pair does not affect the stereo reconstruction as long as the transmittance of the filter is relatively low.



Figure 5.5: 4 sets of point clouds are created from each pair of stereo camera. Each point cloud contains two pixel values for same point in 3D, one from the left camera (without filter) and one from the right camera (with filter).

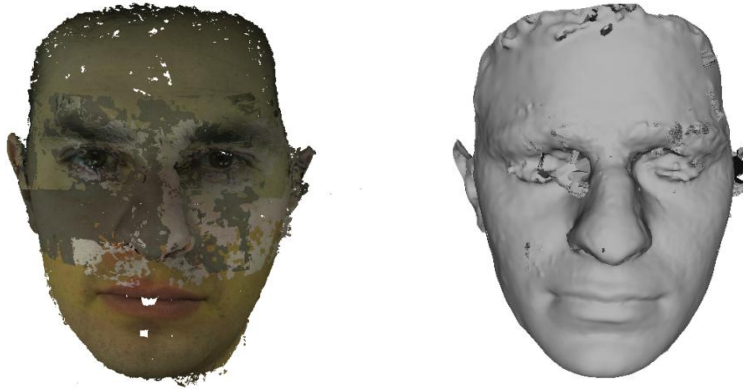


Figure 5.6: Merging four point clouds we generate a single point cloud (left), and convert it to a triangular mesh (Right).

## 5.7 Geometry and Texture Tracking

Once we have per-frame reconstructions, next we reconstruct the motion of the face by tracking the geometry and texture over time using the existing system presented by Bradley et al. [2010b].

### 5.7.1 Reference Mesh

The first step of geometry and texture tracking process is to generate a reference mesh. Each performance of our capture starts with a neutral facial expression. We use this neutral facial geometry to create our template mesh or the reference mesh. The reference mesh is created by cleaning the neutral face geometry for outliers. Then we calculate the 2D parameterization of the geometry using ABF++ [Sheffer et al. 2005]. The 2D parameterization is later used for texture generation. We fill the holes in the mesh by creating small Delaunay triangulations [Shewchuk 1996] and cut a slit in the mesh for mouth.



### 5.7.2 Frame Propagation

We compute optical flow [Bouguet 1999] for all the cameras separately and over the whole capture sequence. Once we have per-frame reconstructions from the multi-view reconstruction stage we use this optical flow and the initial reconstructions to propagate the reference mesh in time. We look up the 2D optical flow vector that corresponds to a pixel in a camera and add the flow to get a new pixel location. Back projecting from this new pixel location onto the initial mesh gives us a new vertex location. In this way we generate a set of compatible meshes that have the same connectivity and exact vertex correspondence (Figure 5.7).

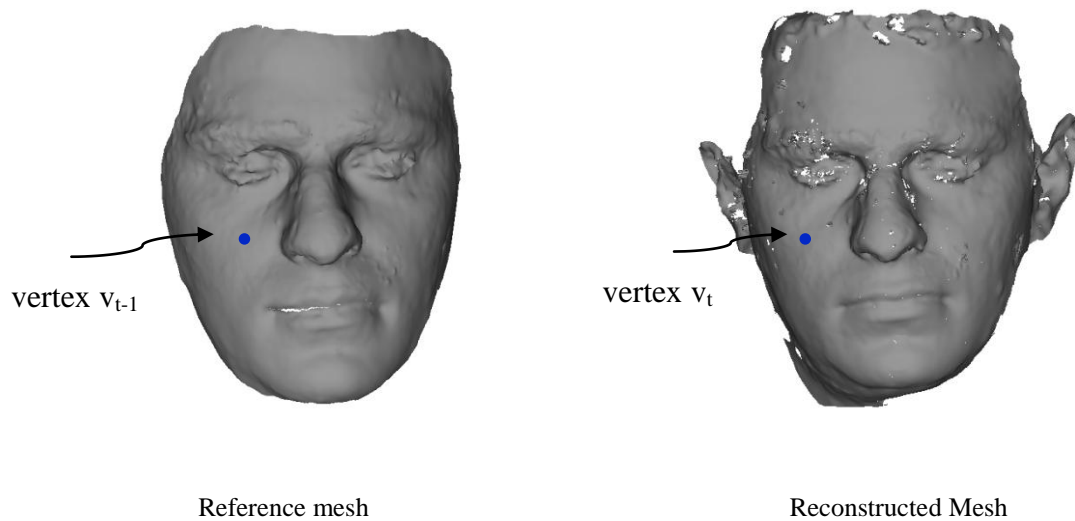


Figure 5.7: Computing vertex positions for the next frame, using per camera optical flow.

### 5.7.3 Computing 2D Texture

For each frame in the capture sequence along with reconstructed geometry we compute a high-resolution 2D texture that covers the entire surface. We create two separate textures for the same sequence. Using all the left side camera frames (with no filter) we create the first texture and the second texture is created using other four right side camera frames (with filter). Every vertex of the 3D reconstructed mesh has a unique 2D coordinates in the parameterized domain. To compute the texture for a specific frame, we start by projecting each triangle of the mesh onto the camera that observes it best determined by the dot product between the triangle normal and the camera direction. Then we copy the camera pixels those correspond the projection to the corresponding texture domain. To track the contribution camera pixel we generate two separate gray scale color coded textures. We compute the final texture using Poisson image editing [Pérez et al. 2003]. Figure 5.8 shows the 2D face texture from left and right cameras and their corresponding contribution map.

### 5.7.4 Smoothing

Spatial noise can appear in the reconstructed geometry while running any stereo algorithm. We smooth the reconstructed face meshes to remove this noise. Smoothing often removes the special features that define the face. We use a saliency based smoothing [Bradley et al. 2010b], which smoothes less-salient regions of the face while preserving more-salient features (Figure 5.9).

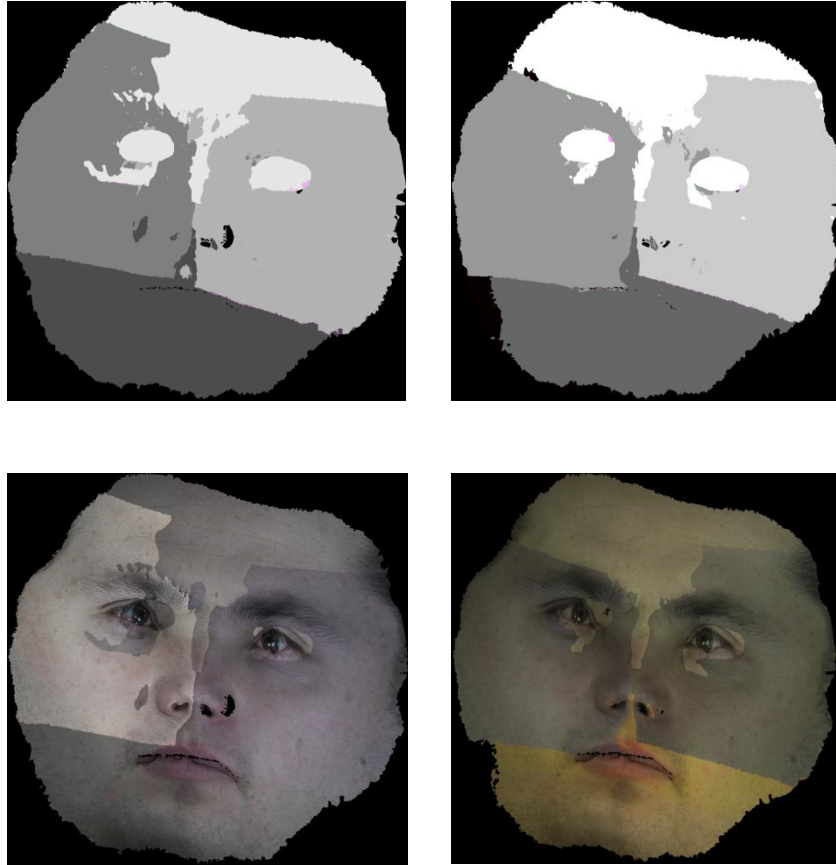


Figure 5.8: 2D texture and color coded contribution map. Top left: Contribution map for left cameras. Top right: Contribution map for right cameras. Bottom left: 2D texture from left cameras. Bottom right: 2D texture from right cameras.

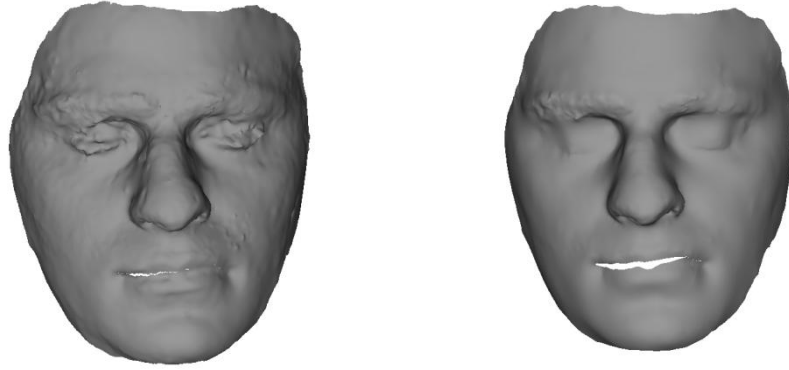


Figure 5.9: Left mesh: Before smoothing, Right mesh: After smoothing.

## 5.8 Output Capture Sequence

Using our passive facial capture system we can generate high resolution geometry and texture of the performer. For each frame of our capture we generate a high resolution 3D geometry, two textures of the face one with and another without a filter, and two color coded camera contribution map for with and without filter generated textures (Figure 5.10).

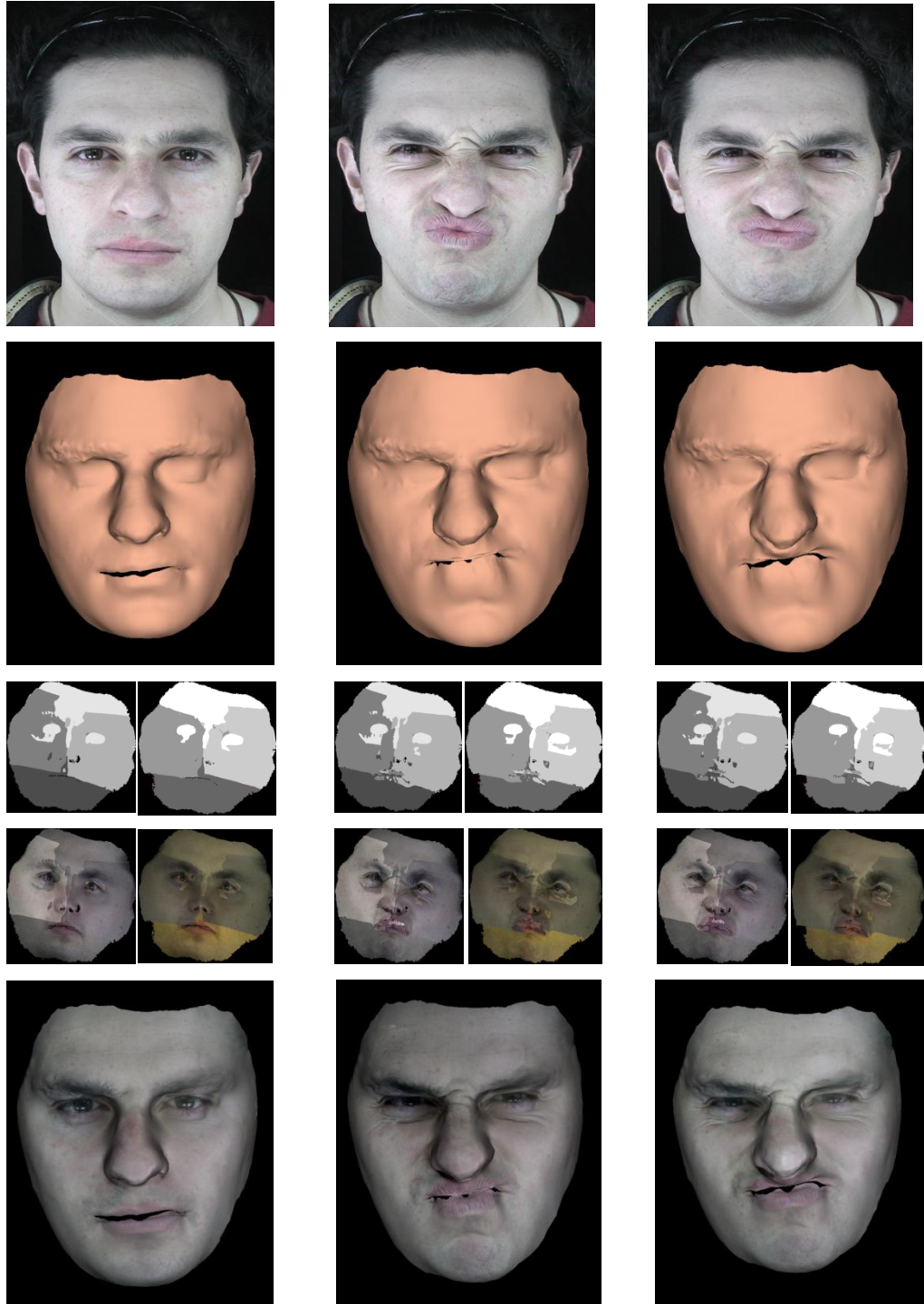


Figure 5.10: Output Capture Sequences. Top row: Reference frames, 2<sup>nd</sup> row: Reconstructed geometry, 3<sup>rd</sup> row: Contribution image of left and right cameras, 4<sup>th</sup> Row: Texture image from left and right cameras, Bottom row: Rendered output.

## Chapter 6

# Results

To capture dynamic facial appearance we needed to capture skin color change over time. Facial appearance change is visible during facial expressions, for example people tend to get red when they become angry. This is difficult to achieve without an experienced actor. To capture realistic facial appearance we chose to capture basic six expressions: anger, joy, disgust, sadness, fear and surprise. We chose seven subjects mostly Caucasian, ageing from 20 to 35 years. They all performed the six basic expressions (Figure 6.1) all starting from a neutral face. Though the performance varied from subject to subject, they all managed to show the basic appearance change realistically. Our subjects did not wear any makeup. Each of them was asked to clean their face before the capture to handle any specularities caused by oily skin.

For all the subjects and their expressions we reconstructed the facial geometry along with six primary textures. Figure 6.2 shows the expression changes for a specific subject from neutral to joy. The top row shows some frames from the performance sequence. The 2<sup>nd</sup> row shows the reconstructed geometry. The 3<sup>rd</sup> row shows the camera contribution texture for left cameras, 4<sup>th</sup> row shows the camera contribution texture for the right cameras, 5<sup>th</sup> row shows the texture from different left contribution cameras and bottom row shows the texture from different right contribution cameras. Figure 6.4, Figure 6.5 and Figure 6.6 show the results from capture sequences consecutively showing expressions Anger, Sad and Fear.

We also captured multi-spectral facial skin data of different subjects from variety of ethnicity (Figure 6.3). Using these data along with the response of the filters we found the best filter that shifts the three new primaries ( $R_f G_f B_f$ ) most from RGB primaries. The transmittance of the filters were another factor while considering the specific filter. Low transmittance rate of the filter, like

the previously mentioned polarizer, reduces incoming energy and results in a failed stereo reconstruction. That's why we selected our filter by optimizing the light transmittance value along with the most shifted primaries.

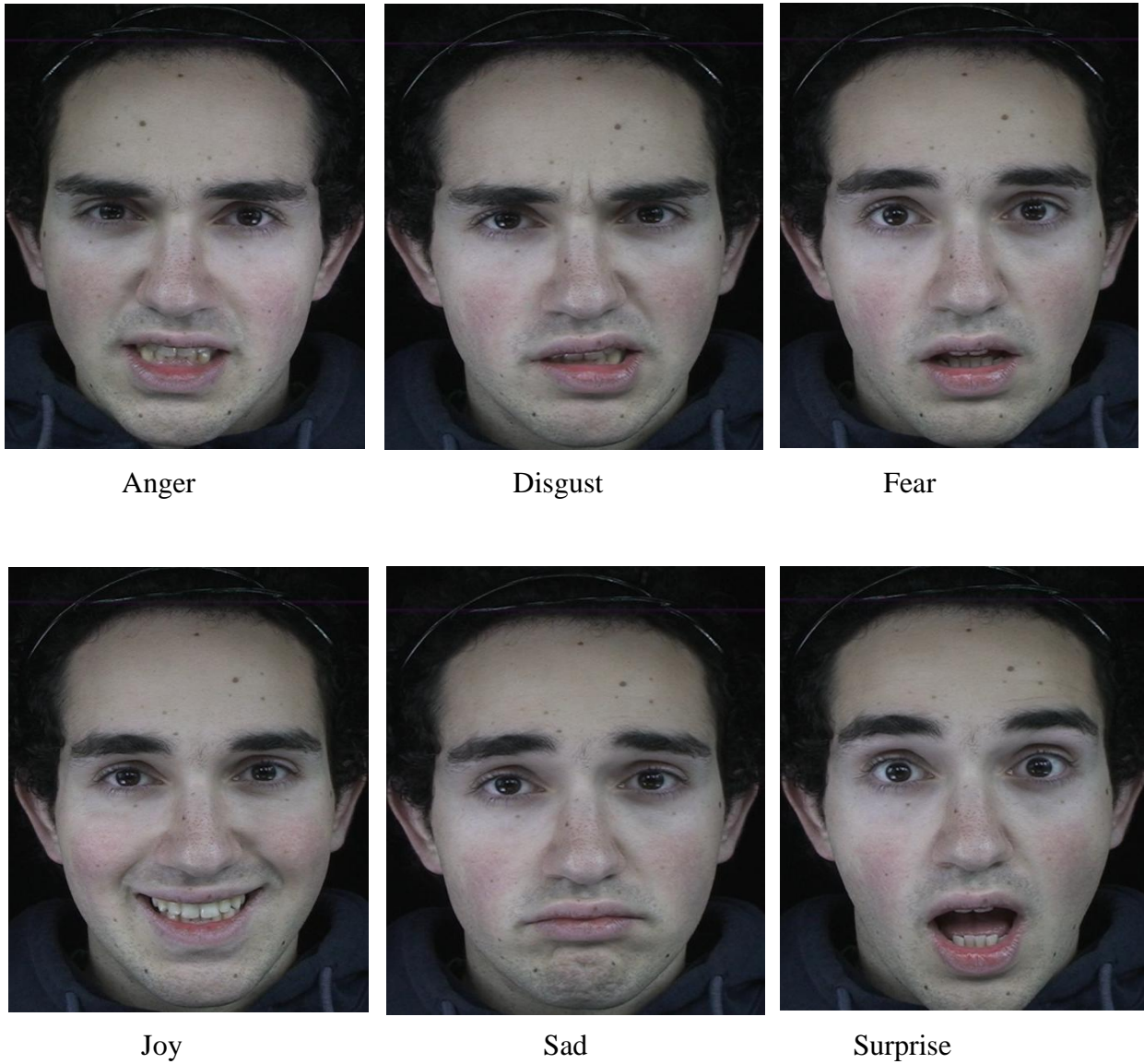


Figure 6.1: Six basic expressions.





Figure 6.2: Results from a sequence capture of the expression “Joy”. Top row: Some frames from the performance sequence. 2<sup>nd</sup> row: Reconstructed geometry. 3<sup>rd</sup> row: Camera contribution texture for left cameras, 4<sup>th</sup> row: Camera contribution texture for the right cameras, 5<sup>th</sup> row: Texture from different left contribution cameras. Bottom row: Texture from different right contribution cameras.



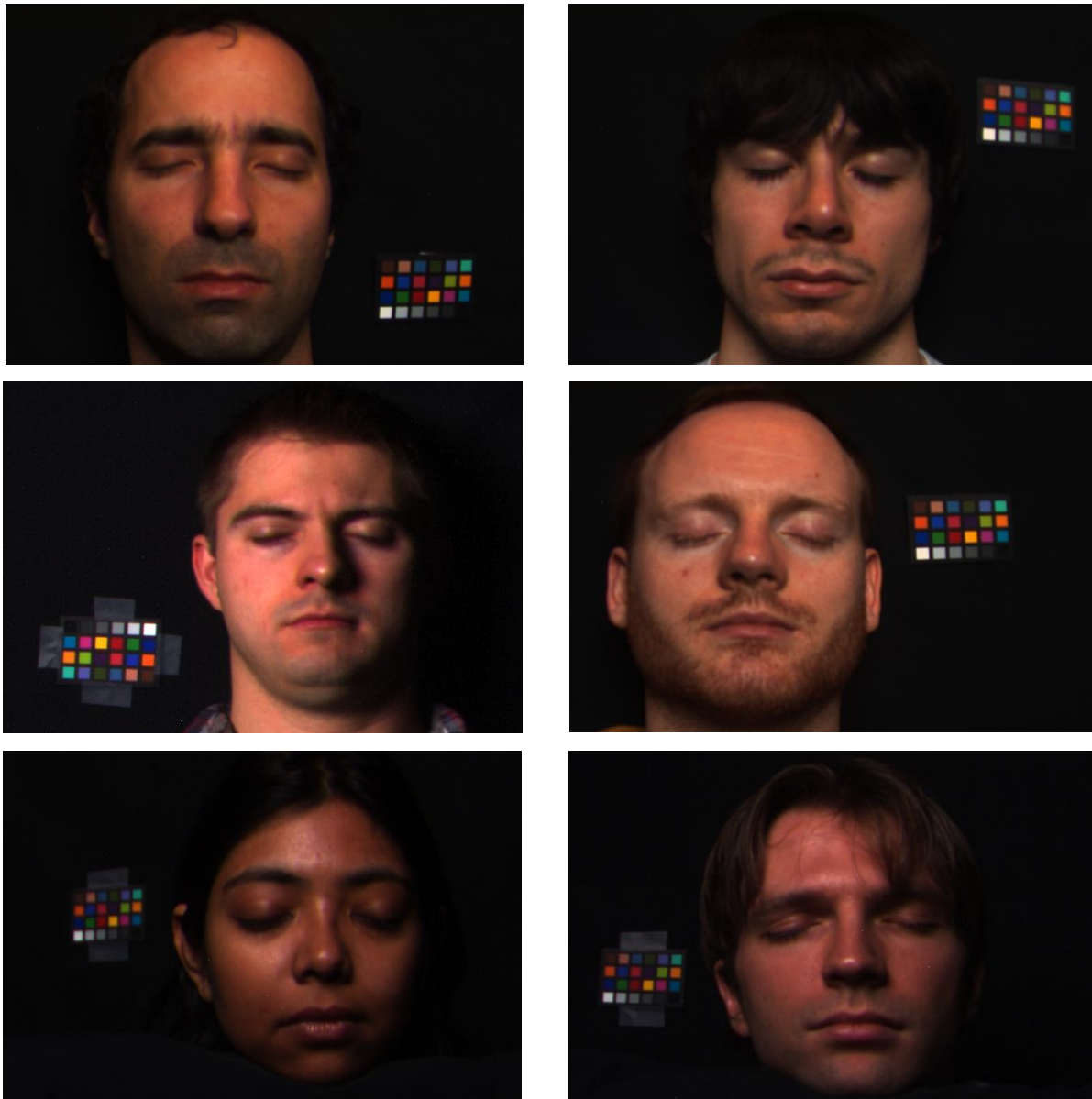


Figure 6.3: Multi-spectral facial skin capture of different subjects from variety of ethnicity. Using these data along with the response of the filters we found the best filter that shifts the three new primaries ( $R_f G_f B_f$ ) most from RGB primaries.



Figure 6.4: Results of the expression “Anger” capture. Top row: Some frames from the performance sequence. 2<sup>nd</sup> row: Reconstructed geometry. 3<sup>rd</sup> row: Texture from different left contribution cameras. Bottom row: Texture from different right contribution cameras.



Figure 6.5: Results of the expression “Sad” capture. Top row: Some frames from the performance sequence. 2<sup>nd</sup> row: Reconstructed geometry. 3<sup>rd</sup> row: Texture from different left contribution cameras. Bottom row: Texture from different right contribution cameras.





Figure 6.6: Results of the expression “Fear” capture. Top row: Some frames from the performance sequence. 2<sup>nd</sup> row: Reconstructed geometry. 3<sup>rd</sup> row: Texture from different left contribution cameras. Bottom row: Texture from different right contribution cameras.

## Chapter 7

# Conclusion and Future Work

In this work we have presented a novel technique to integrate existing facial performance capture system with facial appearance capture without adding any overhead on the existing performance capture pipeline. We proposed a six primary capture system that represents the human skin color in a six primary space. Extra three primaries add extra dimensionality to include specularly. As a result we can capture the facial appearance change over time by extracting the concentration of hemoglobin from the captured data without cross-polarization. To our knowledge our facial capture system is the first to capture facial appearance with off-the-shelf camcorders. We demonstrate that even without the RAW Bayer pattern data the capture device can be radiometrically calibrated.

The main limitation of our system is the skin model we have used in our system turns out to be not so robust after taking specularly into consideration. Though we assumed the directionality of the light from our light panels is uniform but in practice that is not always true. Our camcorders tend to add more gain after it crosses the maximum exposure time. As there is no exact indication on the device when we have reached the maximum exposure level, our manual inspection sometimes tends to be wrong. This leaves our system to be more error prone.

In spite of the lack of robustness of the skin model, our proposed method can be used very easily with existing capture systems and has more scope to improve in future. For future one way to improve our system would be to incorporate a more robust skin model. We could also take the image analysis method to separate melanin and hemoglobin [Tsumura et al. 1999], rather than depending on physiological measurement, as for facial appearance we do not need fine accuracy. This could give us the initial guess about our lookup surface. As we can track the texture color

over time, we could also regularize our skin surface lookup from its initial color. This would give us an optimum color for directional color variance.

# References

ALEXANDER, O., ROGERS, M., LAMBETH, W., CHIANG, M., AND DEBEVEC, P. 2009. “The Digital Emily project: photoreal facial modeling and animation”. In: *ACM SIGGRAPH Courses*, pp.1-15.

ANDERSON, R. R., AND PARRISH, J. A. 1981. The optics of human skin. *Journal of Investigative Dermatology*, 77,1,13–19.

ATCHESON, B., HEIDE, F., AND HEIDRICH, W. 2010. CALTag: High Precision Fiducial Markers for Camera Calibration. *Vision, Modeling, and Visualization (VMV)*.

BEELER, T., BICKEL, B., SUMNER, R., BEARDSLEY, P., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 40.

BEELER, T., HAHN, F., BRADLEY, D., BICKEL, B., BEARDSLEY, P., GOTSMAN, C., SUMNER, R. W., AND GROSS, M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 75.

BESL, P. J., AND MCKAY, N. D. 1992. A method for registration of 3-d shapes. *IEEE Trans. on PAMI*, Vol. 14, No. 2, pp. 239–256.

BICKEL, B., BOTSCH, M., ANGST, R., MATUSIK, W., OTADUY, M., PFISTER, H., AND GROSS, M. 2007. Multi-scale capture of facial geometry and motion. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 33.

BLANZ, V., BASSO, C., VETTER, T., AND POGGIO, T. 2003. Reanimating faces in images and video. *Computer Graphics Forum (Proc. Eurographics)* 22, 3, 641–650.

BOUGUET, J.Y. 1999. Pyramidal implementation of the lucas kanade feature tracker: Description of the algorithm. *Tech. rep., Intel Corporation, Microprocessor Research Labs.*

BRADLEY, D., ATCHESON, B., IHRKE, I., AND HEIDRICH, W. 2009. Synchronization and rolling shutter compensation for consumer video camera arrays. In *International Workshop on Projector-Camera Systems (PROCAMS)*.

BRADLEY, D., AND HEIDRICH, W. 2010. Binocular Camera Calibration Using Rectification Error. *IEEE Conference on Computer and Robot Vision (CRV)*, p.

BRADLEY, D., HEIDRICH, W., POPA, T., AND SHEFFER, A. 2010. High resolution passive facial performance capture. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 41.

CHEUNG, V., WESTLAND, S., LI, C., HARDEBERG, J., AND CONNAH, D. 2005. Characterization of trichromatic color cameras by using a new multispectral imaging technique. *J. Opt. Soc. Am. A* 22, 7 (Jul), 1231–1240.

COTTON, S. D., AND CLARIDGE, E. 1996. Developing a predictive model of human skin colouring. In *Proceedings of the SPIE Medical Imaging 1996*, vol. 2708, 814–825.

COTTON, S. D., CLARIDGE, E., AND HALL, P. N. 1999. A skin imaging method based on a colour formation model and its application to the diagnosis of pigmented skin lesions. In *Proceedings of Medical Image Understanding and Analysis '99*, 49–52.

CULA, O., DANA, K., MURPHY, F., AND RAO, B. 2004. Bidirectional Imaging and Modeling of Skin Texture. *IEEE Trans. On Biomedical Engineering*, Vol. 51, No. 12 (Dec.), 2148–2159.

CULA, O., DANA, K., MURPHY, F., AND RAO, B. 2005. Skin Texture Modeling. *International Journal of Computer Vision*, Vol. 62, No. 1–2 (April–May), 97–119.



- DECARLO, D., AND METAXAS, D. 1996. The integration of optical flow and deformable models with applications to human face shape and motion estimation. In *Proc. CVPR*, 231–238.
- DONNER, C., AND JENSEN, H. W. 2006. A Spectral BSSRDF for Shading Human Skin. In *Rendering Techniques (Proc. EGSR)*, 409–417.
- DONNER, C., WEYRICH, T., D'EON, E., RAMAMOORTHY, R., AND RUSINKIEWICZ, S. 2008. A Layered, Heterogeneous Reflectance Model for Acquiring and Rendering Human Skin. In *Trans. on Graphics (Proc. SIGGRAPH Asia)*, Vol. 27, 10:1– 10:12.
- EBNER, M. 2007. Estimating the spectral sensitivity of a digital sensor using calibration targets. In *Proceedings of the 9<sup>th</sup> annual conference on Genetic and evolutionary computation, ACM, New York, NY, USA, GECCO '07*, 642–649.
- ESSA, I., BASU, S., DARRELL, T., AND PENTLAND, A. 1996. Modeling, tracking and interactive animation of faces and heads using input from video. In *Proc. Computer Animation*, 68.
- FARRELL, J., OKINCHA, M., AND PARMAR, M. 2008. Sensor calibration and simulation. In *Proceedings of the SPIE, Digital Photography IV, SPIE, J. M. DiCarlo and B. G. Rodricks, Eds., no. 1, 68170R*.
- FYFFE, G., HAWKINS, T., WATTS, C., MA, W. C., AND DEBEVEC, P. 2011. Comprehensive facial performance capture. *Comp. Graphics Forum (Proc. Eurographics)* 30, 2, 425–434.
- GHOSH, A., HAWKINS, T., PEERS, P., FREDERIKSEN, S., AND DEBEVEC, P. 2008. Practical Modeling and Acquisition of Layered Facial Reflectance. In *Transactions on Graphics (Proc. SIGGRAPH Asia)*, Vol. 27, 9:1–9:10.
- GUENTER, B., GRIMM, C., WOOD, D., MALVAR, H., AND PIGHIN, F. 1998. Making faces. In *Comp. Graphics*, 55–66.

- HERNÁNDEZ, C., AND VOGIATZIS, G. 2010. Self-calibrating a real-time monocular 3D facial capture system. In *Proceedings International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*.
- HUBEL, P. M., SHERMAN, D., AND FARRELL, J. E. 1994. A comparison of methods of sensor spectral sensitivity estimation. In *Second Color Imaging Conference: Color Science, Systems, and Applications*, 45–48.
- JIMENEZ, J., SCULLY, T., BARBOSA, N., DONNER, C., ALVAREZ, X., VIEIRAK, T., MATTS, P., ORVALHO, V., GUTIERREZ, D., AND WEYRICH, T. 2010. A Practical Appearance Model for Dynamic Facial Color. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 29, 6, 141.
- JUNG, Y., WEBER, C., KEIL, J., AND FRANKE, T. 2009. Real-Time Rendering of Skin Changes Caused by Emotions. In *Proc. of the 9th International Conference on Intelligent Virtual Agents (IVA)*, Springer-Verlag, Berlin, Heidelberg, 504–505.
- KALRA, P., AND MAGNENAT-THALMANN, N. 1994. *Modelling of vascular expressions in facial animation*. In *Proc. of Computer Animation*, 50–58.
- LEVER, W. F., AND SCHAUMBURG-LEVER, G. 1990. Histopathology of the Skin. *J. B. Lippincott Company, seventh edition*.
- LI, H., ROIVAINEN, P., AND FORCHEIMER, R. 1993. 3-D motion estimation in model-based facial image coding. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 6, 545–555.
- LIN, I. C., AND OUHYOUNG, M. 2005. Mirror MoCap: Automatic and efficient capture of dense 3D facial motion parameters from video. *The Visual Computer*, Vol. 21, No. 6, pp. 355–372.
- MA, W. C., JONES, A., CHIANG, J. Y., HAWKINS, T., FREDERIKSEN, S., PEERS, P., VUKOVIC, M., OUHYOUNG, M., AND DEBEVEC, P. 2008. Facial performance synthesis using deformation-driven polynomial displacement maps. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 27, 5, 121.

- MATTS, P. J., DYKES, P. J., AND MARKS, R. 2007. The distribution of melanin in skin determined in vivo. *British Journal of Dermatology*, Vol. 156, No. 4, 620–628.
- MEGLINSKY, I. V., AND MATCHER, S. J. 2001. Modelling the sampling volume for skin blood oxygenation. *Medical & Biological Engineering & Computing*, 39:44–49.
- MELO, C., AND GRATCH, J. 2009. Expression of Emotions Using Wrinkles, Blushing, Sweating and Tears. In *Intelligent Virtual Agents: 9th International Conference, Springer, Ed.*, 188–200.
- PARK, S. B., HUH, C. H., CHOE, Y. B., AND YOUN, J. I. 2002. Time course of ultraviolet-induced skin reactions evaluated by two different reflectance spectrophotometers: *Derma SpectrophotometerR* and *Minolta spectrophotometer CM-2002R*. *Photodermatology, Photoimmunology & Photomedicine*, Vol. 18, 23–28.
- PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. Graph.* 22, 3, 313–318.
- PIGHIN, F. H., SZELISKI, R., AND SALESIN, D. 1999. Resynthesizing facial animation through 3D model-based tracking. In *Proc. ICCV*, 143–150.
- PLUTCHIK, R. 1980. A general psychoevolutionary theory of emotion. *Emotion Theory, Research, And Experience*, Vol. 1.
- POPA, T., SOUTH-DICKINSON, I., BRADLEY, D., SHEFFER, A., AND HEIDRICH, W. 2010. Globally consistent space-time reconstruction. *Comp. Graphics Forum (Proc. SGP)*, 1633–1642.
- ROBERTSON, M. A., BORMAN, S., AND STEVENSON, R. L. 1999. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging* 12.
- ROSS, M. H., AND ROMRELL, L. J. 1989. Histology a text and atlas, *Williams and Wilkins*.

- RUMP, M., ZINKE, A. AND KLEIN, R. 2011. Practical spectral characterization of trichromatic cameras. In *Proceedings of the 2011 SIGGRAPH Asia Conference* 30, 6, 170.
- SHAFFER, S. A. 1985. Using color to separate reflection components. *Color Res. App.*, 10(4): 210-218.
- SHEWCHUK, J. 1996. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In *Applied Computational Geometry: Towards Geometric Engineering, Eds., vol. 1148 of Lecture Notes in Computer Science. Springer-Verlag*, 203–222.
- SHEFFER, A., LÉVY, B., MOGILNITSKY, M., BOGOMYAKOV, A. 2005. ABF++: Fast and Robust Angle Based Flattening, *ACM Transactions on Graphics*, 24(2), 311-330.
- SHEN, H. L., AND XIN, J. H. 2004. Colorimetric and spectral characterization of a color scanner using local statistics. *Journal of Imaging Science and Technology* 48, 4, 342–346.
- SHEN, H. L., AND XIN, J. H. 2004. Spectral characterization of a color scanner by adaptive estimation. *Journal of the Optical Society of America* 21, 7, 1125–1130.
- SHEN, H. L., AND XIN, J. H. 2006. Spectral characterization of a color scanner based on optimized adaptive estimation. *J. Opt. Soc. Am. A* 23, 7 (Jul), 1566–1569.
- SOLLI, M., ANDERSSON, M., LENZ, R., AND KRUSE, B. 2005. Color measurements with a consumer digital camera using spectral estimation techniques. In *Proceedings SCIA 2005*, 105–114.
- TSUMURA, N., HANEISHI, H., AND MIYAKE, Y. 1999. Independent component analysis of skin color image. *Journal of Optical Society of America A* 16, 9, 2169-2176.
- TSUMURA, N., OJIMA, N., SATO, K., SHIRAISHI, M., SHIMIZU, H., NABESHIMA, H., AKAZAKI, S., HORI, K., AND MIYAKE, Y. 2003. Image-based skin color and texture analysis synthesis by

extracting hemoglobin and melanin information in the skin. *Trans. on Graphics (Proc. SIGGRAPH)*, Vol. 22, No. 3, 770–779.

VORA, P. L., FARRELL, J. E., TIETZ, J. D., AND BRAINARD, D. H., 1997. Digital color cameras-2-spectral response.

VRHEL, M. J., AND TRUSSELL, H. J. 1999. Color device calibration: A mathematical formulation. *IEEE TRANS. IMAGE PROCESS* 8, 1796–1806.

WANG, Y., HUANG, X., LEE, C. S., ZHANG, S., LI, Z., SAMARAS, D., METAXAS, D., ELGAMMAL, A., AND HUANG, P. 2004. High resolution acquisition, learning and transfer of dynamic 3-D facial expressions. *Comp. Graphics Forum* 23, 3, 677–686.

WILLIAMS, L. 1990. Performance-driven facial animation. In *Computer Graphics (Proc. SIGGRAPH)*, vol. 24, 235–242.

WILSON, C. A., GHOSH, A., PEERS, P., CHIANG, J. Y., BUSCH, J., AND DEBEVEC, P. 2010. Temporal upsampling of performance geometry using photometric alignment. *ACM Trans. Graphics* 29, 2.

WEYRICH, T., MATUSIK, W., PFISTER, H., BICKEL, B., DONNER, C., TU, C., MCANDLESS, J., LEE, J., NGAN, A., JENSEN, H. W., AND GROSS, M. 2006. Analysis of Human Faces using a Measurement-Based Skin Reflectance Model. *Trans. on Graphics (Proc. SIGGRAPH)*, Vol. 25, 1013–1024.

YAMADA, T., AND WATANABE, T. 2007. Virtual Facial Image Synthesis with Facial Color Enhancement and Expression under Emotional Change of Anger. In *16th IEEE International Conference on Robot & Human Interactive Communication*, 49–54.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: High resolution capture for modeling and animation. *ACM Trans. Graphics* 23, 3, 548–558.

ZHANG, Z. 2000. A flexible new technique for camera calibration. *IEEE Trans. on PAMI*, Vol. 22, No. 11, pp. 1330–1334.