Radiometric Calibration Methods from Image Sequences

Seon Joo Kim

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2008

Approved by:

Marc Pollefeys, Advisor

Jan-Michael Frahm, Co-principal Reader

Leonard McMillan, Reader

Greg Welch, Reader

Svetlana Lazebnik, Reader

Abstract

Seon Joo Kim: Radiometric Calibration Methods from Image Sequences. (Under the direction of Marc Pollefeys.)

In many computer vision systems, an image of a scene is assumed to directly reflect the scene radiance. However, this is not the case for most cameras as the radiometric response function which is a mapping from the scene radiance to the image brightness is nonlinear. In addition, the exposure settings of the camera are adjusted (often in the auto-exposure mode) according to the dynamic range of the scene changing the appearance of the scene in the images. Vignetting effect which refers to the gradual fading-out of an image at points near its periphery also contributes in changing the scene appearance in images.

In this dissertation, I present several algorithms to compute the radiometric properties of a camera which enable us to find the relationship between the image brightness and the scene radiance. First, I introduce an algorithm to compute the vignetting function, the response function, and the exposure values that fully explain the radiometric image formation process from a set of images of a scene taken with different and unknown exposure values. One of the key features of the proposed method is that the movement of the camera is not limited when taking the pictures whereas most existing methods limit the motion of the camera. Then I present a joint feature tracking and radiometric calibration scheme which performs an integrated radiometric calibration in contrast to previous radiometric calibration techniques which require the correspondences as an input which leads to a chicken-and-egg problem as precise tracking requires accurate radiometric calibration. By combining both into an integrated approach we solve this chicken-and-egg problem. Finally, I propose a radiometric calibration method suited for a set of images of an outdoor scene taken at a regular interval over a period of time. This type of data is a challenging problem because the illumination for each image is changing causing the exposure of the camera to change and the conventional radiometric calibration framework cannot be used for this type of data. The proposed methods are applied

to radiometrically align images for seamless mosaics and 3D model textures, to create high dynamic range mosaics, and to build an adaptive stereo system.

To my wife, Jae Young.

Acknowledgments

I would like to acknowledge the enormous amount of help I have received from the teachers, colleagues, and family throughout the course of the Ph.D. program.

First, I would like to thank my advisor, Marc Pollefeys, for his tremendous support during my years in Chapel Hill. It has been a privilege to learn from him and this work would not have been possible without his guidance. I will carry with me his passion for research and the optimistic approach on solving problems.

I would also like to thank the rest of my committee : Jan-Michael Frahm, Leonard McMillan, Greg Welch, and Svetlana Lazebnik. I have been working closely with Jan for the last few years and his advice has been vital in finishing this work. The feedbacks and advice from Leonard, Greg, and Lana during classes and meetings were instrumental for my development as a researcher and for this work.

I am also grateful to the faculty and the staff in the Computer Science Department. I learned so much from them in and out of the classrooms and their support during the years have definitely helped me progress. In addition, I would like to thank my colleagues in the computer vision group which has grown significantly over the years. Very best wishes to Christopher Zach, Philippos Mordohai, Jean-Sebastien Franco, Enrique Dunn, Sudipta Sinha, David Gallup, Brian Clipp, Li Guan, Changchang Wu, Xiaowei Li, Rahul Raguram, Megha Pandey, Ram Krishan Kumar, Paul Merrell, Sashi-Kumar Penta, Jingyu Yan, Jason Repko, and Sriram-Thirthala Venkata. I also thank my students in COMP 110 who taught me the joy of teaching.

I would like to thank Jens Rittscher, Gianfranco Doretto, and Peter Tu at GE Research, with whom I had wonderful experience working as a summer intern. In

addition, I would like to thank Dan Goldman, Anatoly Litvinov, Yoav Schechner, and Nathan Jacobs for providing me with the data and the codes for the experiments.

I also thank my parents, sister, and parents-in-law for their support from back home. Finally, I would like to thank my wife Jae Young for always being there for me and believing in me. Thank you for being my inspiration and I wish you the very best in finishing your Ph.D. degree.

Table of Contents

List of Figures

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
B	Blue
BTF	Brightness Transfer Function
CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide-Semiconductor
DoRF	Database of Response Functions
EMoR	Empirical Model of Response
G	Green
GHz	Gigahertz
GPU	Graphics Processing Uint
HDR	High Dynamic Range
KLT	Kanade Lucas Tomasi
PCA	Principal Component Analysis
R	Red
RMS	Root-Mean-Square
SAD	Sum of Absolute Differences

Chapter 1

Introduction

In many computer vision systems, an image is assumed to represent a photometric measurement of a scene. However, this is not the case for most cameras as the radiometric response function which is a mapping from the scene radiance to the image brightness is nonlinear. In addition, the exposure settings of the camera are adjusted (often in the auto-exposure mode) according to the dynamic range of the scene, thus changing the appearance of the scene in the images. Vignetting effect, which refers to the gradual fading-out of an image at points near its periphery, also contributes in changing the scene's appearance in images. The effects of the exposure change and vignetting on images are shown in Figure 1.1. An image mosaic is created from multiple images, where each image is taken with a different exposure value to capture the high dynamic range of the scene, which is much greater than the camera's dynamic range. While the scene itself was reflecting light consistently during the image capture, the resulting mosaic exhibits significant brightness inconsistency due to the exposure changes and the vignetting effect. Another example of the effect of the exposure change is shown in Figure 1.2, where pixel values of a point over time recorded with auto-exposure are compared with those recorded with a fixed exposure value. In such outdoor scenes, the exposure is adjusted to accommodate the significant lighting variation over the course of a day. In this particular example, the sun is moving away from the surface. As time

Figure 1.1: Effect of exposure and vignetting on images. Due to vignetting and exposure changes between images, there are significant brightness inconsistency in the image mosaic.

goes on the radiance of the points in the scene decrease, as shown by the pixel values of the fixed exposure sequence. But the camera compensates for the decrease in the brightness of the scene by increasing its exposure value, resulting in almost constant pixel values over time.

While the exposure change (or auto-exposure) is desirable to make optimal use of the limited dynamic range of most cameras, it has an ill effect on many computer vision methods along with the nonlinearity of the camera response that rely on the scene radiance measurement such as photometric stereo, color constancy, and on the methods that use image sequences or time-lapse data of a long period of time such as in Jacobs et al. (2006, 2007) and Weiss (2001) since the pixel values do not reflect the actual scene radiance. The radiometric properties of the camera also affect the synthesis of image mosaics and texture-maps for 3-D models from multiple images as seen in Figure 1.1. The goal of this dissertation is to compute the radiometric properties of the cameras including the radiometric response function, the exposure values, and the vignetting function from multiple images which explain the relationship between the image brightness and the scene radiance as well as the radiometric relationship between multiple images.

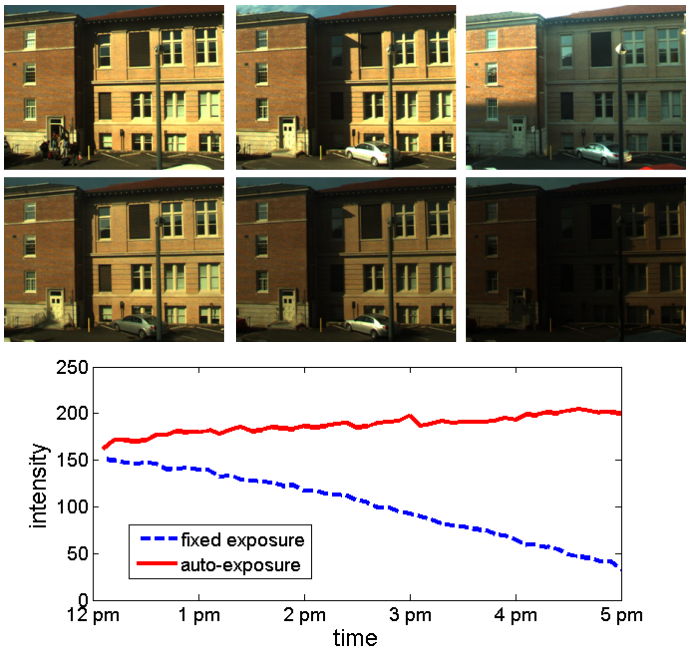


Figure 1.2: Effect of auto-exposure on images. Sample of images taken at different times with (Top) auto-exposure and (Middle) exposure fixed. (Bottom) Pixel values of a point over time.

1.1 Thesis Statement

Given a collection of images of a scene taken under varying conditions, one can compute radiometric properties of the camera (up to some ambiguities) that explain the relationship between the image brightness and the scene radiance as well as the radiometric relationship between multiple images. The radiometric properties include the radiometric response function, exposures, and vignetting.

1.2 Contribution

My research makes the following contributions:

1. **Robust radiometric calibration and vignetting correction from correspondence. (Chapter 3)**

   I introduce an algorithm that robustly estimates the radiometric response func-

tion, exposures, and the vignetting effect given multiple images taken with a freely moving camera. Specifically,

- I present a method to decouple the vignetting effect from the radiometric response function estimation.
- I introduce a novel method to estimate the radiometric response function from correspondences between images which is robust to noise and outliers enabling the use of images from a moving camera.
- I present a vignetting estimation method which is also robust to noise and outliers.
- I demonstrate methods to radiometrically align images and to create high dynamic range (HDR) mosaics using the estimated radiometric properties of the camera.

2. **Joint feature tracking and radiometric calibration. (Chapter 4)**

   I present an algorithm suited for video data taken with auto-exposure where the correspondence (feature tracks) and the radiometric response function along with the exposure values are computed simultaneously. In detail,

   - I present a novel method to simultaneously compute the feature tracks and the camera exposure values from a video taken with a camera with known response.
   - I present a method to simultaneously compute the feature tracks, the radiometric response function, and the exposures from a video taken with a camera with unknown response.
   - I apply the results of the algorithm to build an adaptive stereo system.

3. **Radiometric calibration with illumination change for outdoor scene analysis. (Chapter 5)**

I introduce a new algorithm to compute the radiometric response function and the exposure of images given a sequence of images of a static outdoor scene taken over time where the illumination is changing. In detail,

- I present a method to cluster pixels with same illumination conditions.
- I introduce a method to estimate the radiometric response function using the group of pixels with same illumination conditions.
- I present a novel method to compute the exposure values of images using the illumination model assuming the known motion of the sun.

1.3 Overview

The remainder of the dissertation is organized as follows.

Chapter 2 presents the overview of the image formation process and the related terminology. In addition, previous work on radiometric calibration and vignetting correction is surveyed.

Chapter 3 describes a novel method for robust radiometric calibration and vignetting correction that deals with images taken with a moving camera. Applications including radiometric alignment of images for texture-mapping 3-D models and image mosaics as well as high dynamic range (HDR) mosaic are shown.

Chapter 4 introduces a new framework for feature tracking where radiometric calibration process is combined with feature tracking. The presented method is applied to build an adaptive stereo system.

Chapter 5 presents a new algorithm to compute the radiometric response function and the exposure of images given a sequence of images of a static outdoor scene where the illumination is changing.

Chapter 6 discusses the contributions of this thesis and suggests directions for future work.

Chapter 2

Background

What determines the brightness at a certain point in an image? How is the image brightness related to the actual scene brightness? These are the key questions asked for this dissertation. Before presenting novel methods developed to answer those questions, I first review the related terminology and the image formation process in this chapter. In addition, a survey of previous work on the topic of radiometric calibration and vignetting correction is presented.

2.1 Image Formation

Figure 2.1 summarizes the image formation process. The scene brightness or the amount of light reflected from a surface point (x) to a direction can be defined by the term *radiance* which is the power per unit foreshortened area emitted into a unit solid angle by a surface (Figure 2.1, L) (Horn, 1986). The unit for radiance is watts per square meter per steradian[1] ($Wm^{-2}sr^{-1}$). For a Lambertian surface for which the radiance leaving the surface is independent of the angle, the radiance is proportional to the albedo of the surface point and the dot product between the illumination direction and the surface normal. Albedo is a reflectance term for Lambertian (diffuse) surface ranging from 0 to 1 characterizing the ratio of reflected light to incident light.

[1]Steradian is the unit of solid angular measure. See Horn (1986) for more details.

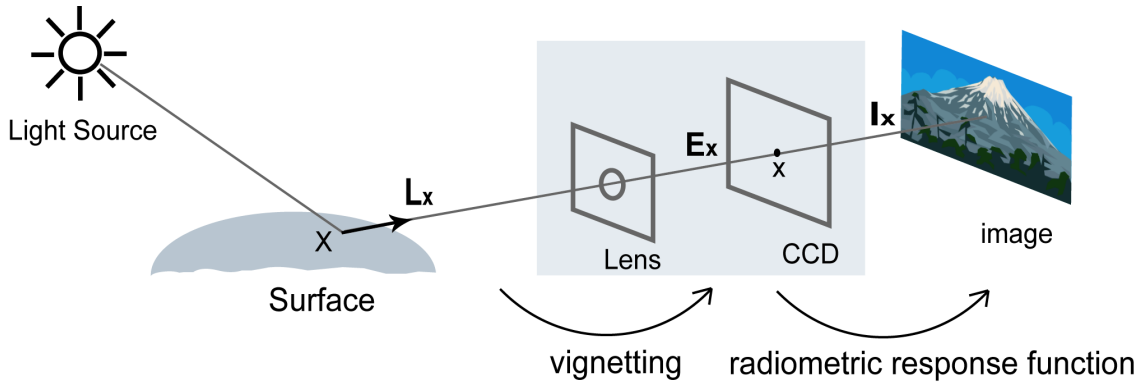


Figure 2.1: Radiometric image formation process. Vignetting affects the transformation from the scene radiance (L) to the image irradiance (E). Then the radiometric response function explains the nonlinear relationship between the image irradiance (E) and the image brightness (I).

After passing through the lens system, the power of radiant energy falling on the image plane is called the *image irradiance* (Figure 2.1, E). The unit for irradiance is watts per square meter (Wm^{-2}). Irradiance is then transformed to image brightness (I). These two steps, radiance to image irradiance and image irradiance to image brightness are explained in more details below.

2.1.1 Radiance to Image Irradiance

The amount of light hitting the image plane (image irradiance, E) is proportional to the scene radiance (L) but varies spatially causing the fade-out in the image periphery due to multiple factors (Figure 2.2). This irradiance fall-off effect often goes unnoticed unless the object in the image is of uniform color / brightness. However, this image distortion can be damaging to photometric methods such as shape from shading, appearance-based techniques such as object recognition, and image mosaicing (Zheng et al., 2006).

One of the factors for the irradiance fall-off in the periphery is the cosine-fourth law that defines the relationship between the radiance (L) and the irradiance (E) which is derived using a simple camera model consisting of a thin lens and an image plane (Horn,

1986). The following equation shows that the irradiance is proportional to the radiance but it decreases as cosine-fourth of the angle θ that a ray makes with the optical axis. In the equation, R is the radius of the lens and d denotes the distance between the lens and the image plane.

$$E = \frac{L\pi R^2 cos^4\theta}{4d^2} \tag{2.1}$$

A more dominant source for the irradiance fall-off is a phenomenon called vignetting. The vignetting effect refers to the gradual darkening of an image towards image corners due to the blocking of a part of the incident ray bundle by the effective aperture size (Yu, 2004). The effect of vignetting increases as the size of the aperture increases and vice versa (Figure 2.2). The white openings in Figure 2.2 indicate effective apertures. For a large aperture size (small F-number), the opening is smaller when viewed from an oblique angle because the view is blocked by the lens barrel. This implies that for large apertures, the lens will collect less light away from its optical axis making the image corners darker than its center. The vignetting effect decreases as the aperture size gets smaller since the opening (effective aperture) gets smaller and no longer blocked.

A phenomenon called the pupil aberration has been described as another cause for the fall in irradiance away from the image center (Aggarwal et al., 2001). The pupil aberration is caused by the nonlinear refraction of the rays which results in a significantly nonuniform light distribution across the aperture.

In this thesis, I view vignetting as the combination of all irradiance fall-off effects including the effect from the cosine-fourth law and the pupil aberration as it is the most dominant factor as well as to conform with the previous work and for generality. Rather than trying to model this radiometric distortion physically by combining the effects from different sources, we use a model that explains the overall irradiance fall-off behavior. The following equation shows the mapping from radiance ($L_{\mathbf{X}}$) to image irradiance ($E_{\mathbf{x}}$) through the vignetting function ($V(r_{\mathbf{x}})$) which is radially symmetric with r being the distance of the image point from the image center.

Figure 2.2: Vignetting Effect. (Top) Image taken with (left) a large aperture (f/1.4) and (right) a small aperture (f/5.6). (Bottom) Lens Images : (left) f/1.4 and (right) f/5.6 . ©Paul van Walree (www.vanwalree.com)

$$E_{\mathbf{x}} = V(r_{\mathbf{x}})L_{\mathbf{X}} \tag{2.2}$$

2.1.2 Image Irradiance to Image Brightness

The amount of light collected by the imaging sensor (irradiance E) is transformed to image brightness value (I) through a function called the radiometric response function or the camera response function. The relationship can be stated as follows :

$$I_{\mathbf{x}} = f(kE_{\mathbf{x}}) \tag{2.3}$$

where $E_{\mathbf{x}}$ is the image irradiance at a point $\mathbf{x}$, k is the exposure value with which the picture was taken, and $I_{\mathbf{x}}$ is the observed image intensity value at the pixel $\mathbf{x}$. Because increasing the irradiance will result in increasing (or keeping constant) the image intensity for cameras, the response function is (semi-) monotonic and can be inverted.

In general, the camera response is a nonlinear function providing means to compress the dynamic range of the scene that far exceeds the dynamic range of the camera. For digital cameras, even though the CCD and the CMOS respond linearly to the image irradiance, nonlinearities are purposely introduced in the cameras electronics to mimic the nonlinearities of film, to mimic the response of the human visual system, or to create a variety of aesthetic effects (Grossberg and Nayar, 2004).

As can be seen in Equation (2.3), the exposure value k plays a big role in deciding the final image intensity value by determining the amount of light exposed on the imaging sensor. Since the dynamic range of the scene usually exceeds that of a camera, the exposure value has to be adjusted to capture the dynamic range of interest by controlling the shutter speed and/or the aperture. The shutter speed controls how long the imaging sensor is exposed and is represented in a fraction of a second like 1/500, 1/250, 1/120,

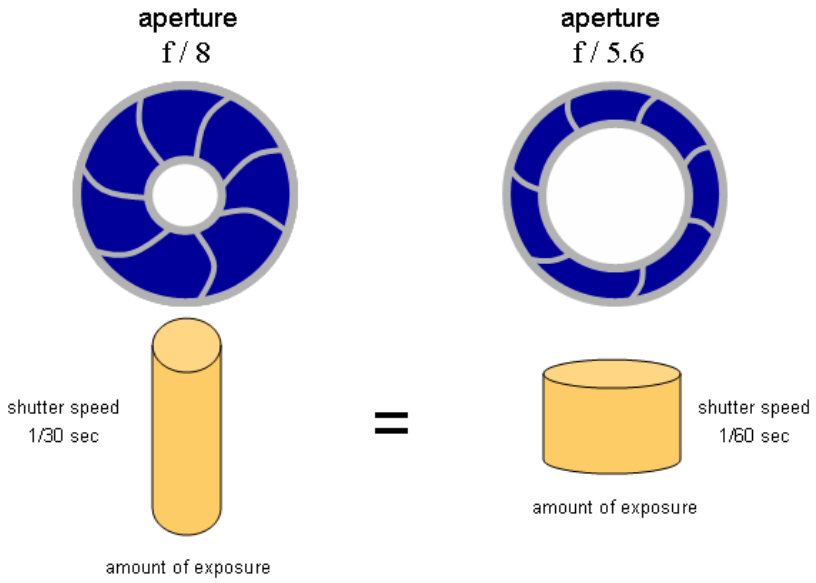


Figure 2.3: Reciprocity : the same amount of light is obtained with an exposure twice as long and an aperture area half as big

1/60, 1/30, 1/15, and 1/8. The shutter speed affects how motions in the scene are captured in the image: a fast shutter speed will freeze the movement and a slow shutter speed will blur the motion. The aperture is the diameter of the lens opening which is expressed as a fraction of focal length (f-number) such as f/2.0, f/2.8, f/4, f/5.6, etc. Smaller f-numbers represent bigger apertures. The aperture is related to the depth of field which means the amount of the picture, from foreground to background, that is in sharp focus. A smaller aperture will give you a greater depth of field and a larger aperture will give you a more restricted depth of field. Many different combinations of the shutter speed and the aperture result in identical exposure[2] and the choice depends on the motion and the depth of field.

Most of the digital cameras provide several ways to set exposure. In the auto-exposure mode, the camera automatically determines the appropriate aperture and shutter speed for the scene. In addition, there are two semi-automatic methods called the aperture priority mode and the shutter priority mode. In case of the aperture priority

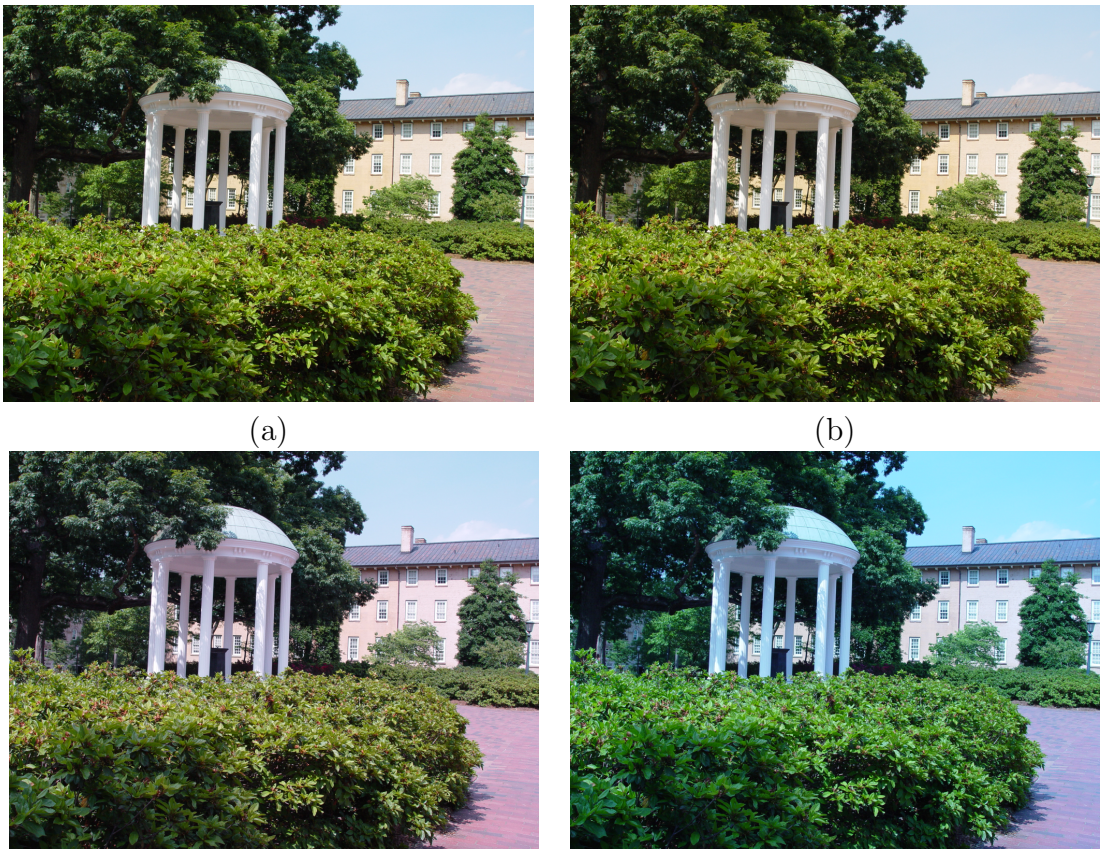
[2]This is called the reciprocity.

mode, the user manually chooses the size of the aperture while the camera automatically determines the shutter speed appropriate for the shooting condition. In the shutter priority mode, the decision on the shutter is made by the user and the aperture is determined by the camera. Finally, there is the manual mode in which both the aperture and the shutter speed is manually chosen by the user.

The effect of exposure change on images is illustrated in Figure 2.4. A set of images were taken around a tree with auto-exposure where some images were taken inside shadows and some in sunlight (images on top in Figure 2.4). The exposure value was adjusted to a high value in shadows to allow more light in the camera and to a low value in sunlight to allow less light. If the images were taken with a fixed low exposure, the images will look as the images in the second row of Figure 2.4. On the other hand, the images will look as the images on the bottom of Figure 2.4 if a fixed high exposure was used. As can be seen, the auto-exposure functionality provides the flexibility of not having to worry about finding the right 8-bit range to avoid over-exposed or under-exposed images. However, the exposure change causes the appearance of the object to change, which may be problematic for some computer vision methods that relate multiple images, such as image matching, feature tracking, and creating image mosaics and texture-maps.

There is another step in the imaging process called the white balance that influences the image brightness. The white balance corresponds to color constancy in human visual system, which is the ability to perceive color of an object independent of the illumination condition. The goal of white balance is usually to make sure that a white object appears white in the image no matter what the illumination condition is (Martinkauppi, 2002). For example, a photograph taken under incandescent illumination will appear unnaturally orange without proper white balance (Hsu et al., 2008). As with the exposure, digital cameras provide automatic white balancing. One of the basic algorithms for automatic white balancing is the gray world assumption which assumes that the

Figure 2.4: Effect of exposure change: (Top) Few sample images of the sequence taken with auto-exposure. Images radiometrically aligned with a fixed (Middle) high exposure value and (Bottom) low exposure value.



(a) (b)

(c) (d)

Figure 2.5: Effect of white balance. Photographs taken with different white balance settings (Sony F717) : (a) automatic white balance, (b) set to sunlight (c) set to fluorescent light (d) set to incandescent light.

average color in a scene is gray (Buchsbaum, 1980). Additionally, users can manually adjust the white balance by setting the illumination condition to one of camera-provided presets, such as daylight, cloudy, fluorescent, and incandescent. Figure 2.5 illustrates the effect of white balancing with images taken with different white balance settings. In this dissertation, the white balance is modeled with different exposure values for each of the color channels (red, green, and blue).

Figure 2.6: Macbeth color chart with 24 color patches with known reflectance.

2.2 Previous Work

While both the radiometric response function and the vignetting problem need to be addressed to fully explain the radiometric image formation process, works on these two problems have been developed separately in most cases. Hence, we can classify previous work on the subject into three categories: methods that deal with the camera response function only, methods that deal with vignetting only, and those that include both problems.

2.2.1 Radiometric Response Function Estimation

We first discuss the works that compute the radiometric response function without considering the vignetting effect. One way to compute the camera response function is to photograph a color chart with known reflectances, such as the Macbeth chart (Figure 2.6), in a uniform illumination condition. A mapping from the known reflectance to the image intensity provides a simple means to find the camera response function. However, the radiometric calibration using the color chart is not practical since the method can be used only when the image of the chart is available.

For chartless recovery of the camera response, the majority of existing radiometric

calibration methods uses multiple images taken with different exposure values to compute the camera response function. Assuming constant irradiance value, which implies constant illumination while the photographs are taken, the change in intensity is explained by the change in exposure. Taking the inverse of the response function on both sides in Equation (2.3), we get

$$f^{-1}(I_{\mathbf{x}}) = kE_{\mathbf{x}} \ . \tag{2.4}$$

With two images taken with different exposure values k_i and k_j, Equation (2.4) becomes

$$\frac{f^{-1}(I_{\mathbf{x_i}})}{f^{-1}(I_{\mathbf{x_j}})} = \frac{k_i}{k_j} \ , \tag{2.5}$$

assuming the irradiance of the point stays constant ($E_{\mathbf{x_i}} = E_{\mathbf{x_j}}$). Let $g = \log f^{-1}$ and $K = \log k$, then we get the following relationship in the log-domain :

$$g(I_{\mathbf{x_i}}) - g(I_{\mathbf{x_j}}) = K_i - K_j \ . \tag{2.6}$$

Equation (2.5) or (2.6) serves as the basis for computing the camera response function in most of the methods that use multiple images taken with different exposures and the early radiometric calibration methods concentrated on using different models of the response function. Mann and Picard (1995) estimated the response curve assuming that the response is a gamma curve and they know the exposure ratios between images. While the method is limited due to the model of the response function, the work by Mann and Picard has significance as the earliest work to introduce radiometric calibration from images and the concept of extending the dynamic range by combining differently exposed images. Debevec and Malik (1997) introduced a nonparametric method for response function recovery by imposing a smoothness constraint and assuming that the exposure values are known. With the computed response function, they created high dynamic

range (HDR) radiance maps from multiple images with different exposures, and showed applications of HDR maps such as synthesizing realistic motion blur and simulating the response of the human visual system. In the work by Mitsunaga and Nayar (1999), the response curve was assumed to be a low degree polynomial and was estimated iteratively with rough exposure ratio estimates. In their work, Mitsunaga and Nayar also introduced a method for automatic rejection of image areas with large vignetting effects and fused multiple images for HDR imaging. Tsin et al. (2001) also introduced an iterative method for computing the response function with the nonparametric response form using a statistical model of the measurement errors. Pal et al. (2004) propose the use of probability models for the imaging system and prior models for the response function to estimate the response function that is modeled differently for each image in the sequence. In Grossberg and Nayar (2004), the authors introduced a new model for the response function called the empirical model of response (EMoR) which is based on applying principal component analysis (PCA) to the database of response functions. This model will be discussed in details later in this chapter. A common limitation of all the mentioned methods above is that both the camera and the scene have to be fixed when multiple images are photographed for the calibration.

Several methods were introduced to loosen the scene and the camera movement restrictions. Mann and Mann (2001) proposed an iterative method with a non-parametric model that computes the response function and the exposures that allows camera rotation. Grossberg and Nayar (2003) explained ambiguities associated with the problem of finding the response function and introduced a response curve estimation method by recovering intensity mapping functions[3] between differently exposed images from histograms using histogram specification. The registration process is unnecessary in this method, allowing small movement of the scene and the camera. In Candocia and Man-

[3]In this dissertation, I use the term *brightness transfer function* instead of the intensity mapping function

darino (2005), the authors present an approach for response function computation by approximating the camera response function with a constrained piecewise linear model. They also incorporate the framework for spatial and tonal image registration (Candocia, 2003) to allow camera rotation.

Methods described so far use multiple images taken with different exposures and assume the irradiance for each image point stays constant, which implies that the illumination condition for all the images in the sequence is the same. A couple of methods were presented for computing the camera response when the illumination is changing. Manders et al. (2004) proposed a radiometric calibration method by using superposition constraints imposed by different combinations of two (or more) lights. It is difficult to apply this method in practice because it requires two or more images, each with different lighting direction, and an image with all the lights combined. Shafique and Shah (2004) also introduced a method that uses differently illuminated images. They estimate the response function by exploiting the fact that the material properties of the scene should remain constant and use cross-ratios of image values of different color channels to compute the response function. The response function is modeled as a gamma curve and a constrained non-linear minimization approach is used for the computation. This method is also limited in practice due to the restricted model for the response function and the algorithm is verified only by synthetic experiments by the authors.

Instead of using multiple images with different exposures or different lighting conditions, there are algorithms that compute the camera response function from a single image. Farid (2001) treats the radiometric nonlinearity as a gamma correction and presents a technique for computing the gamma correction from a single image without any information about the imaging device. His approach exploits the fact that gamma correction introduces specific higher-order correlations in the frequency domain which can be detected using tools from polyspectral analysis. The method by Farid is limited in practice because the response curves can be significantly different from a gamma curve.

Figure 2.7: An image of a flat and textureless Lambertian surface under constant illumination can be used as an reference image for vignetting estimation.

In the work by Lin et al. (2004), a single image was used for computing the response function by looking at the color distributions of local edge regions. Measured colors across edges should form linear distributions in color space due to blending of distinct region colors. However, they actually show nonlinear distributions because of the nonlinear camera response function. Using this idea, Lin et al. (2004) compute the response function which maps the nonlinear distributions of edge colors into linear distributions. Lin and Zhang further extended the method to deal with a single grayscale image by using the histograms of edge regions (Lin and Zhang, 2005). While these methods can be used for cases when multiple images with different exposures are not available, they are susceptible to high levels of image noise. Matsushita and Lin (2007) presented a method to complement the methods in Lin et al. (2004) and Lin and Zhang (2005) by using the asymmetric profiles of measured noise distributions to compute the camera response function which maps the asymmetric noise distribution to a symmetric distribution. This method requires the noise distributions for different image irradiances, which may not be simple, and the assumption on the symmetric noise distribution may not hold in low lighting conditions.

2.2.2 Vignetting Estimation

We now discuss previous work on vignetting. Conventional methods for correcting vignetting involve taking a reference image of a non-specular object such as a white paper with uniform color (Figure 2.7). This reference image is then used to build a correction lookup table or to approximate a parametric correction function. Asada et al. (2001) proposed a camera model using a variable cone that accounts for vignetting effects in a zoom lens system. Parameters of the variable cone model were estimated by taking images of a uniform radiance field. Yu et al. proposed using a hypercosine function to represent the pattern of the vignetting distortion for each scanline (Yu et al., 2004). They expanded their work to a 2D hypercosine model in Yu (2004) and also introduced an anti-vignetting method based on wavelet denoising and decimation. Other vignetting models include a simple form using radial distance and focal length (Uyttendaele et al., 2004), a third-order polynomial model (Bastuscheck, 1987), a first order Taylor expansion (Sawchuk, 1977), and an empirical exponential function (Chen and Mudunuri, 1986). While above methods rely on a reference image of an object of uniform color, Zheng et al. introduced a new method for determining the vignetting function given only a single image of a normal scene (Zheng et al., 2006). To extract vignetting information from an image, they presented adaptations of segmentation techniques that locate image regions for vignetting estimation. In all of the works mentioned above, the radiometric response function was ignored and vignetting was modeled in the image intensity domain rather than in irradiance domain.

In a related work, Schechner and Nayar (2003) exploited the vignetting effect to capture high dynamic range intensity values. In their work, they calibrate the "intended vignetting" using a linear least-squares fit on the image data itself rather than using a reference image. Their work assumes either a linear response function or a known response function. In Kang and Weiss (2000), vignetting effect was used for camera calibration. Their image formation model included the effect of the tilt of the camera in

addition to the vignetting effect. Using an image of a flat and textureless Lambertian surface under constant illumination, the camera intrinsics such as focal length, principal point, aspect ratio, and skew were computed. While the concept was novel, their method was impractical, since it did not yield accurate calibration results with real images.

2.2.3 Radiometric Response Function and Vignetting Estimation

Recently, works that include both the radiometric response function and the vignetting effect have been introduced. Litvinov and Schechner presented an unified framework for simultaneously estimating the unknown response function, exposures, and vignetting from a normal image sequence taken with camera motion (Litvinov and Schechner, 2005a,b). They achieve the goal by a nonparametric linear least squares method using common areas (correspondences) between images. Goldman and Chen (2005) also presented a solution for estimating the response function, the exposures, and vignetting. Using the empirical model of response (EMoR, Grossberg and Nayar (2004)) for the response function and a polynomial model for vignetting, they estimate the model parameters simultaneously by a nonlinear optimization method. In these papers, the recovered response function, exposure, and the vignetting factors were used to radiometrically align images for seamless mosaics. The method presented in Chapter 3 falls into this category and the results will be compared with results from Litvinov and Schechner (2005a) and Goldman and Chen (2005).

2.3 Radiometric Response Function Model

Before introducing different methods for radiometric calibration, I will first introduce the model of the radiometric response function used through out the dissertation. To model the camera response function, the empirical model of response (EMoR) introduced in

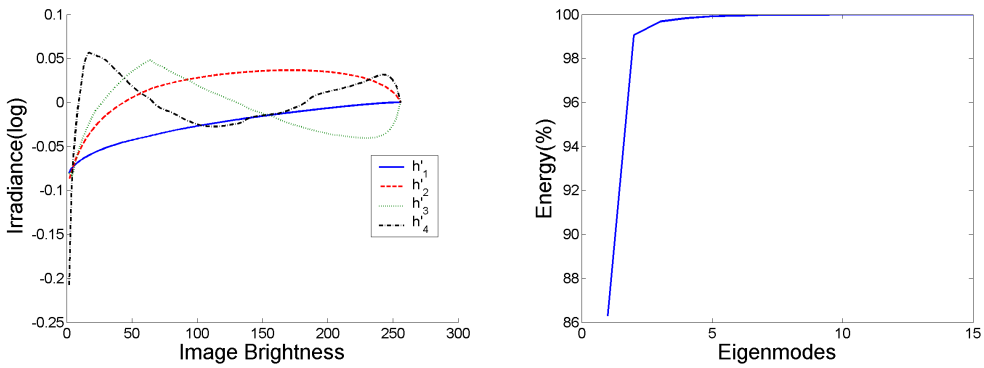


Figure 2.8: EMoR Basis. (Left) First four basis of the DoRF (log space), (Right) The cumulative energy occupied by the first 15 basis

Grossberg and Nayar (2004) will be used. In their work, Grossberg and Nayar first show that all response functions must lie within a convex set that results from the intersection of a hyperplane and a positive cone in function space. They also collected a Database of Response Functions (DoRF) of a variety of imaging systems including film, CCD, and solid-state camera components that are currently used. The database includes a total of 201 real-world response functions. Then they combine the constraints from the theoretical analysis and the data from DoRF to formulate a new model for the camera response function called the Empirical Model of Response (EMoR) which is a low (M^{th}) order approximation :

$$f(E) = f_0(E) + \sum_{n=1}^{M} c_n h_n(E), \tag{2.7}$$

where h_n's are basis functions found by applying PCA to the DoRF and f_0 is the mean function.

In log space, Equation (2.7) becomes :

$$g(I) = g_0(I) + \sum_{n=1}^{M} c_n h'_n(I), \tag{2.8}$$

where $g(I) = \ln f^{-1}(I)$ and h'_n's are basis functions for log inverse response function of

the database. The h'_n's are found by applying PCA to the log space of the DoRF. One thing to notice is that elements of the first column and the first row of the covariance matrix of DoRF in log space are -∞ since data are normalized from zero to one. So, we remove the first column and the first row from the matrix for the PCA. Figure 2.8 shows the first four basis functions of the log space of DoRF and the cumulative energy occupied by first 15 basis. The first three eigenvalues explain more than 99.6%, which suggest that the EMoR model represents the log space of response functions very well.

Chapter 3

Robust Radiometric Calibration and Vignetting Correction from Correspondence

3.1 Introduction

In this chapter, I introduce an algorithm to compute the vignetting function, the response function, and the exposure values that fully explain the radiometric image formation process from a set of images of a scene taken with different and unknown exposure values. One of the key features of the method is that the movement of the camera is not limited when taking the pictures whereas most existing methods limit the motion of the camera. The main application of interest is to radiometrically align images for image mosaics and for texture mapping 3D models where vignetting and exposure changes cause color inconsistency. The proposed approach is essentially different from image blending/feathering methods commonly used in image mosaicing (Brown and Lowe, 2003; Burt and Adelson, 1983; Levin et al., 2004) and other texture correction methods such as the method in Jia and Tang (2005) where the global and the local intensity variation were corrected using tensor voting, the method in Agathos and Fisher (2003) where a color transform was adapted for correcting the color discontinuity, and the method in

Beauchesne and Roy (2003) where a common lighting between textures was derived to relight textures. I also apply the method to create high dynamic range (HDR) mosaics that better represent radiometric measurement of the scene than normal mosaics.

The rest of the chapter is organized as follows. In the next section, a novel method for computing the radiometric response function is introduced. In Section 3.3, an algorithm for vignetting estimation is presented. Associated ambiguities are explained in Section 3.4 and methods for radiometrically aligning images and creating HDR mosaic are presented in Section 3.5. The proposed method is evaluated with various experiments in Section 3.6 and the chapter is concluded with discussions about the algorithm in Section 3.7.

Versions of this work were published in Kim and Pollefeys (2004) and Kim and Pollefeys (2008).

3.2 Radiometric Response Function Estimation

We begin by showing the equations for relating radiance (L) to image irradiance (E) and image irradiance (E) to image brightness (I) as introduced in Chapter 2.

$$E_{\mathbf{x}} = V(r_{\mathbf{x}})L_{\mathbf{X}} \tag{3.1}$$

$$I_{\mathbf{x}} = f(kE_{\mathbf{x}}) \tag{3.2}$$

Combining (3.1) and (3.2), the radiometric process of image formation can be mathematically stated as follows.

$$I_{\mathbf{x}} = f(kV(r_{\mathbf{x}})L_{\mathbf{X}}) \tag{3.3}$$

$L_{\mathbf{X}}$ is the radiance of a scene point $\mathbf{X}$ towards the camera, $I_{\mathbf{x}}$ is the image intensity value at the projected image point $\mathbf{x}$, k is the exposure, $f()$ is the radiometric response function, $V()$ is the vignetting function, and $r_{\mathbf{x}}$ is the normalized radius of the point $\mathbf{x}$

from the center of vignetting. We assume that vignetting is radially symmetric with the center of vignetting being the center of the image. We also assume that the vignetting function is the same for all images in the sequence. Equation (3.3) can be rewritten as follows.

$$\ln(f^{-1}(I_x)) = \ln k + \ln V(r_x) + \ln L_x \tag{3.4}$$

$$g(I_x) = K + \ln V(r_x) + \ln L_x \tag{3.5}$$

The goal of our algorithm is to estimate $f()$ (or $g()$), $V()$, and k (or K) given a set of differently exposed images taken with a non-stationary camera. Our work falls under the last group of existing work (Goldman and Chen (2005); Litvinov and Schechner (2005a)) explained in the previous chapter where both the response function and the vignetting function are recovered. The difference between those methods and our method is that while the camera response function and the vignetting function were estimated simultaneously in Goldman and Chen (2005) and Litvinov and Schechner (2005a,b), we approach the problem differently by robustly computing the response function and the vignetting function independently. Separating the two processes is possible by decoupling the vignetting process from the radiometric response function estimation. By separating the two processes, we derive a solution for each process that is robust against noise and outliers. Thus we are able to get robust estimation even when there is a vast number of outliers due to inaccurate stereo correspondences for the overlap region on the 3D models as well as non-Lambertian reflection. Previous least-squares based approaches are not able to deal with this.

3.2.1 Correspondence

Since we are dealing with images taken with a moving camera, the first thing that we consider is the computation of correspondences. Ideally, only a limited number of points are required to estimate the radiometric response curve, the exposures, and the

vignetting parameters. However, because of a certain number of limitations in finding accurate correspondences, it is best to estimate correspondences for a larger number of points. First, we want corresponding points to cover as many intensity values as possible (and this for each R, G and B channel separately). In addition, matching between images recorded with different exposure settings is in itself hard, thus, we expect a significant number of wrong matches. Finally, because we deal with a moving camera and not all pixels correspond to Lambertian surfaces, we can not always expect the radiance to be constant over varying viewing directions (this would not be a problem for static or purely rotating cameras). Therefore, it is important to obtain as much redundancy as possible so that a robust approach can later be used to estimate the desired camera properties.

If the set of images are captured with a purely rotating camera, we compute the homographies between images to compute the correspondences. We used the software "Autostitch" (Brown and Lowe, 2003)[1] for computing the homographies.

For images taken with a moving camera, the correspondences are computed by estimating the epipolar geometry for each pair of consecutive images (for video, keyframes would be selected so that the estimation of the epipolar geometry would be stable) using tracked or matched features, followed by stereo matching (Pollefeys et al., 2004). To avoid problems with intensity changes it is important to use zero-mean normalized cross-correlation. While we do not explicitly deal with independent motions in the scene, our stereo algorithm combined with our robust joint histogram approach explained in the next subsection will handle those as outliers.

3.2.2 Estimating the radiometric response function

Equation (3.3) shows that the response function $f()$ cannot be recovered without the knowledge about the vignetting function $V()$ and vice versa. Hence, one way to solve

[1] http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html

the problem is to estimate both functions simultaneously either in a linear (Litvinov and Schechner, 2005a,b) or in a nonlinear way (Goldman and Chen, 2005) . But if we use corresponding points affected with the same amount of vignetting, we can decouple the vignetting effect from the process and estimate the response function without worrying about the vignetting effect using Equation (3.3). Let $\mathbf{x}_i$ and $\mathbf{x}_j$ be image points of a scene point $\mathbf{X}$ in image i and image j respectively. If $r_{\mathbf{x}_i} = r_{\mathbf{x}_j}$ then $V(r_{\mathbf{x}_i}) = V(r_{\mathbf{x}_j})$ since we have already made the assumption that the vignetting model is the same for all images in the sequence. Hence by using corresponding points that are of equal distance from the center of each image, we can decouple vignetting from the response function. So after finding all possible correspondences first using the methods described in the previous subsection (homography for rotating camera and stereo matching for moving camera), we then compare the distance of the points in each matching pair from the center of its image in order to select only correspondences with equal distance. In practice, we allowed some tolerance to the constraint by allowing correspondences that are close to equal distance from the center rather than strictly enforcing correspondences to be of exact equal distance. In the case of panoramic images, these correspondences will form a band between images (Figure 3.1). In the case of stereo images, these correspondences will form an arbitrarily deformed shape depending on the geometry of the scene and the motion of the camera (Figure 3.2). Note that while in general there are no problems finding a sufficient amount of such correspondences in stereo cases, there are some cases that may not yield enough correspondences particulary in the case of forward (backward) motion where the radius for all pixels would increase (decrease) (except that even in that case we might still have far away points that stay approximately fixed and allow for the exposure changes to be computed while the response function mostly gets constrained by other images).

By using only those correspondences mentioned above, we obtain the following equation from Equation (3.5) where the vignetting function is now removed from the process.

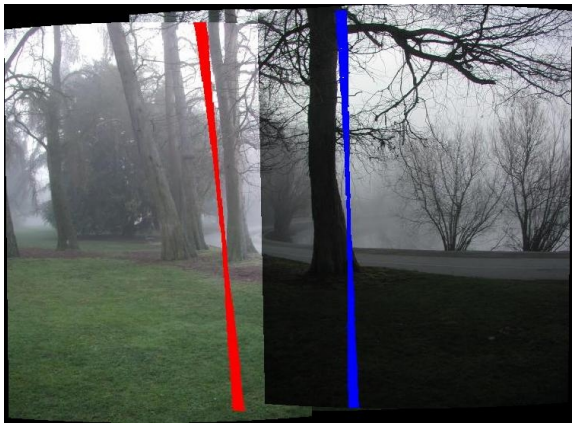


Figure 3.1: Decoupling the vignetting effect (mosaic image) : The figure shows three images stitched to a mosaic. Only corresponding points in the colored band (red for the first pair and blue for the second) are used to decouple the vignetting effect from estimating the radiometric response function .

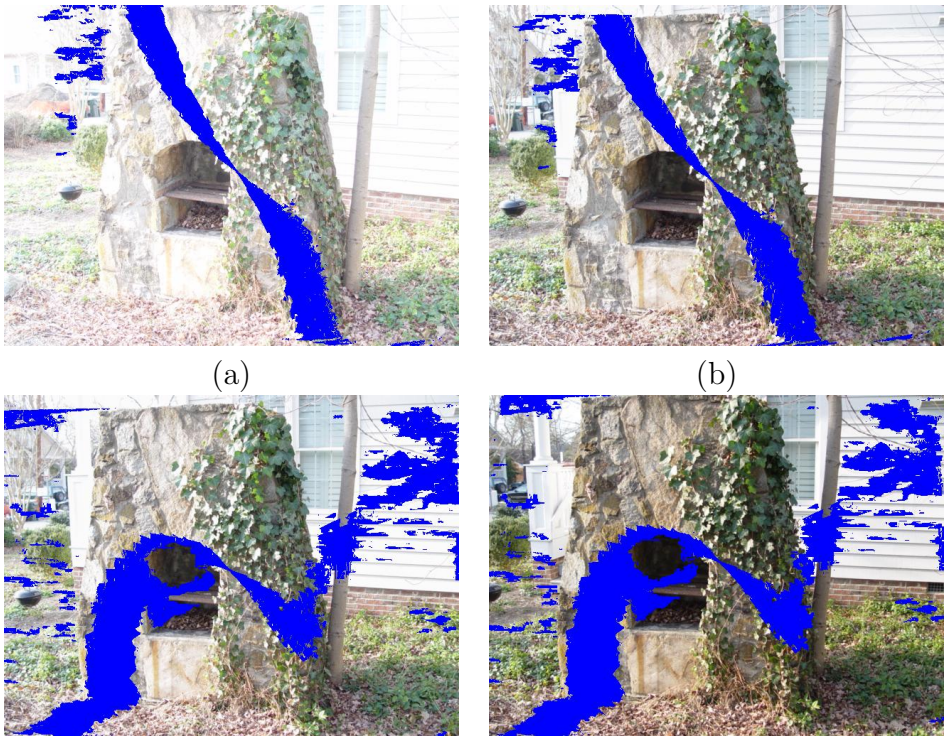


Figure 3.2: Decoupling the vignetting effect (stereo images) : Stereo image pairs (a)-(b), (c)-(d). The colored pixels are the corresponding points between the images that satisfy the equal radius condition (with the tolerance of ± 3 pixels in these examples). By using only these points, we can decouple the vignetting effect from estimating the radiometric response function.

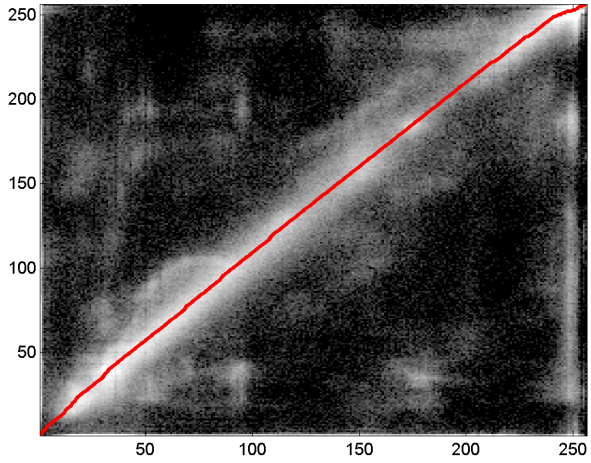


Figure 3.3: An example of a joint histogram with the estimated brightness transfer function (BTF) overlaid on it.

$$g(I_{\mathbf{x}_i}) - g(I_{\mathbf{x}_j}) = K_i - K_j \tag{3.6}$$

While Equation (3.6) is solved for the response function $g()$ in a least squares sense in most previous works, we approach the problem in a robust way to achieve robustness against noise and mismatches. This is very critical since we are dealing with images taken with a moving camera where using least squares would not yield accurate results due to noise and a vast number of outliers. The robust estimation process is explained in details in the following subsections.

Joint Histogram and Brightness Transfer Function

For a pair of images, all the information relevant to our problem is contained in the pair of intensity values of corresponding points. As suggested in Mann (2000), these can all be collected in a two-variable joint histogram which he calls the *comparagram*. For a pair of corresponding intensity values $(I_{\mathbf{x}_i}, I_{\mathbf{x}_j})$, the corresponding value in the joint histogram $J(I_{\mathbf{x}_i}, I_{\mathbf{x}_j})$ indicates for how many pixels the intensity value changes from $I_{\mathbf{x}_i}$

to $I_{\mathbf{x}_j}$.

As noted in Grossberg and Nayar (2003), ideally a function should relate the intensity values between the two images. From Equation (3.6), one immediately obtains

$$I_{\mathbf{x}_j} = \tau(I_{\mathbf{x}_i}) := g^{-1}(g(I_{\mathbf{x}_i}) + \Delta K) \ . \tag{3.7}$$

with $\Delta K = K_j - K_i$. We will call the function τ as the *brightness transfer function* (BTF). It was shown in Grossberg and Nayar (2003) that under reasonable assumptions for g, τ is monotonically increasing, $\tau(0) = 0$ and if $\Delta K > 0$, then $I \leq \tau(I)$. Inversely, if $\Delta K < 0$ then $I \geq \tau(I)$. Ideally, making abstraction of noise and discretisation, if $I_{x_j} \neq \tau(I_{x_i})$, then we should have $J(I_{\mathbf{x}_i}, I_{\mathbf{x}_j}) = 0$. However, real joint histograms are quite different due to image noise, mismatches, view dependent effects and a non-uniform histogram as shown in Figure 3.3. In the presence of large numbers of outliers, least squares solutions for response functions as have been used by others are not viable. We propose to use the following function as an approximation for the likelihood of the BTF passing through a pixel of the joint histogram.

$$P(\tau(I_1) = I_2|\bar{J}) = (G(0, \sigma_1) * \bar{J})(I_1, I_2) + P_o \tag{3.8}$$

where $G(0, \sigma_1)*$ represent the convolution with a zero-mean Gaussian with standard deviation σ_1 to take image noise into account, $\bar{J}$ is the normalized joint histogram, and P_0 is a term that represents the probability for $\tau(I_1) = I_2$ independent of the joint histogram. This term is necessary to be able to deal with the possibility of having the BTF pass through zeros in the joint histogram which could be necessary if for some intensity values no correct correspondence was obtained. Based on these assumptions the most probable solution is the BTF that maximizes

$$\ln P(\tau|\bar{J}) = \iint J_\tau(I_1, I_2) ln P(\tau(I_1) = I_2|\bar{J}) dI_1 dI_2 \tag{3.9}$$

with $J_\tau(I_1, I_2)$ being a function that is one where $I_2 = \tau(I_1)$ and zero otherwise. Using dynamic programming it is possible to compute the BTF that maximizes Equation (3.9) under the constraints discussed above, i.e. semi-monotonicity, $\tau(0) = 0$, $\tau(255) = 255$ and $\tau(I) \geq I$ or $\tau(I) \leq I$ for all I (Figure 3.3).

3.2.3 Radiometric Response Function Estimation

With the computed BTFs, we now estimate the radiometric response function by using the low parameter Empirical Model of Response (EMoR) by Grossberg and Nayar (2004) as the model for the response function which was explained in Section 2.3. The equation for the model in log space is as follows.

$$g(I) = g_0(I) + \sum_{n=1}^{M} c_n h'_n(I), \tag{3.10}$$

where $g(I) = \ln f^{-1}(I)$ and h'_n's are basis functions for log inverse response function of the database.

We estimate the response function and exposure differences between images by using the computed BTFs and combining Equation (3.6) and Equation (3.10).

$$g_0(\tau_{ij}(I)) - g_0(I) + \sum_{n=1}^{M} c_n(h'_n(\tau_{i,j}(I)) - h'_n(I)) - K_{ji} = 0 \tag{3.11}$$

where $K_{ji} = K_j - K_i$ and $\tau_{i,j}()$ is the brightness transfer function from the image i with exposure K_i to the image j with exposure K_j.

To deal with the white balance, we adopt the simplifying assumption that the effect of white-balance corresponds to changing the exposure independently for each color band (Finlayson et al., 1994). Then the unknowns of Equation (3.11) are the coefficients c_n's and the exposure differences K_{ji}'s for each different color channel. The solution for these unknowns at this point will suffer from the exponential ambiguity. The exponential ambiguity means that if a response function g and some exposures K's are solution

to Equation (3.6) then so are γg and γK's. Simply put, there are many response functions and exposures that satisfy the equation as long as they have the same scale factor. As stated in Grossberg and Nayar (2003), it is impossible to recover both the response function (g) and the exposures (K's) simultaneously from BTF alone, without making assumptions on either the response function or the exposures. To resolve this ambiguity problem, we chose to make an assumption on an exposure value by setting the exposure difference of a pair to a constant value (α). For simplicity, we will set the exposure difference of the first image pair K_{12} to a constant value (α) in this work. This serves as fixing the scale of the response function. For many applications including the tonemapping for high dynamic range imaging and the texture alignment application, the choice of the constant is not critical which is an advantage over many other methods which require exact or rough estimate of exposure values. The sign of α should be positive when the exposure increases while it should be negative when the exposure decreases for the chosen pair. Alternatively, the scale can be fixed by setting the value of the response curve at an intensity to an arbitrary value. This alternative method is used in Chapter 4 and Chapter 5 to deal with the ambiguity.

After fixing the value of K_{12}, we now have to solve for the unknown model parameters (c_n) and exposure differences of each image pair for each color channel except the first pair which amounts to $M + 3(N - 2)$ unknowns. The computed BTF (τ) for each color channel of an image pair yields 254 equations (Equation 3.11) for image values (I) 1 to 254. We do not include the value 0 and 255 to avoid under-exposed or saturated data.

To solve the problem in a linear least squares sense, we first build matrices $\mathbf{A}_i^l$ $(254 \times (M + 3(N - 2)))$ and $\mathbf{b}_i^l$ (254×1) for each image pair other than the first pair

$(2 \leq i \leq N-1)$ and each color channel ($l \in \{R, G, B\}$ or $\{0, 1, 2\}$).

$$\mathbf{A}_i^l(m,n) = \begin{cases} w(m)(h'_n(\tau_{i,i+1}^l(m)) - h'_n(m)); & 1 \leq m \leq 254, 1 \leq n \leq M \\ -w(m); & 1 \leq m \leq 254,\ n = M + (N-2) \times l + i - 1 \\ 0; \quad \text{elsewhere} \end{cases} \tag{3.12}$$

$$\mathbf{b}_i^l(m) = w(m)(g_0(m) - g_0(\tau_{i,i+1}^l(m))); \quad 1 \leq m \leq 254 \tag{3.13}$$

For the first pair ($i = 1$),

$$\mathbf{A}_1^l(m,n) = \begin{cases} w(m)(h'_n(\tau_{1,2}^l(m)) - h'_n(m)); & 1 \leq m \leq 254, 1 \leq n \leq M \\ 0; \quad \text{elsewhere} \end{cases} \tag{3.14}$$

$$\mathbf{b}_1^l(m) = w(m)(g_0(m) - g_0(\tau_{1,2}^l(m)) + \alpha); \quad 1 \leq m \leq 254 \tag{3.15}$$

The weight w is as follows.

$$w(m) = w_1(m) w_2(m) \tag{3.16}$$

$$w_1(m) = \begin{cases} 0; & \text{if } J(m, \tau(m)) < \epsilon \text{ or } \tau(m) = 0 \text{ or } 255 \\ 1; & \text{else} \end{cases} \tag{3.17}$$

$$w_2(m) = exp(-\frac{((m-127)/127)^2}{2\sigma_2^2}) \tag{3.18}$$

The weights are included for two reasons. First, joint histograms may not contain data on all the intensity range. So the brightness transfer function (τ) values at the intensity range where there are no data (or very few) may not be accurate. Also, data that are either saturated or under-exposed should be eliminated from the equation. All these factors are reflected in the first weight w_1.

In addition, the response function will typically have a steep slope near I_{max} and I_{min}, so we expect the response function to be less smooth and fit the data more poorly near these extremes (Debevec and Malik, 1997). This is reflected to our algorithm by

the second weight w_2. We used values from 1.0 to 10.0 for σ_2 for the examples in this chapter.

To deal with the discretization problem, we also compute BTFs in the opposite direction (i+1 to i) and build matrices $\mathbf{A}_i^{l'}$ and $\mathbf{b}_i^{l'}$ which is similar to $\mathbf{A}_i^l$ and $\mathbf{b}_i^l$ except that $\tau_{i,i+1}$ is now changed to $\tau_{i+1,i}$ along with following few changes.

$$\mathbf{A}_i^{l'}(m,n) = \begin{cases} w(m)(h'_n(\tau^l_{i+1,i}(m)) - h'_n(m)); & 1 \le m \le 254, 1 \le n \le M \\ w(m); \quad 1 \le m \le 254, \ n{=}M{+}(N-2){\times}l{+}i{-}1 \\ 0; \quad \text{elsewhere} \end{cases} \tag{3.19}$$

$$\mathbf{b}_i^{l'}(m) = w(m)(g_0(m) - g_0(\tau^l_{i+1,i}(m))); \quad 1 \le m \le 254 \tag{3.20}$$

$$\mathbf{A}_1^{l'}(m,n) = \begin{cases} w(m)(h'_n(\tau^l_{2,1}(m)) - h'_n(m)); & 1 \le m \le 254, 1 \le n \le M \\ 0; \quad \text{elsewhere} \end{cases} \tag{3.21}$$

$$\mathbf{b}_1^{l'}(m) = w(m)(g_0(m) - g_0(\tau^l_{2,1}(m)) - \alpha); \quad 1 \le m \le 254 \tag{3.22}$$

After all the matrices above are built, we can solve for the coefficients of the model and the exposure differences linearly ($\mathbf{Au} = \mathbf{b}$) using the singular value decomposition (Equation (3.25)) by combining all the computed matrices to form $\mathbf{A}$ and $\mathbf{b}$ as in Equation (3.23) where each $\mathbf{A}^l$ and $\mathbf{b}^l$ are formed by combining $\mathbf{A}_i^l$ and $\mathbf{b}_i^l$ for all image pairs.

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^R \\ \mathbf{A}^{R'} \\ \mathbf{A}^G \\ \mathbf{A}^{G'} \\ \mathbf{A}^B \\ \mathbf{A}^{B'} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}^R \\ \mathbf{b}^{R'} \\ \mathbf{b}^G \\ \mathbf{b}^{G'} \\ \mathbf{b}^B \\ \mathbf{b}^{B'} \end{bmatrix} \tag{3.23}$$

$$\mathbf{u} = \left[c_1, \ldots, c_M, K_{23}^R, \ldots, K_{23}^G, \ldots, K_{23}^B, \ldots, K_{N-1N}^B\right]^T \tag{3.24}$$

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \| \mathbf{A}\mathbf{u} - \mathbf{b} \|_2. \tag{3.25}$$

3.3 Vignetting Estimation

After estimating the response function and the exposure values, each image intensity value is transformed to an irradiance value E to compute vignetting function V.

$$E_x = \frac{f^{-1}(I_x)}{k} = V(r_{\mathbf{x}})L_{\mathbf{X}} \tag{3.26}$$

Since the scene radiance $L_{\mathbf{X}}$ is the same for the corresponding points $\mathbf{x_i}$ and $\mathbf{x_j}$, we get

$$\frac{E_{x_i}}{V(r_{\mathbf{x_i}})} = \frac{E_{x_j}}{V(r_{\mathbf{x_j}})} \tag{3.27}$$

As presented in Section 2.2.2, many models for vignetting exist. In this dissertation, we chose to use the polynomial model used in Goldman and Chen (2005). In Goldman and Chen (2005), a third order polynomial was used for the vignetting model and it was estimated together with the response function simultaneously by a nonlinear optimization method. By computing the response function independent of vignetting in our first step, we can now compute the polynomial vignetting model linearly. This saves

a great deal of computational time compared to the nonlinear optimization scheme used in Goldman and Chen (2005), avoids issues with potential local minima, and it also enables us to easily use much higher order polynomial function for more accuracy. The vignetting model is given by

$$V(r) = 1 + \sum_{n=1}^{D} \beta_n r^{2n}. \tag{3.28}$$

Let $a = \frac{\mathbf{E_{x_i}}}{\mathbf{E_{x_j}}}$, then combining the model with Equation (3.27) yields the following equation.

$$\sum_{n=1}^{D} \beta_n (a r_{\mathbf{x_j}}^{2n} - r_{\mathbf{x_i}}^{2n}) = 1 - a \tag{3.29}$$

One obvious choice for solving for the D unknown β_n's is to use least squares since each corresponding pair of points in given image pairs provides additional equation in the form of (3.29). But in the presence of many outliers, the least squares solution will not give us a robust solution to the problem.

We propose to approach the problem similar to the way we computed the response function in the first stage. Rather than solving the problem in a least squares sense, we once again solve the problem in a robust fashion. For a pair of $r_{\mathbf{x_i}}$ and $r_{\mathbf{x_j}}$ (discretized), we estimate $\hat{a}(r_{\mathbf{x_i}}, r_{\mathbf{x_j}})$ which is the robust estimate of the ratio a for the given $r_{\mathbf{x_i}}$ and $r_{\mathbf{x_j}}$. For each matching pair of points in the image sequence with radius $r_{\mathbf{x_i}}$ and $r_{\mathbf{x_j}}$ respectively, the irradiance ratio a is computed and stacked at $s(r_{\mathbf{x_i}}, r_{\mathbf{x_j}})$ (Figure 3.4). We only use correspondences where the image intensity of each pixel is within certain range, from 10 to 245 for example. The purpose of this is to exclude saturated or under-exposed pixels as well as pixels in the intensity range where estimate of the response function tends to be less accurate. Also, only pairs with similar ratio for each color channel are added to the stack since the vignetting effect is the same for all color channels. In the end, $\hat{a}(r_{\mathbf{x_i}}, r_{\mathbf{x_j}})$ is computed as the median of stacked values $s(r_{\mathbf{x_i}}, r_{\mathbf{x_j}})$. Notice that we only have to keep track of cases where $r_{\mathbf{x_i}} > r_{\mathbf{x_j}}$ due to symmetry. With

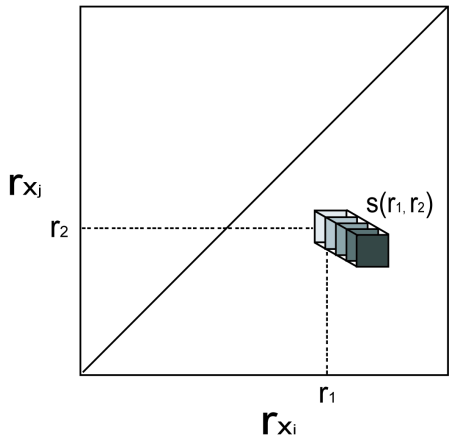


Figure 3.4: Estimating the irradiance ratio $\hat{a}$. For every matching pair with $r_{x_i} = r_1$, $r_{x_j} = r_2$, the value a is stacked in $s(r_1, r_2)$. $\hat{a}(r_1, r_2)$ is computed as the median of stacked values.

the discretisation of 100×100, we have less then 5000 equations in the form of Equation (3.29) instead of having one equations per matching pair of points.

The model coefficients ($\mathbf{v} = [\beta_1, \beta_2, ..., \beta_D]^T$) are estimated by solving the linear equation of the form $\mathbf{Yv} = \mathbf{z}$. The m^{th} r_{x_i} and r_{x_j} pair adds one equation (Equation (3.29)) to the linear equation as follows.

$$\mathbf{Y}(m, n) = w_v(m)(\hat{a}(r_{\mathbf{x_i}}, r_{\mathbf{x_j}}) r_{\mathbf{x_j}}^{2n} - r_{\mathbf{x_i}}^{2n}) \tag{3.30}$$

$$\mathbf{z}(m) = w_v(m)(1 - \hat{a}(r_{\mathbf{x_i}}, r_{\mathbf{x_j}})), \quad 1 \leq n \leq D$$

Note that we weight (w_v) each row of the matrix $\mathbf{Y}$ and $\mathbf{z}$ by the number of elements in the stack $s(r_{\mathbf{x_i}}, r_{\mathbf{x_j}})$. Finally, the model parameter vector $\mathbf{v}$ is the solution to the following least squares problem Equation (3.31) which can be solved using the singular value decomposition (SVD).

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} \| \mathbf{Yv} - \mathbf{z} \|_2. \tag{3.31}$$

Once we have the estimates of the response, the exposures, and the vignetting function from the given set of images, we can add other images that may have different exposure value and vignetting such as zoomed-in images to capture more dynamic range and details. Assuming that the center of vignetting stays at the center of the image, we can first compute the exposure of the added image using the pixels close to the center that are not affected by vignetting (we used pixels within 10% from the center) by Equation (3.5). Then the vignetting function is computed robustly in a similar way as the vignetting estimation described in the previous section. From Equation (3.27), the vignetting function for the zoomed-in image (M_z) is computed with the known response function (f), the vignetting function of the original image (M_i), the exposure of the original image (k_i), and the exposure of the zoomed-in image (k_z) as follows.

$$V_z(r_{x_z}) = \frac{f^{-1}(I_{x_z})k_i V_i(r_{x_i})}{f^{-1}(I_{x_i})k_z} \tag{3.32}$$

Again for the robustness against outliers and noise, we use the median of the right-hand side value of Equation (3.32) for each radius to fit the vignetting model instead of fitting the model to all possible data. An example of adding a zoomed-in image to the sequence is shown in the experiments section (Figure 3.15).

3.4 Ambiguities

As mentioned earlier, the process of radiometric calibration explained thus far is subject to the exponential ambiguity, sometimes called the γ ambiguity as in Grossberg and Nayar (2003) and Litvinov and Schechner (2005a). This ambiguity basically means that if $\hat{f}$, $\hat{k}$, and $\hat{V}$ are the solutions for Equation (3.3), then the whole family of $\hat{f}^\gamma$, $\hat{k}^\gamma$, and $\hat{V}^\gamma$ are also solutions to the problem. In this work, this ambiguity is dealt by setting an exposure ratio of an image pair to a constant value.

There is another ambiguity called the scale ambiguity (Litvinov and Schechner,

2005a) corresponding to arbitrary offsets in Equation (3.5) : $g + S_g$, $K + S_K$, and $\ln V(r_x) + S_V$ would all satisfy the equation with the radiance value being offset accordingly. This ambiguity is dealt with in this dissertation by normalizing the response function and the vignetting function.

Due to these ambiguities, the radiance value $\hat{L}_x$ that we recover using Equation (3.33) with the estimates $f()$, $V()$, and k would not be the true radiance value (L_x). It would be related to the true radiance by an exponential function. However, this is not a problem for many applications including the radiometric alignment and the high dynamic range display explained in the next subsection unless absolute or linearly proportional quantitative measurements are required such as in the simulation of motion blur or lens glare effects (Debevec and Malik, 1997).

$$\hat{L}_x = \frac{f^{-1}(I_x)}{kV(r_x)} \tag{3.33}$$

3.5 Radiometric Alignment and High Dynamic Range Mosaic Imaging

After computing the response function ($f()$), the vignetting function ($V()$), and exposure values (k) for each image, we can radiometrically align images in the sequence so that the vignetting is corrected and all images have a common exposure setting as follows.

$$I_x^{new} = f(k^{new}\hat{L}_x) \tag{3.34}$$

The ambiguities mentioned above will not have any effect on the alignment since the solutions with different γ values will still generate the same intensity values. By radiometrically aligning images, we can make mosaics and textures of 3D models look seamless. Note that even after the alignment, pixels that were either saturated or under-

exposed may still look inconsistent in the resulting mosaic or the 3D model depending on the new exposure value. We cannot find the radiance value if a pixel is either saturated or under-exposed.

Another application of our method is the high dynamic range (HDR) imaging, specifically creation of the high dynamic range mosaic. Radiometrically aligning images has the effect of fixing the exposure which limits the showing of the full dynamic range of the scene. By displaying the estimated scene radiance values ($\hat{L}_x$), we can represent the high dynamic range scene more realistically. While we are not displaying the actual radiance value due to ambiguities, we can alleviate the problem by tuning the value of γ for visual plausibility (Litvinov and Schechner, 2005b). Since most of the displaying systems cannot show high dynamic range images, we have to compress the estimated radiance values ($\hat{L}_x$) appropriately using a method called tonemapping. In this dissertation, we used a software called Photomatix[2] for tonemapping.

For high dynamic range mosaics, we scan the scene changing the exposure accordingly. We have to make sure that every point in the mosaic is at least once correctly exposed meaning it is neither underexposed or saturated. The response function, exposures, and vignetting are computed using our method and the approximate radiance value in Equation (3.33) for each point is computed by averaging the estimated radiance value ($\hat{L}_x$) of the point in multiple images. Pixels that are either saturated or underexposed are excluded in the averaging process. An HDR mosaic example is shown in the next section.

3.6 Experiments

In this section, we evaluate the performance of our proposed method. We test our algorithm by performing experiments with real data as well as synthetic data.

[2]http://www.hdrsoft.com

Figure 3.5: Synthetic Example : (Left) Image mosaic simulated with a response function, a vignetting function, and different exposures. (Right) Image mosaic after images are aligned to a common exposure setting with values estimated with our algorithm

3.6.1 Synthetic Example

We first evaluate our method with a synthetic example. An image is divided into multiple images that overlap and Gaussian white noise is added to each image. The RMS error of corresponding pixels was 5.01 after adding the noise. Then each image was transformed using Equation (3.3) with a known response function, vignetting function, and exposure values. The image mosaic built with these transformed images is shown in Figure 3.5. The RMS value after the transformation was 21.3. Applying our algorithm to this data set, we were able to recover the response function ($g()$), the vignetting function ($V()$), and the exposure values (K) accurately as shown in Figure 3.6. The mosaic built with the images aligned to a common exposure setting using the estimated values is shown in Figure 3.5. The RMS error of corresponding pixels after correction was 5.66.

3.6.2 Real Examples

We now evaluate our algorithm with real data. We run our algorithm on images taken with a rotating camera and a freely moving camera. We then compare our estimation with the ground truth and also with the estimates resulting from algorithms proposed in Goldman and Chen (2005) and in Litvinov and Schechner (2005a).

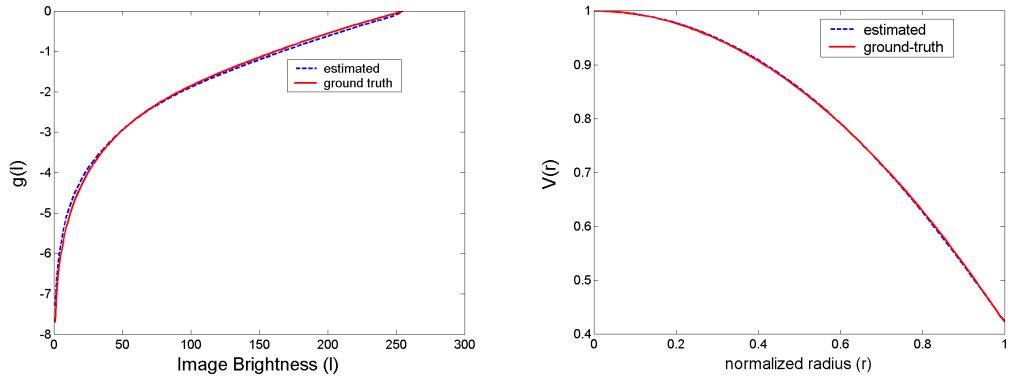


Figure 3.6: Synthetic Experiment. (Left) The response function (g, log inverse response) used for the simulation and the estimate (Right) The vignetting function used for the simulation and the estimate.

To compare the estimates of the response function with the ground truth, the estimated response function is plotted with the measurement from the image of a uniformly lit calibration chart (Macbeth chart) where the reflectance of patches are known. For vignetting, the ground truth is measured by taking images of a light box to image a surface with uniform radiance. The uniformity of the light box was checked with a light meter. The measured image values are then transformed to irradiance value using Equation (3.5) with the response function computed using the EMoR representation on images taken with different exposures where both the camera and the scene were static. The ground truth for the vignetting is then computed by taking the average of the irradiance for each radius value.

Our method assumes, like the methods of Goldman and Chen (2005) and Litvinov and Schechner (2005a), that the vignetting effect is the same for all images in the sequence. This means that the aperture size has to be fixed while taking pictures. To ensure this assumption, the exposures are changed by changing the exposure time with a fixed aperture. In most cameras, this can be done either in aperture priority mode or in manual mode.

The first experiment was carried out with images taken with a rotating camera. Six

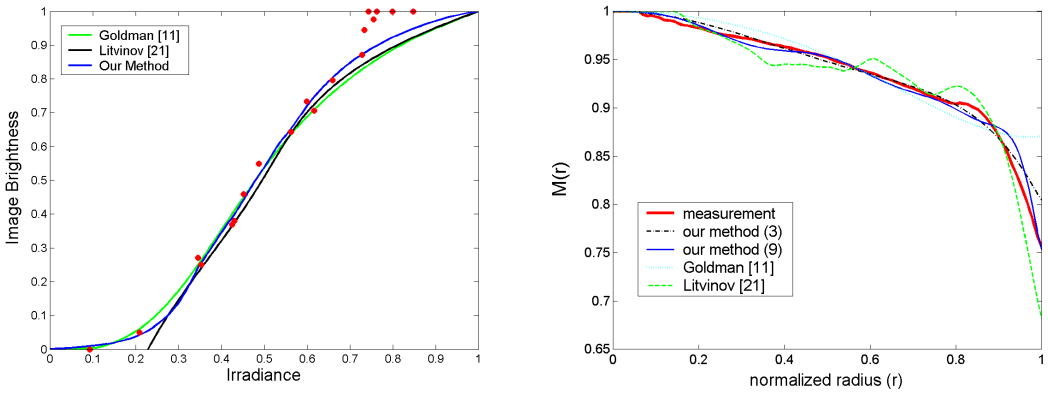


Figure 3.7: Result Comparisons: (Left) Recovered inverse response functions (f^{-1}) and the measurement from the MacBeth chart as ground truth (dots). (Right) Recovered vignetting function and the measured ground truth. Numbers inside the parenthesis indicate the number of parameters used

differently exposed images were taken with a Sony F-717 camera. The image mosaic constructed with this image set is shown in Figure 3.8. Figure 3.7 compares the estimation of the response function and the vignetting function by our method with the ground truth as well as the results from the methods in Goldman and Chen (2005) and Litvinov and Schechner (2005a). Experiments for the methods Goldman and Chen (2005) and Litvinov and Schechner (2005a) were carried out with code provided by the authors. We modified the method of Litvinov and Schechner (2005a) since the method is limited to grayscale images only. We used the same assumption as our method, exposures change independently for each color channel, to explain the white balance. The comparison of image mosaics constructed with radiometrically aligned images by the estimates from each method is shown in Figure 3.8. The error histograms of the corresponding pixels in the mosaic along with the RMS errors are provided in Figure 3.9. The error histograms provide more information about the performance in this case since the RMS errors can be largely affected by mismatches. We also ran another experiment with the data set used in Goldman and Chen (2005) and the results are shown in Figure 3.10 and Figure 3.11.

From the experiments, the estimations using our method were the closest to the ground truth and also yielded the minimum errors resulting in most seamless mosaics. Note that all methods tend to be less accurate in the high/low intensity region for the response function and the high radius value regions due to lack of data in the region. The method in Litvinov and Schechner (2005a) showed lack of robustness against noise and mismatches. As can be seen from the second example (Figure 3.10), there were a large amount of mismatches in this sequence because there were a lot of high frequency components such as tree branches in this sequence. Use of a nonparametric model resulted in lack of accuracy due to the large number of outliers. While the method in Goldman and Chen (2005) gave better results than the method in Litvinov and Schechner (2005a), the problem with this method was the speed of the estimation. Since it relies on a nonlinear optimization, the process is very slow compared to the other two methods. Hence, it is very difficult to increase the number of parameters or samples which will increase the estimation time even more. The estimation by our method was very accurate even against the outliers and the speed of our algorithm is nearly as fast as the method in Litvinov and Schechner (2005a) since the solution is in large part acquired linearly.

The goal of the next experiment was to evaluate our algorithm with the data captured with a moving camera. We used the same camera (Sony F-717) to capture images of an object while freely moving the camera and changing the exposure. We ran the algorithm in Pollefeys et al. (2004) on the data set for the stereo matching and building the 3D model which is shown in Figure 3.12. The radiometric response function, the vignetting function, and the exposures were estimated by running our algorithm using the correspondences from the stereo matching. As shown in Figure 3.3, there are significant amount of outliers in stereo data sets which will make getting robust estimation difficult using existing methods. However, using our method, we were able to get robust estimations as shown in Figure 3.13 which shows that the recovered response function

Figure 3.8: (First) Image mosaic constructed with differently exposed images. All images are aligned to the mean exposure value and vignetting corrected using (second) our method, (third) the method by Goldman and Chen (2005), and (last) the method by Litvinov and Schechner (2005a). Note that the discrepancy in the sky after the alignment is due to saturation.

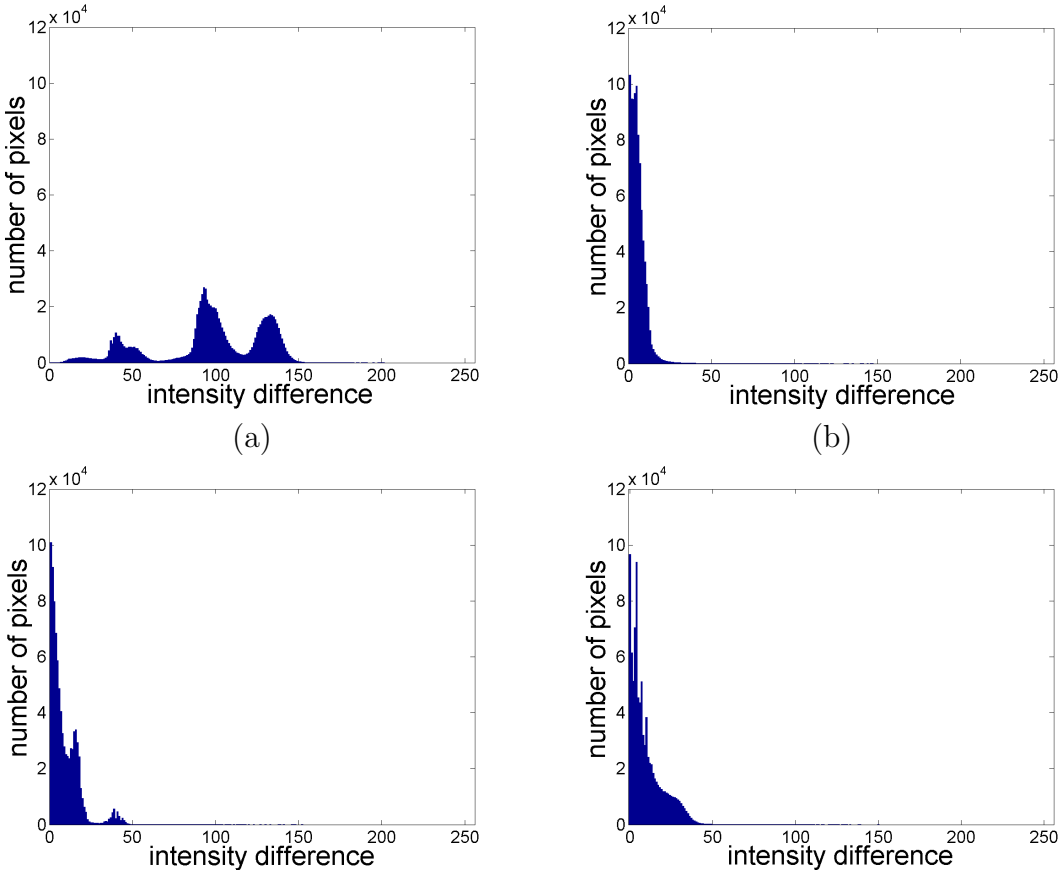


Figure 3.9: Error histograms of corresponding pixels in the mosaics shown in Figure 3.8. (a) original (RMS error = 94.25), (b) our method (RMS error = 8.48), (c) method in Goldman and Chen (2005) (RMS error = 17.26), (d) method in Litvinov and Schechner (2005a) (RMS error = 16.29).

Figure 3.10: (First) Image mosaic constructed with differently exposed images. All images are aligned to the mean exposure value and vignetting corrected using our method (second), the method in Goldman and Chen (2005) (third), and the method Litvinov and Schechner (2005b) (last). The images are provided by Dan Goldman (Goldman and Chen, 2005).

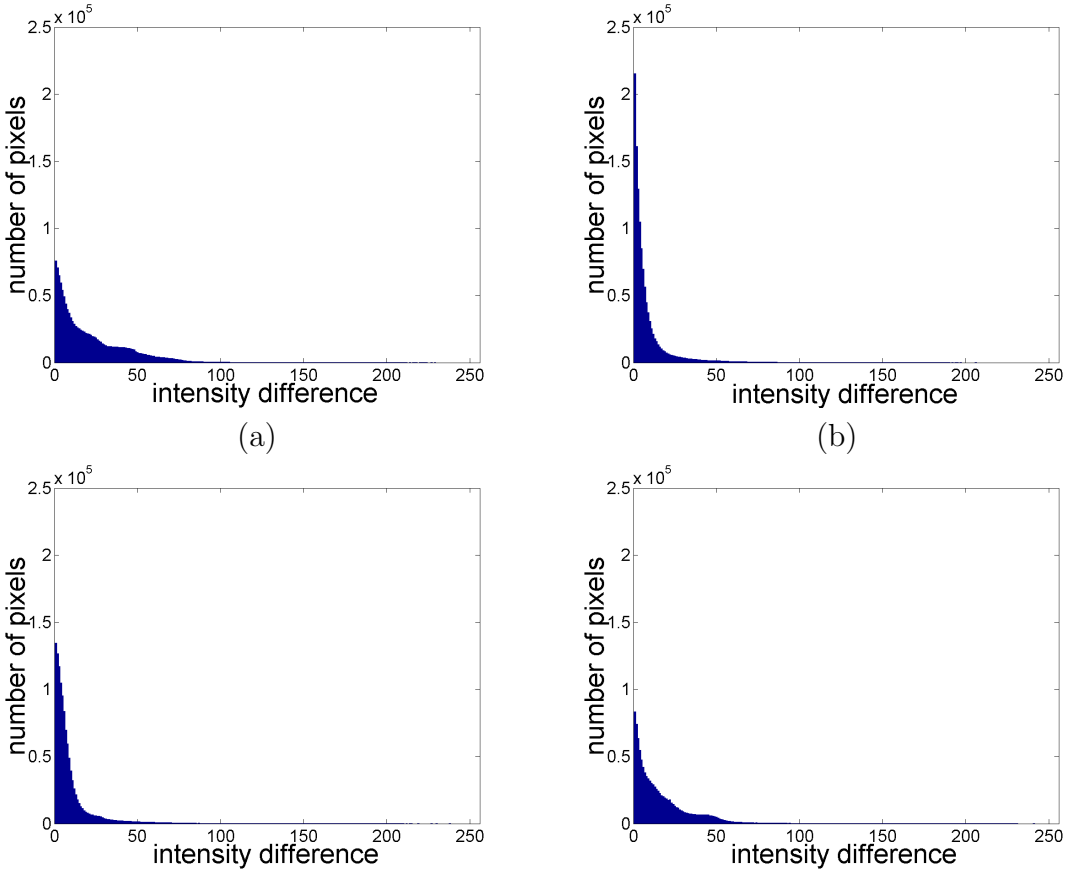


Figure 3.11: Error histograms of corresponding pixels in the mosaics shown in Figure 3.10. (a) original (RMS error = 31.17), (b) our method (RMS error = 16.85), (c) method in Goldman and Chen (2005) (RMS error = 17.97), (d) method in Litvinov and Schechner (2005a) (RMS error = 23.45).

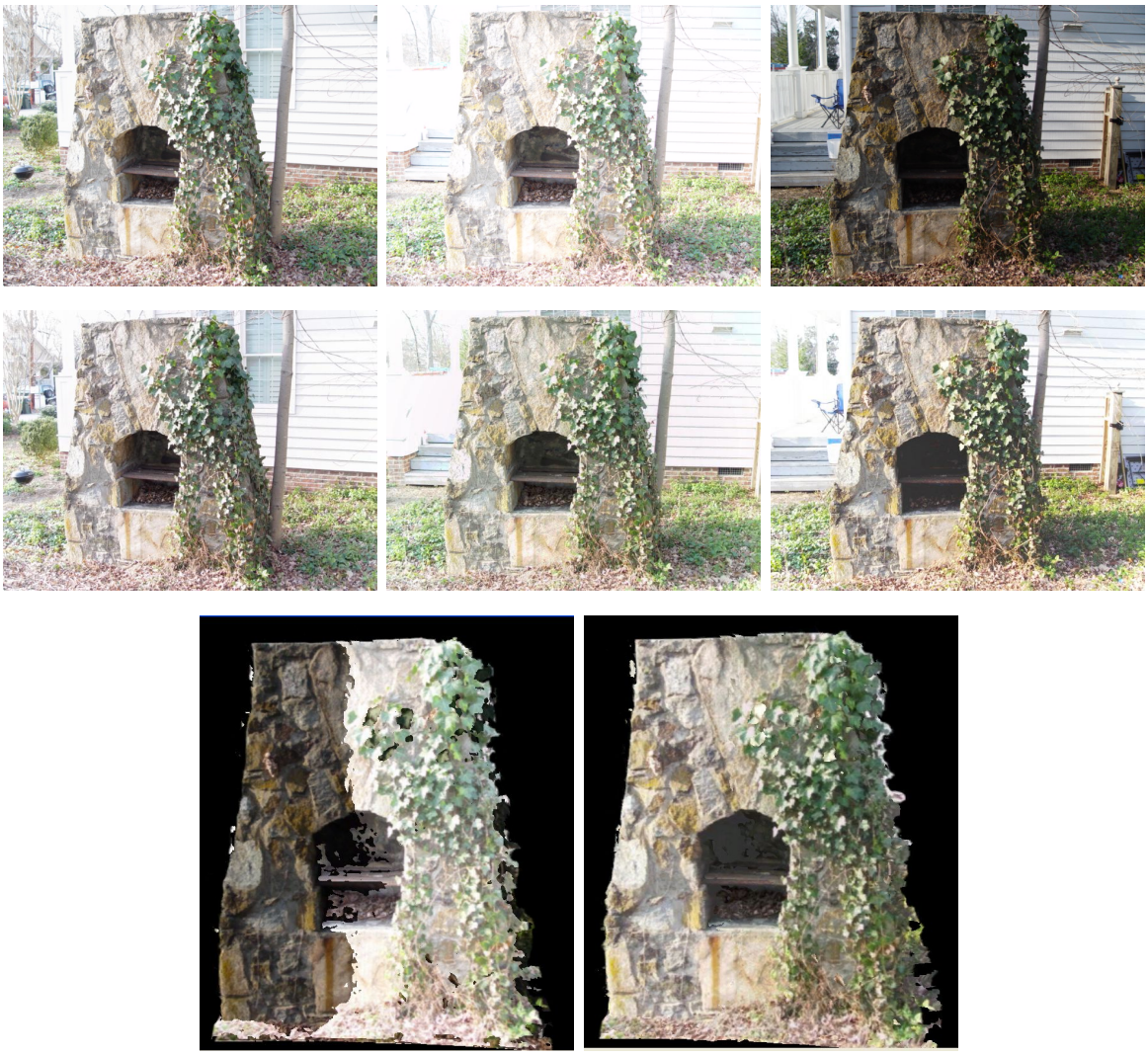


Figure 3.12: (Top) Some samples from the stereo sequence, (middle) radiometrically aligned images using our method, (bottom) texturemapped 3D models before and after the alignment

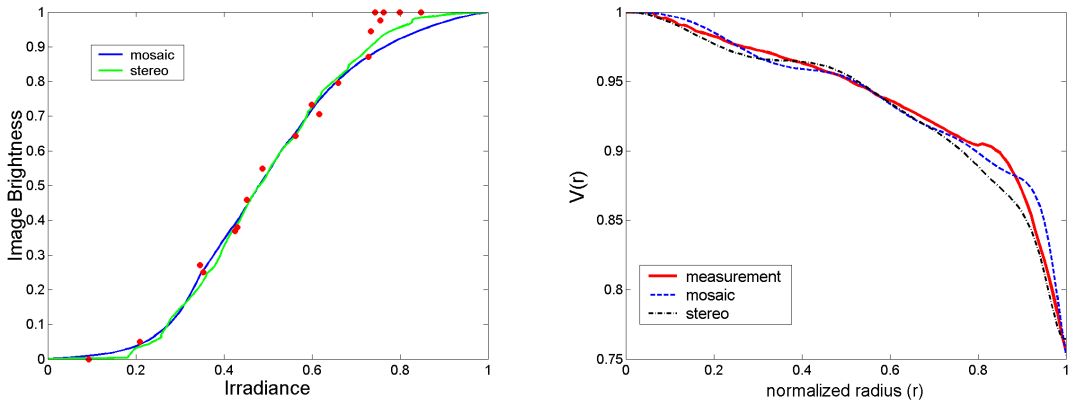


Figure 3.13: Comparison of the results from mosaic sequence and stereo sequence. (Left) The recovered inverse response functions and the vignetting functions (right)

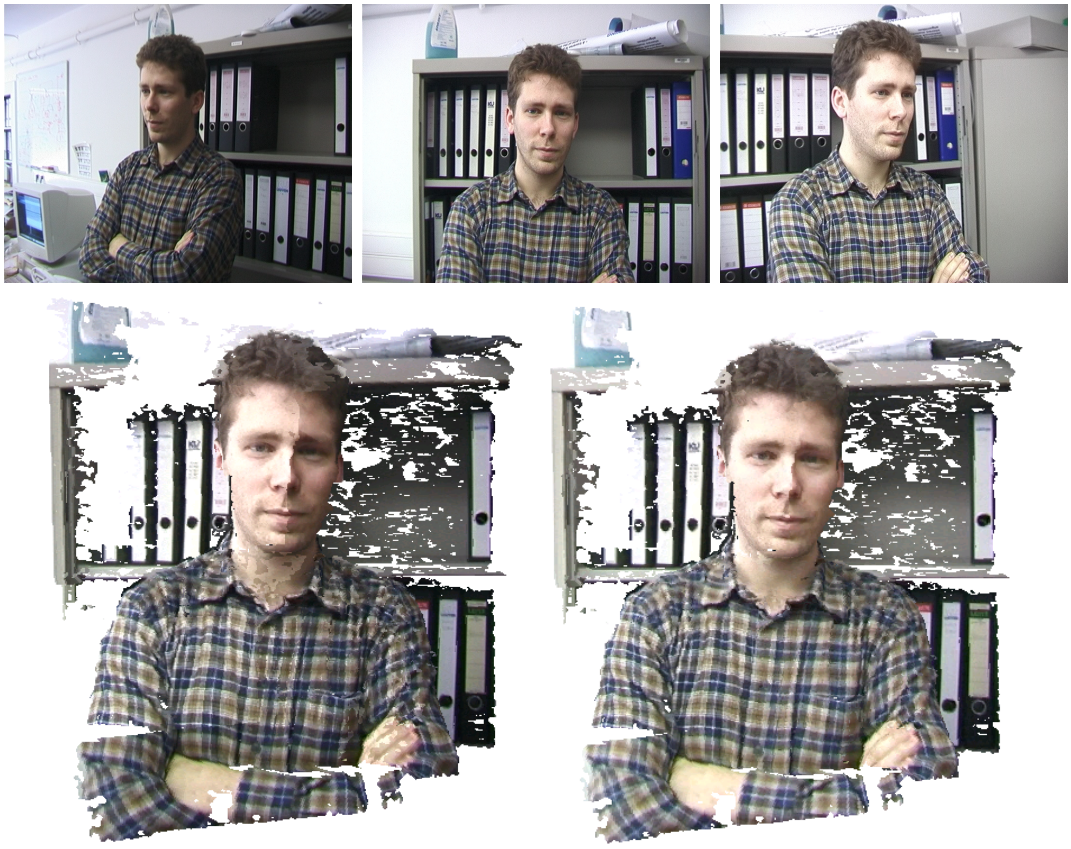


Figure 3.14: Another example of stereo sequence alignment. (Top) Some samples from the stereo sequence. (Bottom) Texturemapped 3D model before the alignment (left) and after the alignment (right). In this example, there are still some artifacts remaining (most visibly on the nose) after the alignment due to view-dependent highlights which are not compensated for by our method.

and the vignetting function is very close to the ones recovered from the mosaic sequence and hence to the real data. Figure 3.12 shows a few samples of radiometrically aligned images as well as the 3D model texturemapped using those images. Another stereo example is shown in Figure 3.14.

3.6.3 High Dynamic Range Mosaic

As the final experiment, we show an HDR mosaic example as explained in the previous section. With the images shown on the top of Figure 3.15, we first estimated the response function, the vignetting function, and the exposures of each image. When the estimation is complete, the approximate radiance value of each point in the mosaic is computed and the radiance map for the example is tonemapped for displaying purposes. Figure 3.15 shows the difference between the radiometric image alignment and the HDR imaging. While the mosaic is seamless after the alignment, the outside scene is still saturated and parts of the inside are too dark to recognize. The tone-mapped HDR mosaic is able to represent the scene in front of the camera more realistically than the normal mosaic.

3.7 Discussion

In this chapter, we have proposed a novel radiometric calibration method for estimating the radiometric response function, exposures, and vignetting. By decoupling vignetting from the response function, we can approach each problem with a robust estimation method. The robustness of our method was validated synthetically and also with real examples. Our method accurately estimates the parameters even in the presence of large noise and mismatches including matches from stereo sequence whereas other existing methods were not effective against noise and outliers. We applied the estimation results to radiometrically align images for seamless mosaics and 3D model textures. We also used our method for creating the HDR mosaic which is more representative of the scene

Figure 3.15: HDR mosaic. (First) Original mosaic, (Second) A zoomed-in image added to the mosaic. The exposure and the vignetting function for this image is computed using the method described in Section 3.3. (Third) Radiometrically aligned mosaic (Last) HDR mosaic

than the normal mosaic.

Some may question the accuracy of our response function estimation process which uses only corresponding pixels between images that are of equal distance (with some tolerance) from the image center. Even though it may seem like we are using fewer samples than other methods, we are actually using more samples than the existing methods which rely on random sampling of points. For the method of Goldman and Chen (2005), only 1000-2000 samples were used for each sequence shown in this chapter. It is very difficult to use more samples because it will slow down the computation which is already very slow. While more samples can be used for the method of Litvinov and Schechner (2005a), it is still limited due to memory constraints when solving the linear equation. It is important to note that we do not use the pixel values directly to compute the response function as in other methods but rather use them to compute the brightness transfer function by dynamic programming. Given the robustness of this process along with the power of the model (EMoR) we use, the number of points we use in our method is not much of a problem. One case that would be problematic for our method is when the distribution of pixel values is very limited such as when the regions we use are of an uniform color. But this is usually not the case in practice especially since we can easily expand our method to include correspondences from more images, not just the next image in the sequence.

Chapter 4

Joint Feature Tracking and Radiometric Calibration from Auto-Exposure Video

4.1 Introduction

Extracting and tracking features is a fundamental step in many computer vision systems since it provides means to relate one image to another spatially. One of the most commonly used feature tracker, especially for processing videos, is the KLT (Kanade-Lucas-Tomasi) tracker (Lucas and Kanade (1981); Shi and Tomasi (1994)) due to its robustness and efficiency. However, there are cases that pose problems for the KLT tracker, mainly when images of a high dynamic range scene are captured. In order to capture the full brightness range of natural scenes, where parts are in the shadow and others are in bright sunlight for example, the camera has to adjust the exposure accordingly. As a result, the appearance of the same scene point varies throughout the video sequence, breaking the basic assumption for the KLT tracker that the brightness of the scene points stays constant. Hence, we need methods to find radiometric relationships between image features in addition to the spatial relationships.

In this chapter, I introduce a new method that models the changes in image brightness between images globally and nonlinearly rather than treating the variation locally and linearly by comparing local regions independently. The brightness change can be

explained by the radiometric response function which defines the mapping from the image irradiance to the image brightness. We first introduce a method for tracking features and estimating the exposure changes between frames when the camera's radiometric response function is known. In many cases the radiometric response function is not known a priori, so I also present a method for joint feature tracking and radiometric calibration by formulating the estimation of the response function within a linear feature tracking scheme that can deal with varying intensity values of features due to exposure changes. The novel framework presented here performs an integrated radiometric calibration in contrast to previous radiometric calibration techniques (including the method presented in Chapter 3) which require the correspondences as an input to the system which leads to a chicken-and-egg problem as precise tracking requires accurate radiometric calibration. By combining both into an integrated approach we solve this chicken-and-egg problem.

The remainder of the chapter is organized as follows. In the next section, a review of the related work will be presented. In Section 4.3, the KLT tracker algorithm will be reviewed. In Section 4.4, a method for tracking features when the response function is known will be introduced first, and then a method for simultaneous tracking and the response function estimation will be explained. We evaluate our method with experiments in Section 4.5, including an application of the method for an adaptive stereo system, and conclude with discussion about our algorithm in Section 4.6.

This work was originally presented in Kim et al. (2007) and the extended version of the original work is currently under review (Kim et al., 2008a).

4.2 Related Work

A problem that is similar to our work is the joint domain and range registration of images. In Mann (2000), Mann introduced a method for jointly computing the projective coordinate transform (domain) and the brightness change (range) between a pair of

images of a static scene taken with a purely rotating camera. The brightness transform which he calls the *comparametric function* is approximated by using a gamma function as the model for the response function. The joint registration process is linearized by the Taylor expansion and the least squares solution is acquired. In similar work, Candocia proposed a method for the joint registration by using a piecewise linear model for the comparametric function (Candocia, 2003).

Our method is different from the methods in Mann (2000) and Candocia (2003) in that we are interested in tracking of features that are allowed to move in an unconstrained manner, rather than estimating a global projective transform between images. Although this involves estimation of significantly more parameters, our algorithm is able to deal with it efficiently. In addition, we do not restrict the movement of the camera, and can also deal with moving objects in the scene. We are also different in that we compute the actual response function of the camera and the exposures, rather than just finding out the brightness transform between images.

One application of our work is a stereo system that is adaptive to brightness changes between images. Several stereo methods have employed matching metrics which achieve invariance to brightness changes. A comparison of these techniques is presented in Hirschmuller and Scharstein (2007). Normalized cross-correlation is effective for dealing with locally linear changes, while mutual information is invariant to arbitrary one-to-one mappings. However, mutual information has been only been successfully implemented as a global mapping between images. The rank transform has been shown to be robust even to local illumination changes. In general, more invariance can lead to ambiguity in some cases, and overfitting is possible. Furthermore, all these methods require known camera poses, or at least rectified images. In our system, we recover the camera poses from feature tracks. Using our method, the radiometric calibration is recovered jointly with the feature tracks, and therefore brightness invariance in stereo is unnecessary.

4.3 Kanade-Lucas-Tomasi (KLT) Tracker

We now review the KLT tracker (Lucas and Kanade (1981); Shi and Tomasi (1994)). The algorithm is based on the assumptions that the motion of the camera is small and the appearance of features stays constant between consecutive frames in the video sequence. The brightness constancy assumption is stated as follows :

$$J(\mathbf{x} + \mathbf{dx}) - I(\mathbf{x}) = 0 \tag{4.1}$$

where J and I are images at time $t+1$ and t respectively, $\mathbf{x} = [x, y]^T$ is the feature location, and $\mathbf{dx} = [dx, dy]^T$ is the displacement vector.

Linearizing Equation 4.1 using the Taylor expansion, we get

$$J_x dx + J_y dy + J_t = 0 \tag{4.2}$$

where $J_x = \frac{\partial J}{\partial x}$, $J_y = \frac{\partial J}{\partial y}$, and $J_t = J(\mathbf{x}) - I(\mathbf{x})$. Computing the displacement is under-constrained since there are two unknowns (dx and dy) with one equation (Equation 4.2). This is the aperture problem in which the component of the optical flow perpendicular to the gradient is unknown. To overcome this problem, Lucas and Kanade (1981) proposed to use the spatial coherence constraint which assumes that the neighbors of the pixel ($\mathbf{x}$) have the same motion $\mathbf{dx}$. Thus the displacements can be computed by minimizing the following energy :

$$E = \sum_{\mathbf{x} \in P} (J_x dx + J_y dy + J_t)^2 \tag{4.3}$$

Notice that the summation is over the patch (or neighborhood) P surrounding the feature. The displacements dx and dy are solved by minimizing the energy in Equation 4.3 ($\frac{\partial E}{\partial dx} = 0$ and $\frac{\partial E}{\partial dy} = 0$) as follows.

$$\begin{bmatrix} \sum_{\mathbf{x}\in P} J_x^2 & \sum_{\mathbf{x}\in P} J_x J_y \\ \sum_{\mathbf{x}\in P} J_x J_y & \sum_{\mathbf{x}\in P} J_y^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = - \begin{bmatrix} \sum_{\mathbf{x}\in P} J_x J_t \\ \sum_{\mathbf{x}\in P} J_y J_t \end{bmatrix} \tag{4.4}$$

The summation in Equation 4.4 is over a patch centered at the feature. In the KLT tracker, the features are extracted so that the matrix on the left side of the Equation 4.4 is well-conditioned and above the image noise level. This is determined by the two eigenvalues of the matrix and the points with two large eigenvalues are selected as features (Shi and Tomasi, 1994).

In practice, Equation 4.1 is rewritten as follows to make the equation symmetric with respect to both images :

$$J(\mathbf{x} + \frac{\mathbf{dx}}{2}) - I(\mathbf{x} - \frac{\mathbf{dx}}{2}) = 0 \tag{4.5}$$

Linearizing Equation 4.5 using the Taylor expansion and minimizing the error over a patch P as explained above, the displacement for each feature is computed as follows.

$$\begin{bmatrix} \sum_{\mathbf{x}\in P} {s_x}^2 & \sum_{\mathbf{x}\in P} s_x s_y \\ \sum_{\mathbf{x}\in P} s_x s_y & \sum_{\mathbf{x}\in P} {s_y}^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = 2 \begin{bmatrix} \sum_{\mathbf{x}\in P} (I(\mathbf{x}) - J(\mathbf{x})) s_x \\ \sum_{\mathbf{x}\in P} (I(\mathbf{x}) - J(\mathbf{x})) s_y \end{bmatrix} \tag{4.6}$$

where $s_x = J_x + I_x$ and $s_y = J_y + I_y$.

The dynamic range of cameras is usually too small to accommodate the large dynamic range of natural scenes. Accordingly, the exposure of the camera is adjusted causing the appearance of the features to change breaking the brightness constancy assumption. In the implementation by Birchfield, a simple method is used to account for the gain change between images (Birchfield, 1997). For each feature patch P in the first image, an individual gain is computed using the current estimate of the location of the patch P' in the second image. The gain ratio is computed by the ratio of mean intensity values of the two patches. The estimated ratio is used to normalize the intensity of the

neighborhoods of the point in the second image to proceed with the tracking process. In Baker et al. (2003) and Jin et al. (2001), illumination invariance is also achieved by solving for a gain and bias factor in each individually tracked patch. In all of these approaches, the change in intensity is treated locally for each individual feature. Also, the intensity change which is a nonlinear process is linearly approximated.

4.4 Joint Tracking and Radiometric Calibration Algorithm

We now introduce our method for brightness-invariant feature tracking and radiometric calibration. Given a video sequence with varying exposure, we estimate the radiometric response function of the camera, the exposure difference between frames, and the feature tracks from frame to frame. Our feature tracking, in contrast to previous approaches, models the global and nonlinear process that is responsible for changes in image brightness rather than adapting to the changes locally and linearly. Our radiometric calibration is different from previous calibration works because the correspondences are an output of our system rather than being an input to the system. Our method is an on-line process not a batch process, which allows subsequent algorithms such as stereo matching to compensate for brightness changes.

We will first start by explaining the method for tracking features when the response function is known, and then we will proceed to the method for the joint feature tracking and radiometric calibration.

4.4.1 Tracking Features with Known Response Function

We first explain the method for tracking features and estimating the exposure difference K between two images when the response function of the camera g (inverse response function in the log-domain) is known. From Equation 2.6, the brightness change for a

feature located at $\mathbf{x}$ with the displacement $\mathbf{dx}$ is explained as follows.

$$g(J(\mathbf{x}+\frac{\mathbf{dx}}{2}))-g(I(\mathbf{x}-\frac{\mathbf{dx}}{2}))=K. \tag{4.7}$$

We apply a Taylor expansion to the images (Equation (4.8)) and then to the response function (Equation (4.9)) to linearize the equation above.

$$g(J(\mathbf{x})+\nabla J(\mathbf{x})^T\frac{\mathbf{dx}}{2})-g(I(\mathbf{x})-\nabla I(\mathbf{x})^T\frac{\mathbf{dx}}{2})=K \tag{4.8}$$

Let $J(\mathbf{x})=J$, $I(\mathbf{x})=I$, and g' be the derivative of the response function g, we get

$$g(J)+g'(J)\nabla J^T\frac{\mathbf{dx}}{2}-\left[g(I)-g'(I)\nabla I^T\frac{\mathbf{dx}}{2}\right]-K=0. \tag{4.9}$$

Assuming equal displacement for all pixels of a patch P_i around each feature, the displacements for each feature $[dx_i, dy_i]^T$ and the exposure difference K are estimated by minimizing the following error function :

$$E(dx_i,dy_i,K)=\sum_{\mathbf{x}\in P_i}(\beta+a\frac{dx_i}{2}+b\frac{dy_i}{2}-K)^2 \tag{4.10}$$

with

$$a=g'(J(\mathbf{x}))J_x+g'(I(\mathbf{x}))I_x \tag{4.11}$$

$$b=g'(J(\mathbf{x}))J_y+g'(I(\mathbf{x}))I_y \tag{4.12}$$

$$\beta=g(J(\mathbf{x}))-g(I(\mathbf{x})). \tag{4.13}$$

The error function is minimized when all partial derivatives with respect to the unknowns are zero. Accordingly, the following equation needs to be solved for each

feature.

$$\begin{bmatrix} \mathbf{U_i} & \mathbf{w_i} \\ \mathbf{w_i}^T & \lambda_i \end{bmatrix} \mathbf{z_i} = \begin{bmatrix} \mathbf{v_i} \\ m_i \end{bmatrix} \tag{4.14}$$

where,

$$\mathbf{U_i} = \begin{bmatrix} \frac{1}{2}\sum_{P_i} a^2 & \frac{1}{2}\sum_{P_i} ab \\ \frac{1}{2}\sum_{P_i} ab & \frac{1}{2}\sum_{P_i} b^2 \end{bmatrix} \tag{4.15}$$

$$\mathbf{w_i} = \begin{bmatrix} -\sum_{P_i} a \\ -\sum_{P_i} b \end{bmatrix}, \quad \lambda_i = \sum_{P_i} 2 \tag{4.16}$$

$$\mathbf{v_i} = \begin{bmatrix} -\sum_{P_i} \beta a \\ -\sum_{P_i} \beta b \end{bmatrix}, \quad m_i = 2\sum_{P_i} \beta \tag{4.17}$$

$$\mathbf{z_i} = [dx_i, dy_i, K]^T \tag{4.18}$$

Note that the exposure difference K is global for all features and we can estimate the unknown displacements for all features (dx_i, dy_i, $i = 1$ to n) and the exposure K simultaneously by minimizing the following error.

$$E(dx_1, dy_1, ..., dx_n, dy_n, K) = \sum_{i=1}^{n} E(dx_i, dy_i, K) \tag{4.19}$$

Accordingly the unknowns ($\mathbf{z}$) are found by solving the following linear equation.

$$\mathbf{Az} = \begin{bmatrix} \mathbf{U} & \mathbf{w} \\ \mathbf{w}^T & \lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ m \end{bmatrix} \tag{4.20}$$

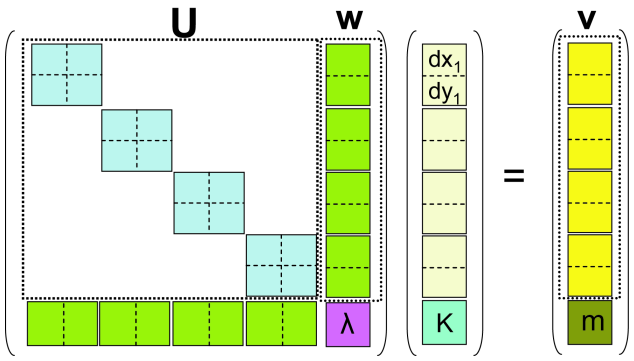


Figure 4.1: Solving for the displacements and the exposure : Illustration of Equation (4.20)

with

$$\mathbf{U} = \begin{bmatrix} \mathbf{U_1} & 0 & \dots & 0 \\ 0 & \mathbf{U_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & \mathbf{U_n} \end{bmatrix}, \quad \mathbf{w} = [\mathbf{w_1}, \dots, \mathbf{w_n}]^T \tag{4.21}$$

$$\lambda = \sum_{i=1}^{n} \lambda_i, \quad m = \sum_{i=1}^{n} m_i, \quad \mathbf{v} = [v_1, \dots, v_n]^T \tag{4.22}$$

$$\mathbf{z} = [dx_1, dy_1, \dots, dx_n, dy_n, K]^T \tag{4.23}$$

Figure 4.1 shows the structure of Equation (4.20). The matrix $\mathbf{A}$ is a sparse matrix and we can take advantage of its structure to find a computationally efficient solutions. Both sides of the Equation (4.20) are multiplied on the left by $\begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{w}^T\mathbf{U}^{-1} & 1 \end{bmatrix}$ resulting in

$$\begin{bmatrix} \mathbf{U} & \mathbf{w} \\ \mathbf{0} & -\mathbf{w}^T\mathbf{U}^{-1}\mathbf{w} + \lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ -\mathbf{w}^T\mathbf{U}^{-1}\mathbf{v} + m \end{bmatrix} \tag{4.24}$$

where $(-\mathbf{w}^T\mathbf{U}^{-1}\mathbf{w} + \lambda)$ is the Schur complement of the matrix $\mathbf{U}$. Since the inverse of $\mathbf{U}$ can be computed efficiently as it is a 2×2 block diagonal matrix (this inversion

corresponds to the amount of work necessary for the standard KLT) and its Schur complement is a scalar, Equation (4.24) can be solved very efficiently. The exposure difference K is given by

$$(-\mathbf{w}^T\mathbf{U}^{-1}\mathbf{w} + \lambda)K = -\mathbf{w}^T\mathbf{U}^{-1}\mathbf{v} + m \tag{4.25}$$

Once K is found, we can solve for the displacements. For each patch i, dx_i and dy_i are computed by back substituting K as in Equation (4.26). Hence the proposed estimation adds one additional equation (Equation (4.25)) to solve to the standard KLT tracking equations.

$$\mathbf{U_i}\begin{bmatrix} dx_i \\ dy_i \end{bmatrix} = \mathbf{v_i} - K\mathbf{w_i} \tag{4.26}$$

4.4.2 Joint Tracking and Radiometric Calibration

We now discuss the case of unknown response function. Given a video sequence, we automatically compute the radiometric response function g, the exposure difference between frames K, and the feature tracks.

We again use the Empirical Model of Response (EMoR) introduced in Grossberg and Nayar (2004) for the camera response function.

$$g(I) = \mathbf{g}_0(I) + \sum_{k=1}^{M} c_k \mathbf{h}_k(I) \tag{4.27}$$

where g_0 is the mean function and c_k's are the coefficients for the basis functions $\mathbf{h}_k$'s. For the work introduced in this chapter, we used a third order approximation ($M = 3$) since the first three basis functions explain more than 99.6% of the energy (Grossberg and Nayar, 2004). The derivative of the response function is similarly a linear combination of the derivatives of the basis functions.

$$g'(I) = \mathbf{g}'_0(I) + \sum_{k=1}^{M} c_k \mathbf{h}'_k(I) \tag{4.28}$$

Substituting g and g' in Equation (4.9) with Equation (4.27) and Equation (4.28), we get the following equation.

$$d + a \cdot dx + b \cdot dy + \sum_{k=1}^{M} c_k r_k + \sum_{k=1}^{M} \alpha_k p_k + \sum_{k=1}^{M} \beta_k q_k - K = 0 \tag{4.29}$$

The known variables for Equation (4.29) are :

$$a = \frac{g'_0(J)J_x + g'_0(I)I_x}{2} \tag{4.30}$$

$$b = \frac{g'_0(J)J_y + g'_0(I)I_y}{2} \tag{4.31}$$

$$r_k = h_k(J) - h_k(I) \tag{4.32}$$

$$p_k = \frac{h'_k(J)J_x + h'_k(I)I_x}{2} \tag{4.33}$$

$$q_k = \frac{h'_k(J)J_y + h'_k(I)I_y}{2} \tag{4.34}$$

$$d = g_0(J) - g_0(I) \tag{4.35}$$

The unknowns are the displacements dx and dy, the coefficients for the response function c_k ($k = 1$ to M), the exposure difference K, and variables introduced for linearization $\alpha_k = c_k dx$ and $\beta_k = c_k dy$.

Again, we assume equal displacement for all pixels in a patch around each feature and minimize the following error function to solve for the unknowns.

$$E(dx_i, dy_i, c_1, \ldots, c_M, \alpha_{i1}, \ldots, \alpha_{iM}, \beta_{i1}, \ldots, \beta_{iM}, K) =$$

$$\sum_{P_i} (d + adx_i + bdy_i + \sum_{k=1}^{M} c_k r_k + \sum_{k=1}^{M} \alpha_{ik} p_k + \sum_{k=1}^{M} \beta_{ik} q_k - K)^2 \tag{4.36}$$

Setting all partial derivatives towards the unknowns to zero, we get following equation for each feature.

$$\underbrace{\begin{bmatrix} \mathbf{U_i} & \mathbf{W_i} \\ \mathbf{W_i}^T & \mathbf{\Lambda_i} \end{bmatrix}}_{\mathbf{A_i}} \mathbf{z_i} = \begin{bmatrix} \mathbf{v_i} \\ \mathbf{m_i} \end{bmatrix} \tag{4.37}$$

Now we can solve for all feature tracks and the global parameters for the response function and the exposure difference similar to the case of known response function.

$$\mathbf{Az} = \begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^T & \Lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ \mathbf{m} \end{bmatrix} \tag{4.38}$$

with

$$\mathbf{U} = \begin{bmatrix} \mathbf{U_1} & 0 & \dots & 0 \\ 0 & \mathbf{U_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & \mathbf{U_n} \end{bmatrix}, \quad \mathbf{W} = [\mathbf{W_1}, \dots \mathbf{W_n}]^T \tag{4.39}$$

$$\mathbf{\Lambda} = \sum_{i=1}^{n} \mathbf{\Lambda_i}, \quad \mathbf{v} = [\mathbf{v_1}, \dots, \mathbf{v_n}]^T, \mathbf{m} = \sum_{i=1}^{n} \mathbf{m_i} \tag{4.40}$$

$$\mathbf{z} = [\varphi_1, \dots, \varphi_n, c_1, \dots, c_M, K]^T \tag{4.41}$$

where

$$\varphi_i = [dx_i, \alpha_{i1}, \dots, \alpha_{iM}, dy_i, \beta_{i1}, \dots, \beta_{iM}]^T \tag{4.42}$$

Notice that Equation (4.38) has the same structure as Equation (4.20) (Figure 4.2) except that the size of each sub-matrices are bigger. $\mathbf{U_i}$'s are $(2M+2) \times (2M+2)$, $\mathbf{W_i}$'s are $(2M+2) \times (M+1)$, and $\mathbf{\Lambda_i}$'s are $(M+1) \times (M+1)$. Multiplying both sides on

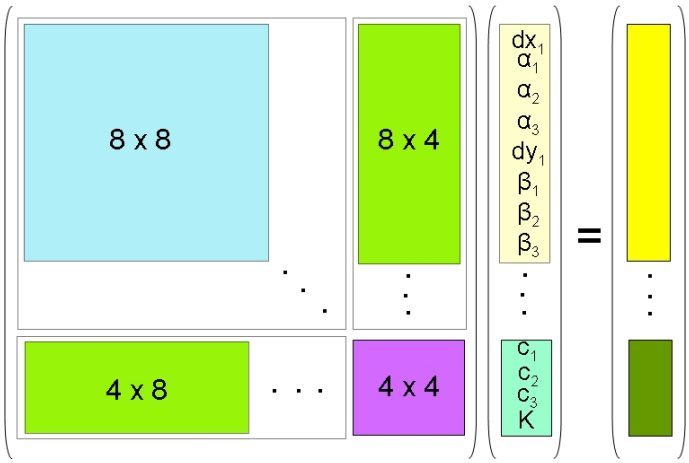


Figure 4.2: Solving for the radiometric response function (3 basis functions), exposures, and the feature displacements : Illustration of Equation (4.38).

the left by $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{W}^T\mathbf{U}^{-1} & \mathbf{I} \end{bmatrix}$ results in

$$\begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{0} & -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{W} + \Lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{v} + \mathbf{m} \end{bmatrix} \tag{4.43}$$

The coefficients of the response function and the exposure can be solved by

$$(-\mathbf{W}^T\mathbf{U}^{-1}\mathbf{W} + \Lambda)\upsilon = -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{v} + \mathbf{m} \tag{4.44}$$

where

$$\upsilon = [c_1, \ldots, c_M, K] \tag{4.45}$$

The solution to Equation (4.44) will suffer from the exponential ambiguity (or γ ambiguity) explained in the previous chapter which means that if a response function g and an exposure K are the solution to the problem so are γg and γK (Grossberg and Nayar, 2004). In Chapter 3, we fixed the scale by fixing the exposure ratio of an image pair. In this work, we chose to set the value of the response function at the image value

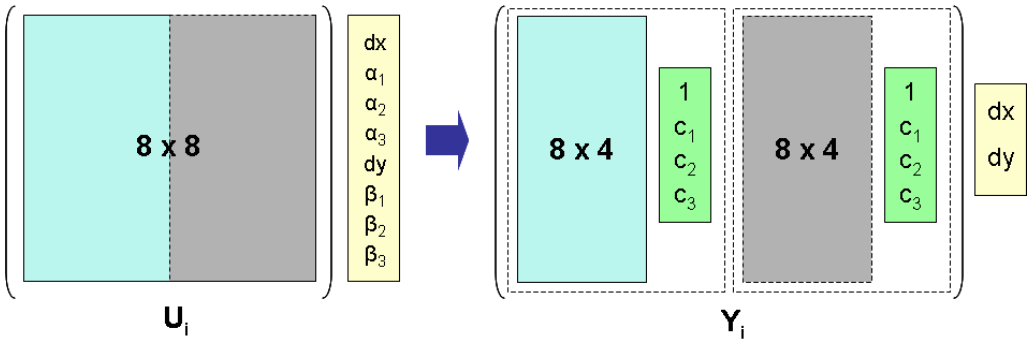


Figure 4.3: Factorization for estimating the tracks (M=3).

at 128 to a value τ to deal with the exponential ambiguity. This is done by adding the following equation to Equation (4.44).

$$\omega \sum_{k=1}^{M} c_k \mathbf{h_k}(128) = \omega(\tau - \mathbf{g_0}(128)) \tag{4.46}$$

The value ω in the equation controls the strength of the constraint.

The displacement for each feature can then solved by back substituting the solution υ to Equation (4.37).

$$\mathbf{U_i}\varphi_i = \mathbf{v_i} - \mathbf{W_i}\upsilon \tag{4.47}$$

Notice that α_{ik}'s and β_{ik}'s in φ_i are the products of the displacement and the response function coefficients: $\alpha_{ik} = c_k dx_i$ and $\beta_{ik} = c_k dy_i$. Since we have already estimated the coefficients c_k's, we can factorize the unknowns in φ_i as follows.

$$\begin{bmatrix} dx_i & \alpha_{i1} & \dots & \alpha_{iM} \\ dy_i & \beta_{i1} & \dots & \beta_{iM} \end{bmatrix} = \begin{bmatrix} dx_i \\ dy_i \end{bmatrix} \underbrace{\begin{bmatrix} 1 & c_1 & \dots & c_M \end{bmatrix}}_{\underline{\mathbf{c}}} \tag{4.48}$$

Using the factorization, Equation (4.47) can be put into a simpler form as follows (Fig-

ure 4.3).

$$\mathbf{Y}_i \begin{bmatrix} dx_i \\ dy_i \end{bmatrix} = \mathbf{v_i} - \mathbf{W_i}\upsilon \tag{4.49}$$

$$\begin{aligned} \mathbf{Y}_i(1:2M+2,1) &= \mathbf{U}_i(1:2M+2,1:M+1)\underline{\mathbf{c}}^T \\ \mathbf{Y}_i(1:2M+2,2) &= \mathbf{U}_i(1:2M+2,M+2:2M+2)\underline{\mathbf{c}}^T \end{aligned} \tag{4.50}$$

4.4.3 Updating the Response Function Estimate

In Section 4.4.2, we introduced the method for computing the response function, the exposure difference, and the feature tracks at the same time given an image pair from a video sequence. We now explain how we can integrate the estimates of the response function from each pair of images using a Kalman filter (Kalman, 1960; Welch and Bishop, 1995). The state is the coefficients of the response function ($\phi = [c_1, \ldots, c_M]^T$) and it is assumed to remain constant. Hence the process noise covariance was set to zero and the time update equations used are

$$\begin{aligned} \widehat{\phi}_k^- &= \widehat{\phi}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{P}_{k-1} \end{aligned} \tag{4.51}$$

where $\widehat{\phi}$ is the estimate of the state and $\mathbf{P}$ is the estimate error covariance matrix. The measurement update equations are

$$\begin{aligned} \kappa_k &= \mathbf{P}_k^-(\mathbf{P}_k^- + \mathbf{R})^{-1} \\ \widehat{\phi}_k &= \widehat{\phi}_k^- + \kappa_k(\mathbf{z}_k - \widehat{\phi}_k^-) \\ \mathbf{P}_k &= (\mathbf{I} - \kappa_k)\mathbf{P}_k^- \end{aligned} \tag{4.52}$$

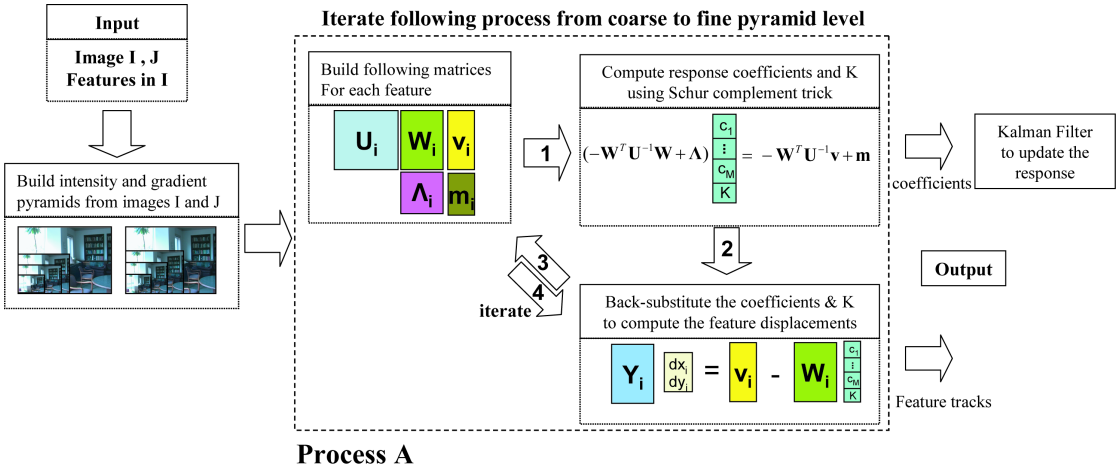


Figure 4.4: Overview of our algorithm

where κ is the Kalman gain, $\mathbf{z}_k$ is the measurement which is the pair-wise estimate of the response function in our case, and $\mathbf{R}$ is the measurement noise covariance.

Let $\mathbf{D} = (-\mathbf{W}^T\mathbf{U}^{-1}\mathbf{W} + \Lambda)$ and $\mathbf{b} = -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{v} + \mathbf{m}$ from Equation (4.44), the covariance matrix $\mathbf{R}$ is computed as follows.

$$\mathbf{R} = (\mathbf{D}^T\mathbf{D})^{-1}((\mathbf{D}\upsilon - \mathbf{b})^T(\mathbf{D}\upsilon - \mathbf{b})) \tag{4.53}$$

The Kalman estimate of the response function $\widehat{\phi} = [\hat{c_1}, \ldots, \hat{c_M}]^T$ is incorporated into the response function estimation in the next frame in the sequence where the problem becomes estimating Δc_k as follows.

$$g(I) = \mathbf{g}_0(I) + \sum_{k=1}^{M}(\hat{c_k} + \Delta c_k)\mathbf{h}_k(I) \tag{4.54}$$

4.4.4 Multi-scale Iterative Algorithm

Figure 4.4 shows the overview of our algorithm for the method explained in Section 4.4.2. As with the standard KLT tracker implementation, our algorithm runs iteratively on multiple scales. Image intensity and gradient pyramids are first built and the computa-

tion (process A in Figure 4.4) starts from the coarsest level pyramid to the finest level. The process A in Figure 4.4 is iterated multiple times for each pyramid level. Strictly speaking, the response estimate in the pyramid level other than the pixel level is an approximation of the real response function since the smoothing operation for building pyramids is done with the image values affected by the nonlinear response function. However, this does not affect our final estimates since the final estimation is done at the pixel level. The output of the algorithm are the coefficients for the response function which are fed to the Kalman filter (Section 4.4.3), the exposure difference K, and the tracked features which become input for the next pair of frames. Notice that we can start the tracking process with unknown response function and switch to tracking with known response function explained in Section 4.4.1 when the estimate of the response function gets stable.

4.5 Experiments

4.5.1 Experiment with Synthetic Data

We first evaluate our proposed methods with synthetic examples using evaluation images from Birchfield (1997). The brightness of an image can be changed from I to I' using Equation (4.55) with a response function g together with an exposure difference of K.

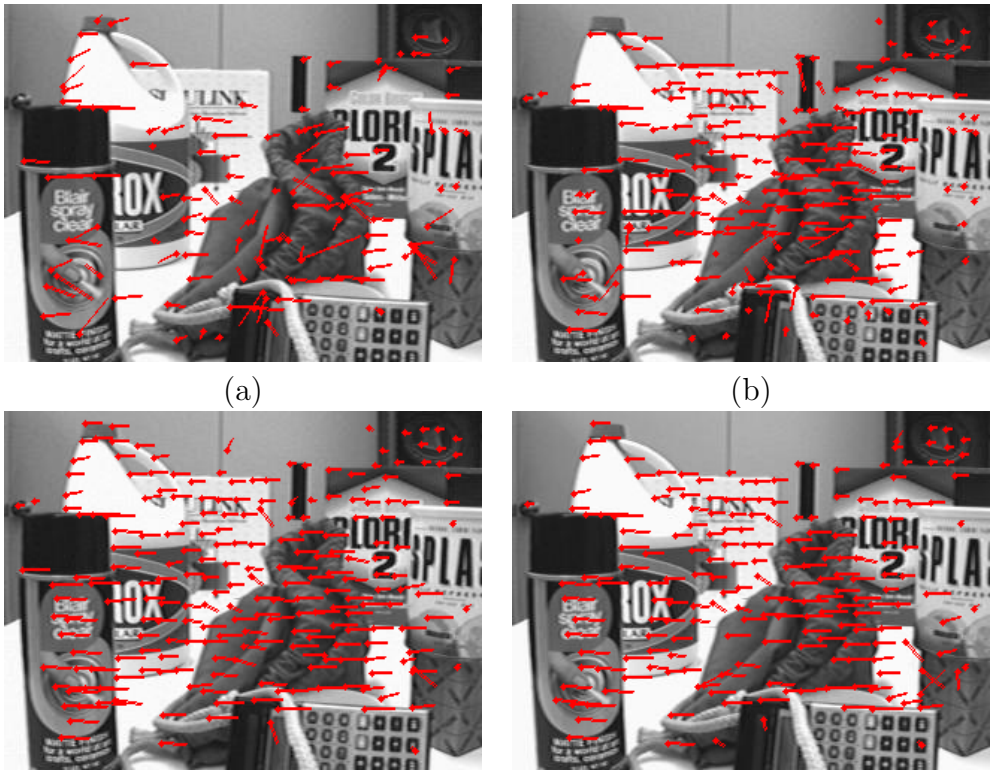
$$I' = g^{-1}(g(I) + K) \ . \tag{4.55}$$

The response function used for the evaluation with the synthetic data is shown in Figure 4.6. The exposure value applied for the examples from Figure 4.5 was 0.4. The feature tracking results using the standard KLT (Lucas and Kanade, 1981; Shi and Tomasi, 1994), the local adaptive KLT (Birchfield, 1997), our method with known response function in Section 4.4.1, and our method with unknown response function in

Section 4.4.2 are shown in Figure 4.5. As expected, the standard KLT does not perform well under the brightness change. Our experiments show that the local adaptive KLT mostly performs well when the camera motion and the brightness change are small (Figure 4.5). However, the performance significantly degrades when the change in motion or brightness increases as demonstrated in Figure 4.8. Tracking results using our methods, both with and without the knowledge of the response function, show superior results even with significant change in brightness which poses some problems for other tracking methods. The exposure value computed by our method was 0.404 with the known response function method and 0.408 with the unknown response function method (The scale of the response function was set with the known scale). We further tested our response function estimation algorithm by creating a synthetic sequence with 9 images with varying exposure values. Figure 4.6 shows some samples of the successive response function estimates and the final estimate along with the ground truth. Some estimates are less accurate in the lower intensity regions because the exposure difference was small in those image pairs. When the exposure difference is small, there are no changes in the brightness in the lower brightness regions giving no constraints to the estimation problem.

4.5.2 Experiments with Real Data

Similar results were observed in an experiment with a real video sequence. It was taken in a high dynamic range scene with a Canon GL2 camera. The exposure was automatically adjusted to a high value when the camera pointed to the dark inside area and it changed to a low value as the camera turned to the bright outside area. The comparison of tracks using the local-adaptive KLT, our method with known response function, and our method with unknown response function is shown in Figure 4.8. Both of our methods are able to track more features with significantly fewer errors when the changes in motion and brightness are relatively large as shown in the example.



(a) (b) (c) (d)

Figure 4.5: Feature tracking results (synthetic example) using : (first) standard KLT (second) local-adaptive KLT (third) our method with known response (fourth) our method with unknown response. Each line shows the movement of the feature from the previous frame to the current frame. Consistent feature displacements indicate good tracking. Images are from Birchfield (1997)

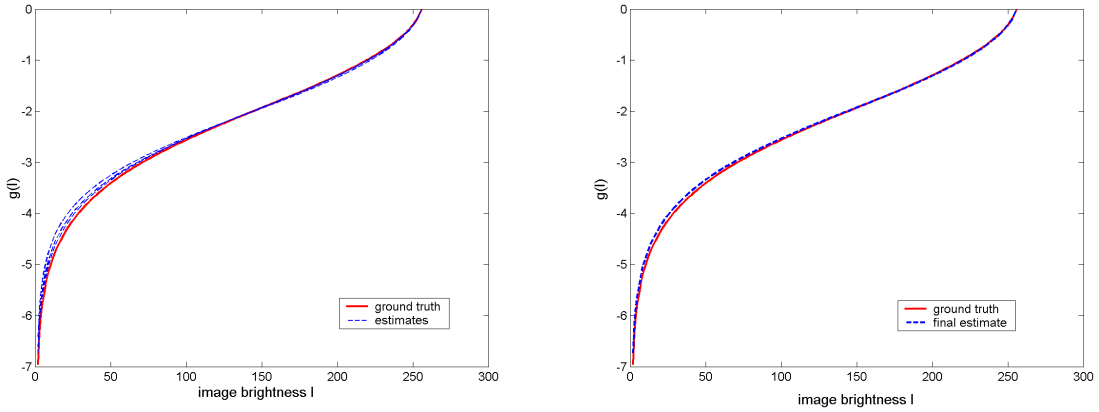


Figure 4.6: Camera response function estimation results. (First) Samples of response functions estimated from the synthetic sequence. (Second) Final estimate of the response function.

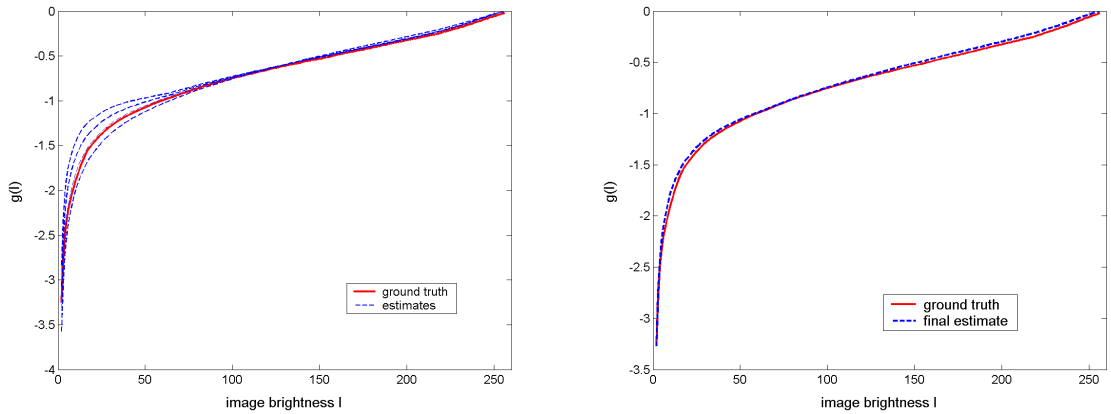


Figure 4.7: Camera Response Function Estimation (Real Data). (First) Samples of response functions estimated from the real video sequence (20 frames). (Second) Final estimate of the response function. The video can be seen at http://www.cs.unc.edu/∼sjkim/klt/track-response.wmv

Figure 4.7 shows the result of our response function estimation from this video. For the ground truth, we took multiple images of a static scene with a fixed camera changing the exposure value and fit the empirical model of response (EMoR) to the data as the method in the previous chapter. Samples of the response function estimates and the final estimate are compared with the ground truth in Figure 4.7.

We further verified our exposure estimation by comparing our estimates with the ground truth. Using a Point Grey Flea camera which has a linear response function, we took videos of scenes where the camera goes in and out of shadows causing the exposure to change frequently. The computed exposure values were compared with the values reported by the camera in Figure 4.9 and Figure 4.10.

The execution time for tracking 500 features in 720x480 images on a Pentium 4 processor (2.80 GHz) was 5 frames/second for the standard KLT, the local-adaptive KLT, and our method with known response. For our method with unknown response, the execution time was 0.2 frames/second which includes camera response and exposure estimation in addition to tracking. Only a few frames are necessary to compute the

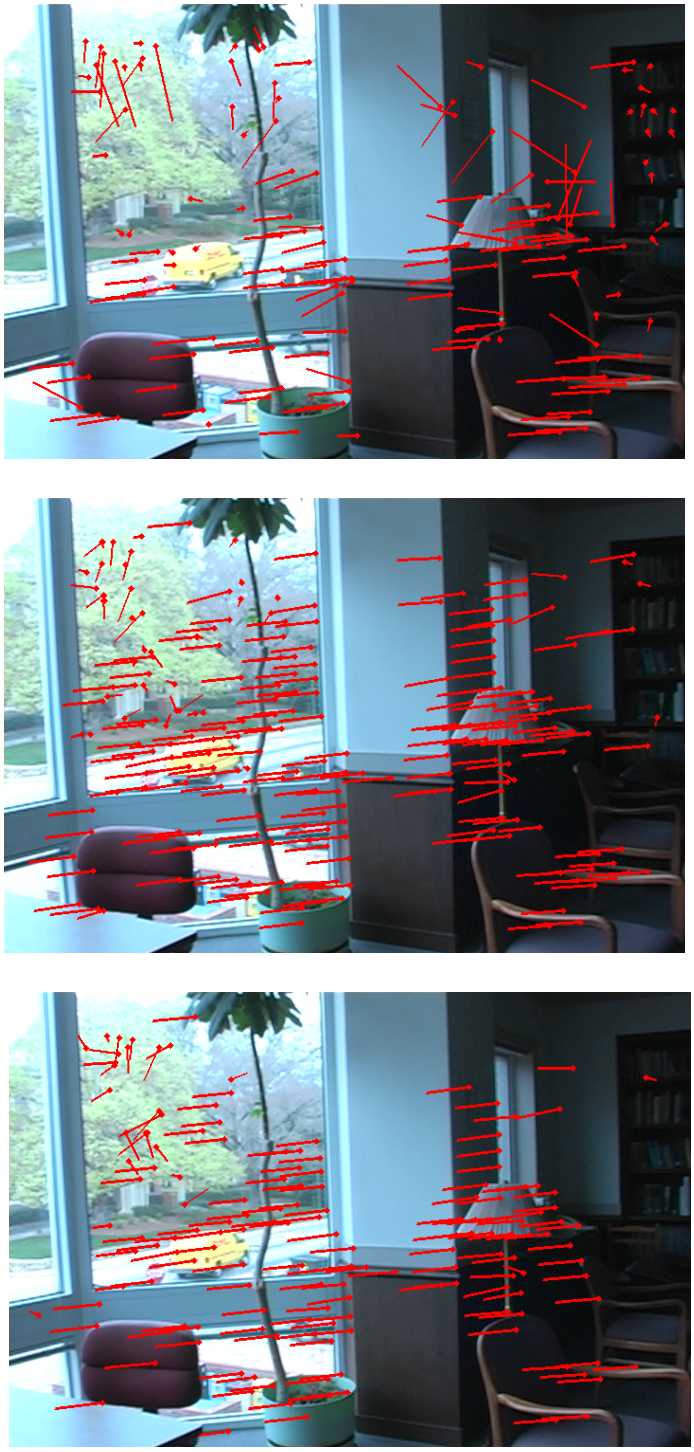


Figure 4.8: Feature Tracking using (First)Local-adaptive KLT (Second) Our method with known response (Third) Our method with unknown response. The video can be seen at http://www.cs.unc.edu/∼sjkim/klt/tracks.wmv

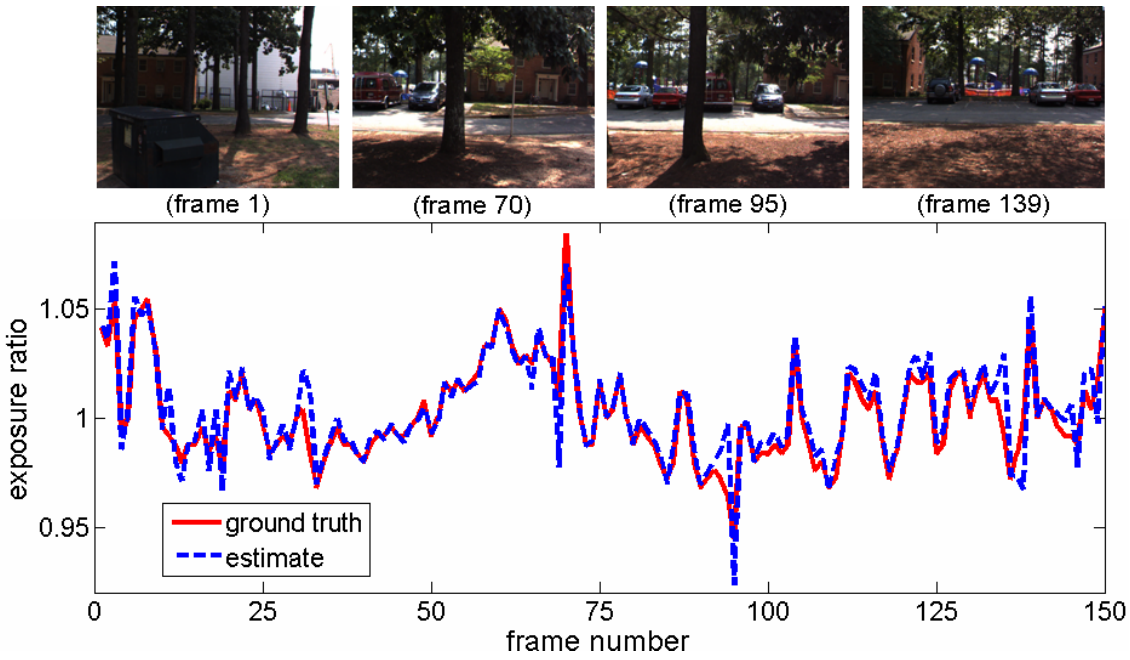


Figure 4.9: Exposure Estimation 1. The estimated exposure values are compared with the values reported by the camera.

response function and our method with the known response can be used for tracking afterwards. The overhead would be about 5% to 10% when tracking a 1-minute video. Zach et al. implemented our joint tracker with known (linear) response in GPU where the processing time is 213 frames/second for similar data sets (Zach et al., 2008).

Our method for joint feature tracking and radiometric calibration can be applied to perform structure from motion and stereo to recover a dense 3D surface. Using the tracked features, the camera motion can be computed using the technique presented in Pollefeys et al. (2004)[1]. Using our method, we are able to recover camera motion despite passing in and out of heavy shadows and even entering fully enclosed areas. Furthermore, feature tracks are continued over a larger number of frames, which is important for reducing drift in bundle adjustment. By using the recovered camera response function and the exposure differences between frames, we can use the simple stereo matching metric such as sum of absolute differences (SAD) instead of using metrics

[1]Camera intrinsics are precomputed manually.

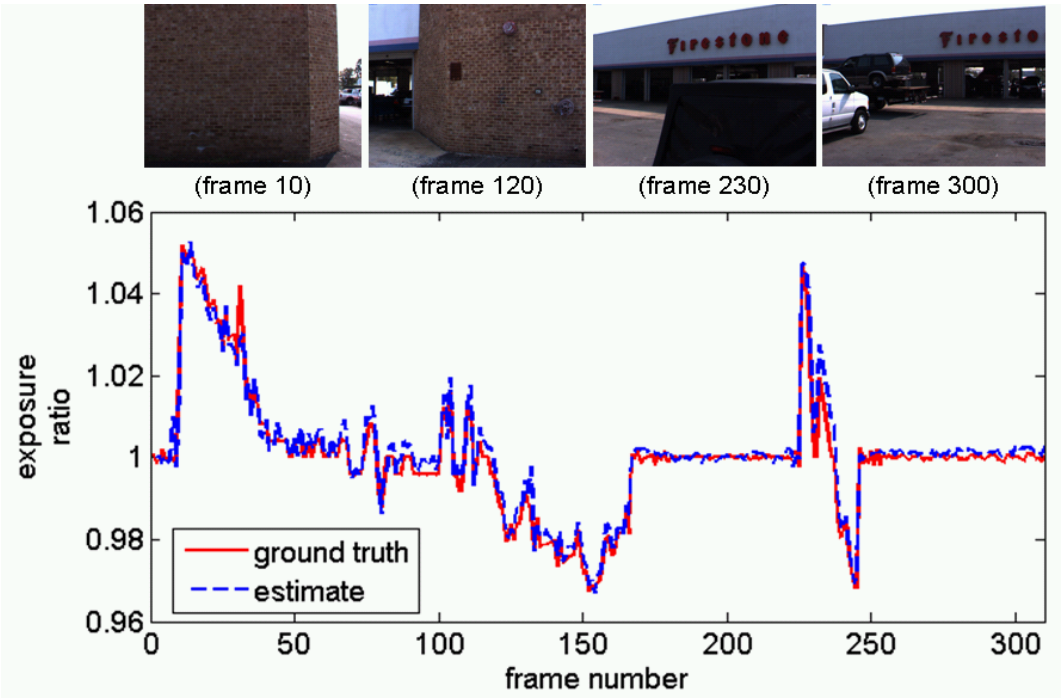


Figure 4.10: Exposure Estimation 2. The estimated exposure values are compared with the values reported by the camera.

that are invariant to brightness changes. Our matching function for a given pixel (x, y) and disparity (depth) d is as follows.

$$C(x, y, d) = \sum_{i,j \in W} |g(I(x + i, y + j))) - g(J(x + i - d, y + j)) - K| \tag{4.56}$$

The cost function is aggregated over a window W, and the disparity (depth) with minimum cost is selected. A plane-sweeping approach is used to handle multiple views simultaneously. More detailed explanation on the stereo system can be found in Pollefeys et al. (2008).

To evaluate the stereo algorithm, we used videos from two scenes with high dynamic range which caused the exposure to change significantly. Some sample images of the videos are shown in Figure 4.11 and Figure 4.12. For the first example, the exposure changes because the camera moves from a shadow to sunlight. Depth map computed

with our method which adapts for the exposure change shows superior results compared with the depth map computed without compensating for the exposure change as shown in the last row of Figure 4.11. The texture-mapped 3D model of the scene generated with our stereo system are also shown in Figure 4.11. The textures are radiometrically aligned to a constant exposure values using Equation 4.55. For the second stereo example, a video of a tunnel-like structure is taken starting from outside. This example is more challenging due to bigger exposure changes and more complex geometry of the scene. Depth map comparison and 3D models in Figure 4.12 show similar result as the first example. Note that some textures are radiometrically distorted in the model because the original pixels for those regions were saturated. An additional stereo result is shown in Figure 4.13.

4.6 Discussion

We have introduced a novel method that unifies the problems of feature tracking and radiometric calibration which includes exposure computation into a common framework. For feature tracking, it is commonly required that the brightness of features stays constant or the variations are dealt locally and linearly when the change is actually global and nonlinear. This limitation is not acceptable in many applications like ground reconaissance video for large scale outdoor scene modeling which needs to capture a high dynamic environment with a low dynamic camera system. To overcome these limitations, we proposed a joint feature tracking, radiometric response function and exposure estimation framework. This solves the chicken-and-egg problem in which the tracking requires accurate radiometric calibration for accuracy which in turn relies on precise tracks. Our computationally efficient algorithm takes advantage of the structure of the estimation problem which leads to a minimal computational overhead. With our joint estimation, we were able to advance the quality and robustness of the known structure

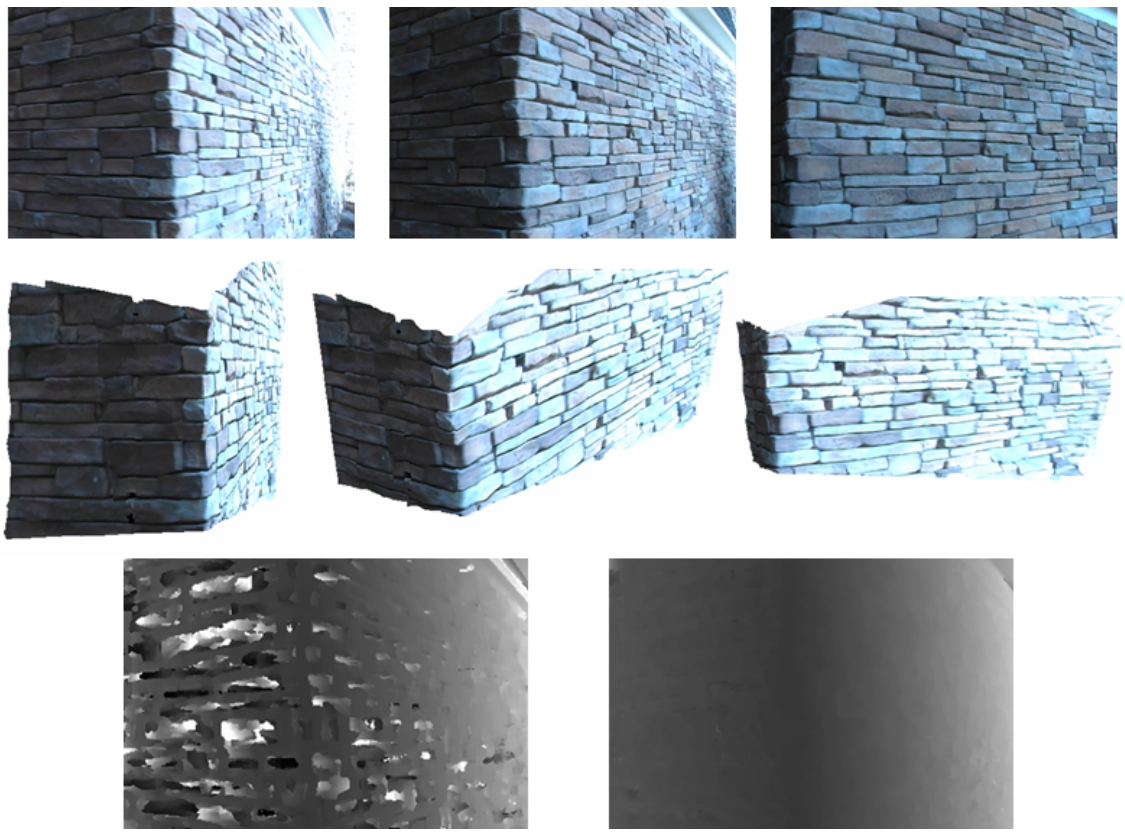


Figure 4.11: First stereo example. (Top) Sample images from the video sequence (Middle) Generated 3D model with radiometrically aligned textures (Bottom) Depth maps computed without exposure compensation (left) and with our method (right)

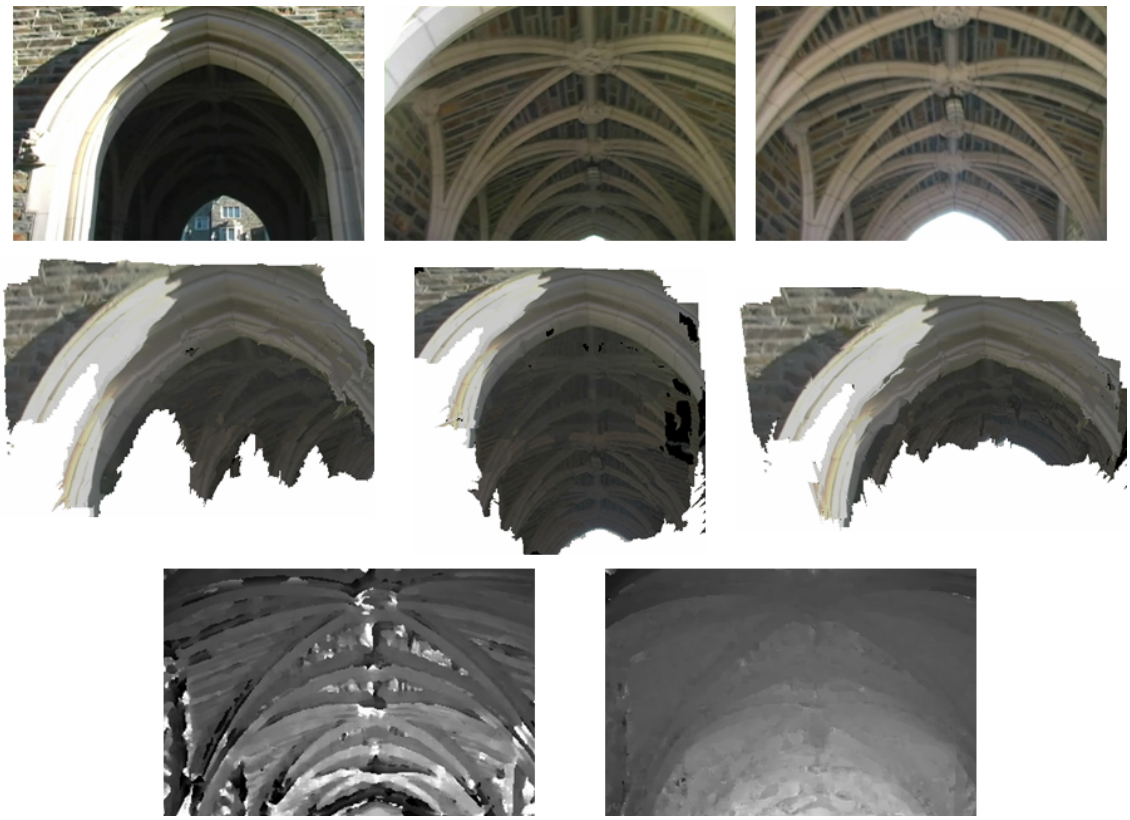
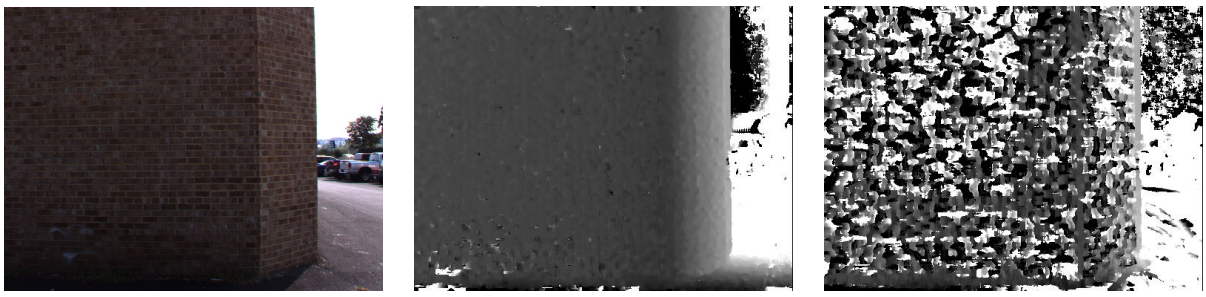


Figure 4.12: Second stereo example. (Top) Sample images from the video sequence (Middle) Generated 3D model with radiometrically aligned textures (Bottom) Depth maps computed without exposure compensation (left) and with our method (right)



(a) (b) (c)

Figure 4.13: Additional Stereo Example. (a) video frame, (b) novel gain corrected stereo, (c) standard stereo.

from motion techniques by incorporating the information for 3D camera tracking, the depth from stereo and providing radiometrically aligned images for texture-mapping.

Chapter 5

Radiometric Calibration with Illumination Change

5.1 Introduction

There are millions of webcams worldwide providing videos of streets, buildings, natural sites such as mountains and beaches, and etc. The images of an outdoor scene collected over time provide a rich source of information and can lead to novel computer vision applications such as computing intrinsic images (Weiss, 2001), building webcam synopsis (Pritch et al., 2007), and geolocating webcams (Jacobs et al., 2007). They are also valuable in studying scene appearance variations, which can help develop more interesting applications and enhance existing computer vision methods that were constrained to controlled indoor environments. For this purpose, Narasimhan et al. (2002) introduced a database of images of a fixed outdoor scene with various weather conditions captured every hour for over 5 months. Their database covers a wide range of illumination conditions (both day and night), weather conditions, and seasons. Another database of images were introduced by Jacobs et al. (2006) where they collected more than 17 million images over 6 months from more than 500 webcams across the United States. In their work, it was shown that the image sets from different static cameras have consistent correlations over large spatial and temporal extents.

The scene appearance depends on multiple factors including the scene geometry and reflectance, illumination geometry and spectrum, and the viewing geometry. In addition, the weather has a large effect on the scene appearance for outdoor scenes. An important factor for determining the image appearance of a scene that is often not considered is the radiometric properties of the camera. In many computer vision systems, an image of a scene is assumed to directly reflect the appearance of the scene. However, this is not the case for most cameras as the camera response function is nonlinear. In addition, cameras usually operate in the auto-exposure mode where the exposure settings are automatically adjusted according to the dynamic range of the scene which may change the appearance of the scene in the images. Note also that this is often a necessity for outdoor scenes undergoing significant lighting variation during the day. The effect of the auto-exposure on the images is illustrated in Figure 5.1 where pixel values of a point over time recorded with auto-exposure are compared with those recorded with a fixed exposure value. In this particular example, the sun is moving away from the scene so the radiance of the points in the scene are decreasing as shown by the pixel values of the fixed exposure sequence. In the auto-exposure mode however, the camera compensates for the decrease in the overall brightness of the scene resulting in the increase of the pixel values. As can be seen in the auto-exposure sequence example, the pixel values stay almost constant even though the scene radiance is actually decreasing. While this behavior is good for the viewing purposes, it has an ill effect on many computer vision methods that rely on the scene radiance measurement such as photometric stereo, color constancy, and on the methods that use image sequences or time-lapse data of a long period of time such as in Jacobs et al. (2006), Jacobs et al. (2007), and Weiss (2001) since the pixel values do not reflect the actual scene radiance.

In this chapter, we introduce a new algorithm to compute the radiometric response function of the camera and the exposure values of images given a sequence of images of a static outdoor scene taken at a regular interval for a period of time. While the

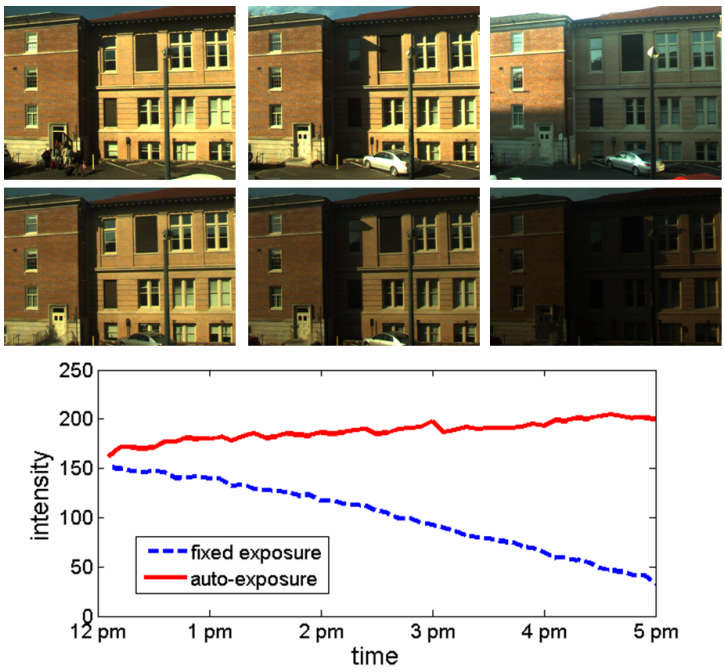


Figure 5.1: Effect of auto-exposure. (Top) Images taken at different times with auto-exposure (Middle) Images taken with exposure fixed (Bottom) Pixel values of a point over time.

underlying assumption for our method is that the surfaces are Lambertian, the proposed method deals with non-Lambertian surfaces such as windows and specular materials by automatically filtering out those points. Radiometric calibration on this type of data is a challenging problem because the illumination for each image is changing, causing the exposure of the camera to change. Most of the previous radiometric calibration methods cannot be applied because they are based on using differently exposed images taken with constant illumination. In particular, exposures will only change in response to lighting changes which makes it hard to separate the effect of both. We solve the problem of lighting change by first selecting groups of pixels that have constant behaviors with regard to illumination change. This means that the pixels in a group are either all sunlit or in a shadow at a certain time in addition to having the same surface normal. The effect of the exposure and the lighting is constant for the selected pixels and the intensity differences between these pixels are due to their albedo differences which should

remain constant over time since the albedo is a property of the material. This property is exploited to compute the response function using images with varying illumination. Estimating the exposure value for each image in the sequence after linearizing the images with the computed response function is still a difficult problem because the change in the intensity is due to the change in both the exposure and the illumination. There are countless combinations of the exposure and the illumination change that results in the same intensity change. To solve this problem, we model the illumination variation according to the motion of the sun since we are dealing with outdoor scenes.

As mentioned in Chapter 2, radiometric calibration methods that deal with illumination change were introduced in Manders et al. (2004) and Shafique and Shah (2004). However, both methods are limited in practice and cannot be used for outdoor images. For the method in Manders et al. (2004) which uses superposition constraints imposed by different combinations of two (or more) lights, the control over lighting is necessary as you need images with different light sources and an image with all the light sources on. The method in Shafique and Shah (2004) estimates the response function by exploiting the fact that the material properties of the scene should remain constant and use cross-ratios of image values of different color channels to compute the response function. The response function for this method is limited to a gamma function making this method difficult to use in practice. Compared to these works, the algorithm proposed in this chapter is more general in that the we use natural lighting conditions and allow exposure changes. In addition, we allow a more general model of the response function, do not require information across different color channels, and compute the response function linearly unlike the method of Shafique and Shah (2004).

The remainder of the chapter is organized as follows. In Section 5.2, we introduce a method for computing the camera response function using images with illumination change. Then we develop methods for computing the exposure value for each image in a sequence in Section 5.3. We evaluate our methods with experiments in Section 5.4 and

conclude with discussion about our algorithm in Section 5.5.

This work was presented in Kim et al. (2008b).

5.2 Computing the Radiometric Response Function with Illumination Change

In this section, we first introduce a method for computing the response function of a camera given multiple images of a static scene which is assumed to consist predominantly of Lambertian surfaces with illumination change. We model the image formation process as

$$I_{it} = f(k_t a_i M_{it}) \ , \tag{5.1}$$

where the response function f transforms the product of the exposure value k, the illumination M, and the albedo a to the image intensity I. The indexes i and t denote pixel location and time respectively. The product of the albedo (a) and the illumination (M) is the irradiance (E) in the image formation equations presented in previous chapters (Equation 2.3 and Equation 3.2). The illumination M is the inner product between the surface normal $\mathbf{N}$ and the directional light $\mathbf{L}$ which in our case is the sunlight. The illumination also includes indirect lighting[1] ($L_{indirect}$) such as from sky illumination and reflections from different surfaces.

$$M_{it} = \mathbf{N}_i \cdot \mathbf{L}_t + L_{indirect} \tag{5.2}$$

Equation (5.1) can also be written as follows :

$$f^{-1}(I_{it}) = k_t a_i M_{it} \ . \tag{5.3}$$

[1]This lighting term is referred to as ambient lighting in computer graphics

Let $g = logf^{-1}$, $K = log(k)$, and $\alpha = log(a)$, then

$$g(I_{it}) = K_t + \alpha_i + log(M_{it}) \ . \tag{5.4}$$

If two points in the image have the same surface normal and both points are either both in a shadow or a non-shadow region, the amount of lighting is the same for the two points ($M_{it} = M_{jt}$)[2]. The exposure K is a global factor for all points in an image so the relationship between the two points can be stated as follows :

$$g(I_{jt}) - g(I_{it}) = \alpha_j - \alpha_i \ . \tag{5.5}$$

By using the points with same lighting conditions, the relationships between the image intensities of the points are explained only with the albedos of the points. Since the albedo of a point is constant over time, we can use Equation (5.5) to compute the response function g as well as the albedo differences between points with same lighting condition ($\alpha_j - \alpha_i$) from multiple images with different illumination.

5.2.1 Finding Pixels with Same Lighting Conditions

The first step necessary to compute the camera response function is to find pixels that have the same lighting conditions in all images that are used for the radiometric calibration. For different pixels to have the same lighting conditions, the surface normals of the points have to be the same and if one point is in the shadows, the other points also have to be in the shadows at that time. We modify the method proposed in Koppal and Narasimhan (2006) in which the appearance of the scene is clustered according to the surface normals. The key observation is that appearance profiles for iso-normal points exhibit similar behaviors over time (Figure 5.2). An appearance profile is a vector of

[2]We assume that the indirect lighting is the same for all points within a patch at a specific time.

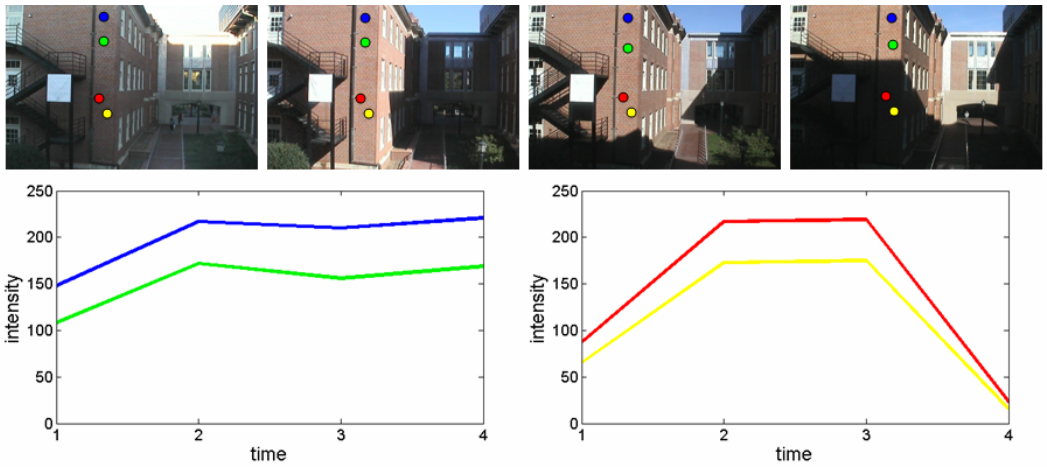


Figure 5.2: Using appearance profile to cluster pixels with same lighting conditions : (Top) Images used to compute the camera response function (Bottom) Appearance profiles of points with the same lighting conditions. Note that even though all the points have the same normal in the example, they have different profiles due to shadows.

measured intensities at a pixel over time and the extrema location in the profiles are used to cluster the appearance in Koppal and Narasimhan (2006). In this work, we compute the similarity of the lighting between two pixels by simply computing the normalized correlation between the appearance profiles of the two points. With this similarity measure, we use the k-means algorithm to cluster pixels with same lighting conditions over time (Figure 5.3). The clusters at this point may contain errors due to non-Lambertian regions, motions in the scene, and reflections. To deal with these errors, we first divide the image into blocks of the same size and filter out regions where all the pixels do not fall into the same cluster as illustrated in Figure 5.3. Blocks with uniform intensity such as in sky are also filtered out since they don't provide valuable information for the radiometric calibration.

5.2.2 Pixel Selection

After clustering the pixels, we then select pixels from each cluster for the response function estimation. First, we randomly pick a number of points (300 points in our

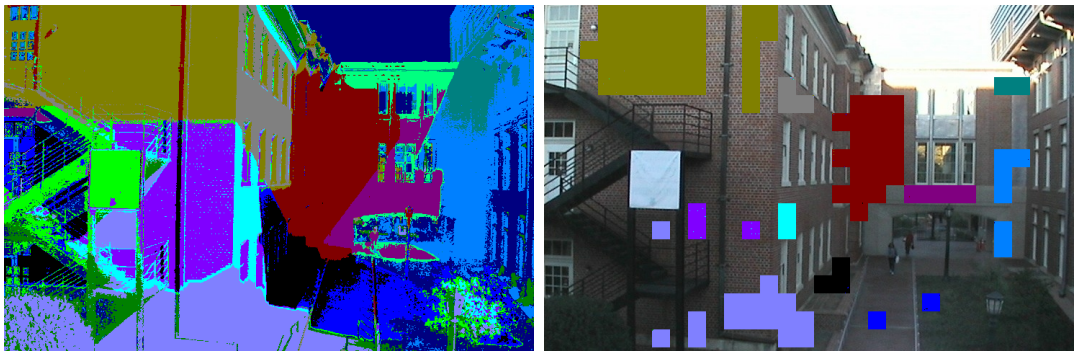


Figure 5.3: (Left) Pixels clustered by the appearance profile with k-means algorithm using 4 images shown in Figure 5.2. Clusters are identified with different colors in the image. (Right) Regions (blocks) with non-uniform clusters are filtered out. Most of the non-Lambertian regions are filtered out at this stage.

experiments) from each cluster. Due to image noise and non-uniform surface, the appearance profiles for the selected pixels will be significantly disturbed by noise as shown in Figure 5.4. Profiles of two pixels under the same lighting conditions crossing each other means that the albedo difference between the two points changed even though it should stay constant throughout. It is essential to filter out these outliers which can otherwise have a serious effect on the estimation results.

To remove outliers from the selected pixels for each cluster, we use the order of the pixel intensities as the cue. The idea is that if a pixel has the lowest intensity in one frame, the intensity of that pixel should also be the lowest in the following frames. Assuming that there are n points selected for a cluster, we build a vector d_{it} of size n for each pixel i at time t where each element is :

$$\mathbf{d}_{it}(j) = \begin{cases} +1 & \text{if } I_{it} > I_{jt} \\ -1 & \text{if } I_{it} < I_{jt} \\ 0 & \text{if } I_{it} = I_{jt} \end{cases} \tag{5.6}$$

The dot product between d_{it} and d_{it+1} gives us how much support the pixel i has in terms of orders from other pixels in the cluster. We iteratively remove pixels with

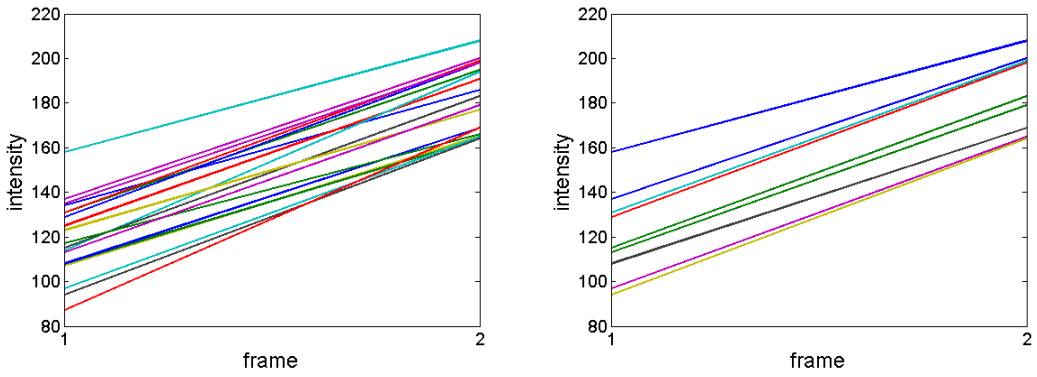


Figure 5.4: Pixel profiles for two frames (Left) Originally selected pixels and their profiles (Right) Profiles after postprocessing.

the worst support until all the pixels are in order between frames. An example of this process is shown in Figure 5.4.

5.2.3 Radiometric Response Function Estimation

To model the response function g, we again use the Empirical Model of Response (EMoR) explained in Section 2.3.

$$g(I) = \mathbf{g}_0(I) + \sum_{s=1}^{m} c_s \mathbf{h}_s(I) \tag{5.7}$$

where the g_0 is the mean function and the c_k's are the coefficients for the basis functions $\mathbf{h}_k$'s. Combining Equation (5.5) and Equation (5.7), we have

$$\sum_{s=1}^{m} c_s(\mathbf{h}_s(I_{jt}) - \mathbf{h}_s(I_{it})) - \alpha_{ji} = g_0(I_{it}) - g_0(I_{jt}) \tag{5.8}$$

where $\alpha_{ji} = \alpha_j - \alpha_i$.

For n pixels in the same cluster l at time t, we have $n-1$ linear equations $\mathbf{A}_{tl}\mathbf{x_t} = \mathbf{b}_{tl}$ as follows.

$$\mathbf{A}_{tl} = [\mathbf{A'}_{tl} \quad \mathbf{I}_{n-1}] \tag{5.9}$$

$$\mathbf{A'}_{tl}(y,x) = \mathbf{h}_x(I_{y+1,t}) - \mathbf{h}_x(I_{1t}), \quad {\scriptstyle 1\le y\le n-1,\ 1\le x\le m,} \tag{5.10}$$

$$\mathbf{b}_{tl}(y) = g_0(I_{1t}) - g_0(I_{y+1,t}), \quad 1 \le y \le n-1 \tag{5.11}$$

$$\mathbf{x_t} = [\mathbf{c}, \mathbf{a}_l]^T \tag{5.12}$$

where $\mathbf{I}_{n-1}$ is an identity matrix of size $n-1$ by $n-1$, $\mathbf{c} = [c_1, c_2, \ldots, c_m]$ and $\mathbf{a}_l = [\alpha_{21}, \alpha_{31}, \ldots, \alpha_{n1}]$.

Since we have m+n-1 unknowns with n-1 equations, the system above is under-constrained. We can add more equations to the system by incorporating the temporal information of multiple frames. The number of points n is typically bigger than the number of basis functions ($m = 5$ in this chapter), so as few as two frames are enough to solve for the response function. Since one cluster typically does not provide enough range of intensities for accurate estimation, we combine equations from different clusters. Adding multiple clusters at multiple frames, we can compute the response function by solving the following least squares problem $\mathbf{Ax} = \mathbf{b}$ with (assuming we are using 3 clusters from 2 frames for simplicity)

$$\mathbf{A} = \begin{pmatrix} \mathbf{A'}_{11} & \mathbf{I}_{n-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{A'}_{21} & \mathbf{I}_{n-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{A'}_{12} & \mathbf{0} & \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{A'}_{22} & \mathbf{0} & \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{A'}_{13} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n-1} \\ \mathbf{A'}_{23} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n-1} \end{pmatrix}, \tag{5.13}$$

$$\mathbf{b} = [\mathbf{b}_{11}, \mathbf{b}_{21}, \mathbf{b}_{12}, \mathbf{b}_{22}, \mathbf{b}_{13}, \mathbf{b}_{23}]^T \ , \tag{5.14}$$

$$\mathbf{x} = [\mathbf{c}, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]^T \ . \tag{5.15}$$

In practice, the rows of $\mathbf{A}$ and $\mathbf{b}$ are weighted according to the intensity of the pixel used for the row as explained in Chapter 3 (Equation 3.18).

The solution to the problem $\mathbf{Ax} = \mathbf{b}$ above suffers from the exponential (or scale in the log space) ambiguity explained in the earlier chapters. To fix the scale of the response function, we set the value of the response function at the image value 128 to a value τ. We will discuss this ambiguity later in Section 5.3.

5.3 Exposure Estimation from Images with Different Illumination

By using the computed response function, we can linearize the images as in Equations (5.3). While the images taken at different times are now linearly related, the images may not reflect the true appearance of the scene due to the exposure change in the camera. However, there is an inherent ambiguity in computing the exposures from images with different illumination similar to the exponential ambiguity mentioned in the previous section. As can be seen from Equation (5.3), there is an infinite number of combinations of the exposure and the lighting that result in the same image intensity. To compute the exposure, assumptions on the lighting have to be made.

In this section, we introduce a method to estimate the exposure values given a sequence of images of an outdoor scene taken over a period of time. For the outdoor scenes, the dominant source of lighting in general is the sun. We model the lighting change according to the motion of the sun and use the model to compute the exposures. We make the assumption that the sunlight was not blocked by clouds when the images were taken.

5.3.1 Modeling the Illumination with the Motion of the Sun

The direction of the sunlight ($\mathbf{L}_t$) at time t and the surface normal of a point i ($\mathbf{N}_i$) can be expressed in Cartesian coordinates as in the following equation where θ's are the

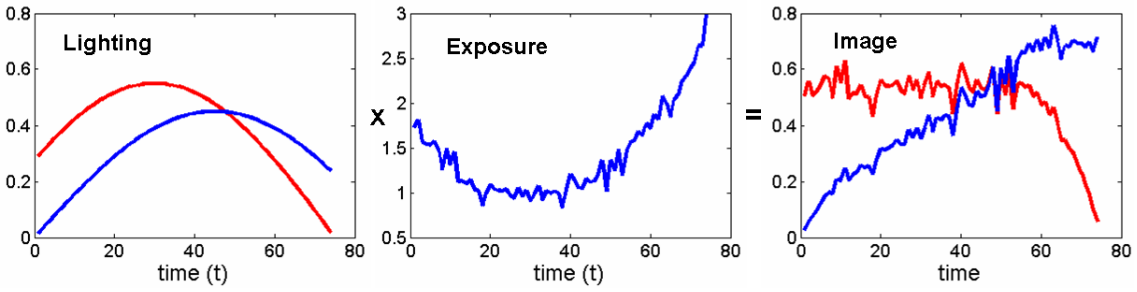


Figure 5.5: Relationship between lighting, exposures, and image appearance. We need at least two pixel profiles to compute exposures since many combination of lighting and exposure can result in a same profile.

azimuth angles and ϕ's are the elevation angles :

$$\begin{aligned} \mathbf{L}_t &= [\cos\phi_t \cos\theta_t, \cos\phi_t \sin\theta_t, \sin\phi_t]^T \ , \\ \mathbf{N}_i &= [\cos\phi_i \cos\theta_i, \cos\phi_i \sin\theta_i, \sin\phi_i]^T \ . \end{aligned} \tag{5.16}$$

The lighting due to the sun at point i is then

$$\mathbf{N}_i \cdot \mathbf{L}_t = \cos\phi_t \cos\phi_i \cos(\theta_t - \theta_i) + \sin\phi_t \sin\phi_i \ . \tag{5.17}$$

Without loss of generality we rotate $\mathbf{L}_t$ and $\mathbf{N}_i$ so that $\phi_t = 0$, Equation (5.17) becomes

$$\begin{aligned} \mathbf{N}_i \cdot \mathbf{L}_t &= \cos\phi'_i \cos(\theta_t - \theta'_i) \\ &= \cos\phi'_i (\cos\theta'_i \cos\theta_t + \sin\theta'_i \sin\theta_t) \\ &= p_i \cos\theta_t + q_i \sin\theta_t \end{aligned} \tag{5.18}$$

where $p_i = \cos\phi'_i \cos\theta'_i$ and $q_i = \cos\phi'_i \sin\theta'_i$. According to Equation (5.18), the lighting variation at a point due to the sun over time is a sinusoidal function with the scale and the phase being the parameter.

5.3.2 Exposure Estimation

Combining Equations (5.2), (5.3), and (5.18) we have

$$\frac{1}{k_t} f^{-1}(I_{it}) - p'_i \cos\theta_t - q'_i sin\theta_t = 0. \tag{5.19}$$

In the above equation, p'_i and q'_i are considered to include the albedo term a from Equation (5.3). Additionally, we assumed that the effect of indirect lighting is constant over time. With a sequence of η images, the exposure for each image (k_t, $1 \leq t \leq \eta$) and the lighting parameters (p'_i and q'_i) for a point i can be computed as follows :

$$\begin{pmatrix} \cos\theta_1 & \sin\theta_1 & f^{-1}(I_{i1}) & 0 & 0 & \cdots & 0 \\ \cos\theta_2 & \sin\theta_2 & 0 & f^{-1}(I_{i2}) & 0 & \cdots & 0 \\ & & & \vdots & & & \\ \cos\theta_\eta & \sin\theta_\eta & 0 & 0 & 0 & \cdots & f^{-1}(I_{i\eta}) \end{pmatrix} \begin{pmatrix} p'_i \\ q'_i \\ k'_1 \\ k'_2 \\ \vdots \\ k'_\eta \end{pmatrix} = \mathbf{0} \tag{5.20}$$

where $k' = 1/k$.

As can be seen, Equation (5.20) is under-constrained with $\eta + 2$ unknowns and η equations. Therefore, at least two pixel profiles of different surface normals are necessary as shown in Figure 5.5.

Let

$$\mathbf{S} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ \cos\theta_2 & \sin\theta_2 \\ \vdots & \vdots \\ \cos\theta_\eta & \sin\theta_\eta \end{pmatrix}, \tag{5.21}$$

$$\mathbf{F}_i = \begin{pmatrix} f^{-1}(I_{i1}) & 0 & \cdots & 0 \\ 0 & f^{-1}(I_{i2}) & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & f^{-1}(I_{i\eta}) \end{pmatrix} \tag{5.22}$$

with $\theta_t = 2\pi t(\frac{1}{24}\frac{s_{min}}{60})$ (s_{min} is the sampling interval in minutes).

Using a total of ζ pixels with different surface normals, we can solve for the exposures k_t ($1 \leq t \leq \eta$) and the lighting parameters (p'_i and q'_i, $1 \leq i \leq \zeta$) by solving a linear least squares problem $\mathbf{U}\mathbf{y} = \mathbf{0}$ with

$$\mathbf{U} = \begin{pmatrix} \mathbf{S} & \mathbf{0}_{\eta\times 2} & \mathbf{0}_{\eta\times 2} & \dots & \mathbf{0}_{\eta\times 2} & \mathbf{F}_1 \\ \mathbf{0}_{\eta\times 2} & \mathbf{S} & \mathbf{0}_{\eta\times 2} & \dots & \mathbf{0}_{\eta\times 2} & \mathbf{F}_2 \\ & & \ddots & & & \\ \mathbf{0}_{\eta\times 2} & \mathbf{0}_{\eta\times 2} & \mathbf{0}_{\eta\times 2} & \dots & \mathbf{S} & \mathbf{F}_\zeta \end{pmatrix}, \tag{5.23}$$

$$\mathbf{y} = [p'_1, q'_1, p'_2, q'_2, \cdots, p'_\zeta, q'_\zeta, k'_1, k'_2, \cdots, k'_\eta]^T. \tag{5.24}$$

More pixels can be added to the problem easily by putting the sub-matrices $\mathbf{S}$ and $\mathbf{0}_{\eta\times 2}$ in $\mathbf{U}$ in the same order as the pixels with the same surface normal.

A set of pixels used to solve the equation above are randomly selected from the clusters used for the response function estimation (Section 5.2). It is important not to use pixel values at time t in the above equation when the pixels fall into shadows since the lighting model does not apply to shadow regions. From the appearance profile of a pixel, we detect whether the pixel is in shadow by a simple thresholding as in Sunkavalli et al. (2006) (Figure 5.6). We also remove the pixels from the equation if the average intensity is too low meaning that the pixels were probably always in the shadow.

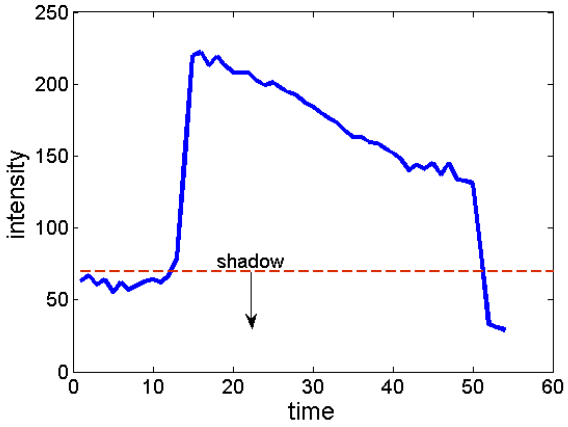


Figure 5.6: Appearance profile of a pixel : a simple thresholding scheme is used to detect whether a pixel is in shadow or not. See Sunkavalli et al. (2006) for the thresholding scheme.

5.3.3 Exponential Ambiguity

In Section 5.2, we discussed the inherent ambiguity in computing the response function where the elements in Equation (5.3) are related exponentially as follows.

$$(f^{-1}(I_{it}))^{\gamma} = k_t^{\gamma} a_i^{\gamma} M_{it}^{\gamma} \tag{5.25}$$

We resolved this exponential ambiguity by arbitrarily fixing the scale of the response function which is not a problem for applications that require image intensity alignment since different γ's still result in the same intensity value. However the ambiguity affects our exposure estimation process since our method is based on having the right scale for the response function f. If the scale of the response function is incorrect, then the system is trying to fit a sine function to a measurement that is the exponent of a sine. Ideally, the error $\|\mathbf{Uy}\|$ in Section 5.3.2 gives us the information about the γ. It should be the minimum when the correct scale of the response function is used. However, the error is not distinctive due to image noise and lack of time interval when surfaces of different normals are both in the sunlight as shown in Figure 5.7. Alternatively, we need

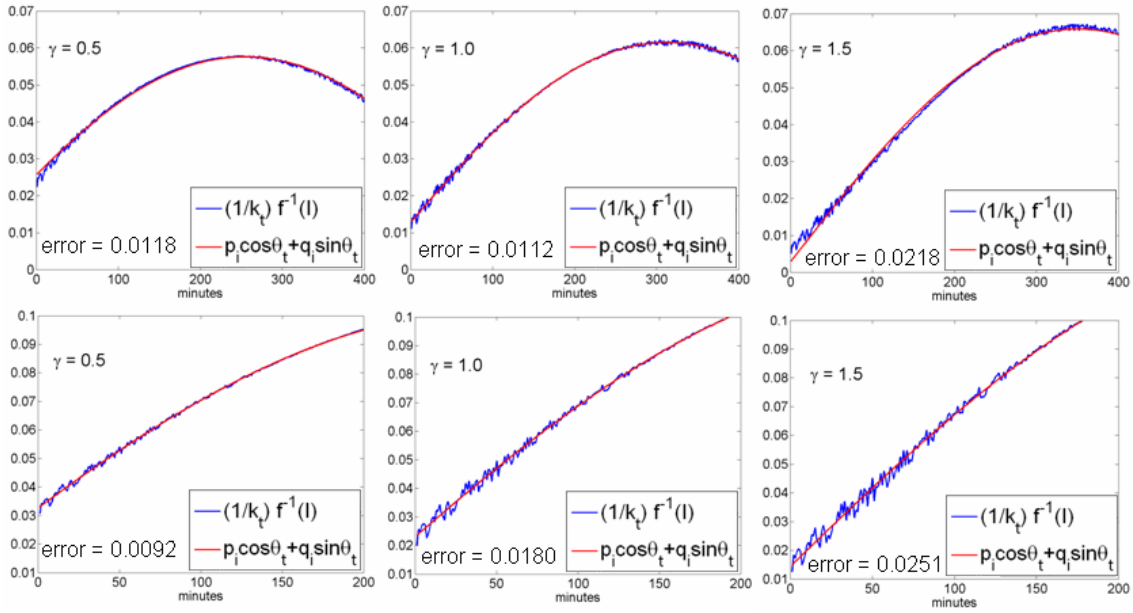


Figure 5.7: Simulation of the effect of the exponential ambiguity. The exposures and the lighting changes were estimated on the two synthetically generated image profiles similar to Figure 5.5 using 400 minutes of data (top) and 200 minutes of data (bottom). The correct γ is 1.0.

information about the camera or the scene to find the right scale. In this work, we first estimate the exposures (k_t) and the lighting functions ($p'_i \cos\theta_t + q'_i sin\theta_t$ in Equation 5.19) using multiple γ values. The recovered lighting functions will have different phases with different γ's as shown in Figure 5.7. We then manually select the γ value that yields the lighting functions to have the peaks at the right time of the day which can be inferred from the orientations of shadows in the image sequence. We plan to add a step for computing the γ automatically in the future.

5.4 Experiments

We first evaluate our response function estimation method introduced in Section 5.2. Two cameras used for the experiments are Sony SNC-RZ30N PTZ camera and Point Grey Dragonfly camera. For the Sony camera, we first computed the response function

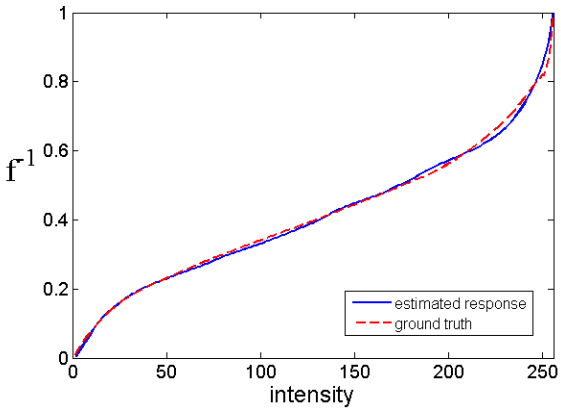


Figure 5.8: Response function estimation result for Sony SNC-RZ30N PTZ camera. Images used are shown in Figure 5.2

by using the method introduced in Chapter 3 with multiple images of a static scene with constant illumination to test our method. We then computed the response function with our method using four images shown in Figure 5.2 and the comparison of the computed response functions is shown in Figure. 5.8. While only the green channel was used for this example, we can easily combine all channels if necessary. For the Point Grey Dragonfly camera, we compare our result computed with two images with the known linear response function (Figure 5.9). The number of images for accurate estimation depends on the intensity range of each image. While the method does not need a large number of points, it is important to have well distributed pixel intensities for accurate estimation.

To evaluate our exposure estimation method, we recorded images of a scene with the Point Grey camera every minute for a little more than 4 hours when we could observe surfaces with different normals being illuminated by the sun. Some sample images as well as some of the pixel profiles used for the estimation are shown in Figure 5.10. Our exposure estimates are compared to the ground truth exposures reported by the camera in Figure 5.11. Notice that the exposure estimates start to deviate from the ground truth starting around hour 1400. The cause for this is the change in indirect lighting as

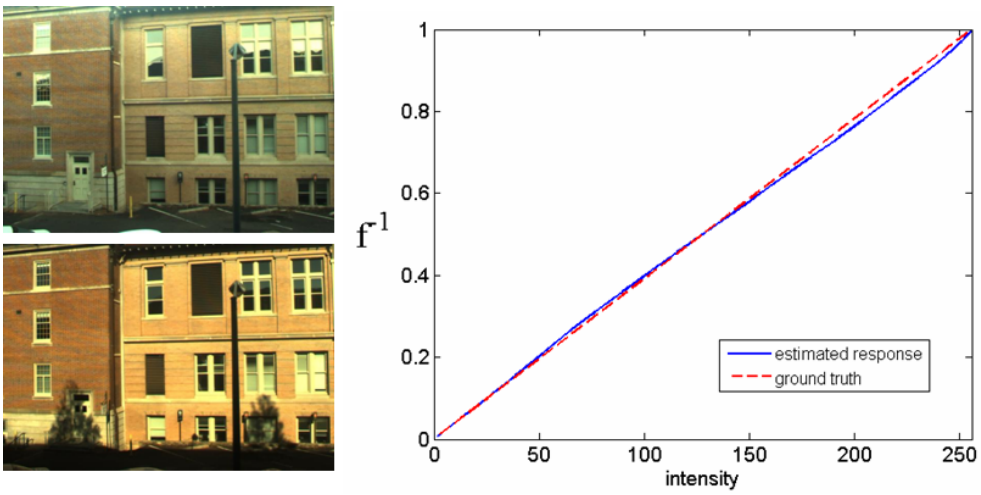


Figure 5.9: Response function estimation result for Point Grey Dragonfly camera with two images used for our estimation.

a building in front of the scene started to cast a large shadow at that time reducing the amount of light in the scene. Since our method is based on constant indirect lighting, the change in the indirect lighting caused errors in the exposure estimation. However, for a long period of time when the indirect lighting was close to being constant, our estimation was accurate as shown in the figure. We can observe the function of the auto-exposure from Figure 5.10. The camera adjusts to the brightness change by trying to fix the intensity of dominant pixels. This function prohibits images from being under-exposed or saturated as can be seen from the exposure-compensated images in the figure. While this is good for viewing, this could affect vision algorithms that rely on photometric measurements since the image intensities do not reflect the true radiance of the points. By computing the response function and exposures using our method, we can convert the image intensities to their actual radiance enabling further analysis of the scene.

As our last experiment, we used one of the webcam datasets introduced in Jacobs et al. (2006) as shown in Figure 5.13. The images we used were captured every 30 minutes for 11 hours. The estimated response function and the exposures are shown in Figure 5.12. Note that we do not have the ground truth for this data since the camera is

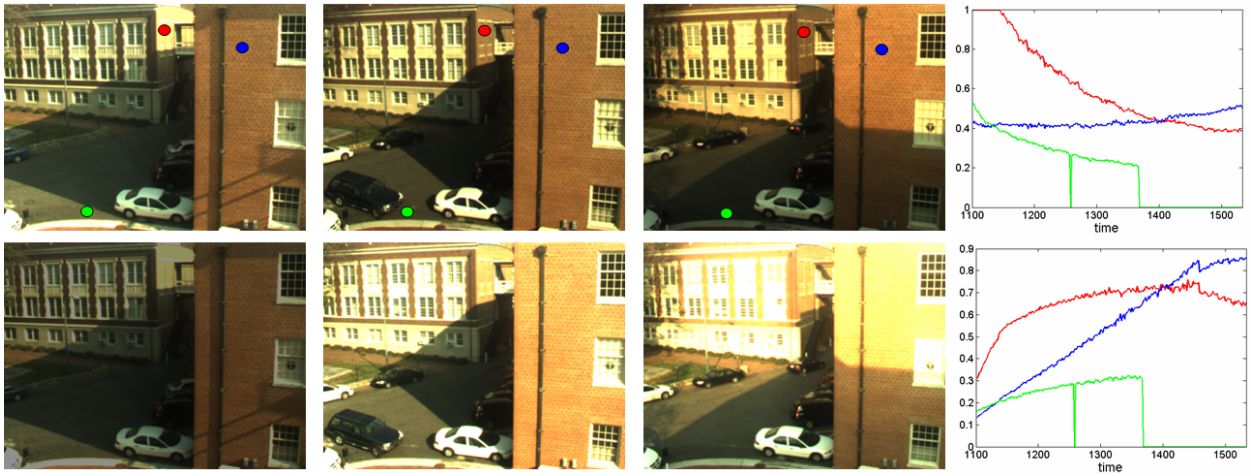


Figure 5.10: Exposure Estimation. (Top) Sample images from the input sequence and the pixel profiles of the dotted points (Bottom) Images and profiles normalized to a fixed exposure. The 0 values in the profiles represent shadow.

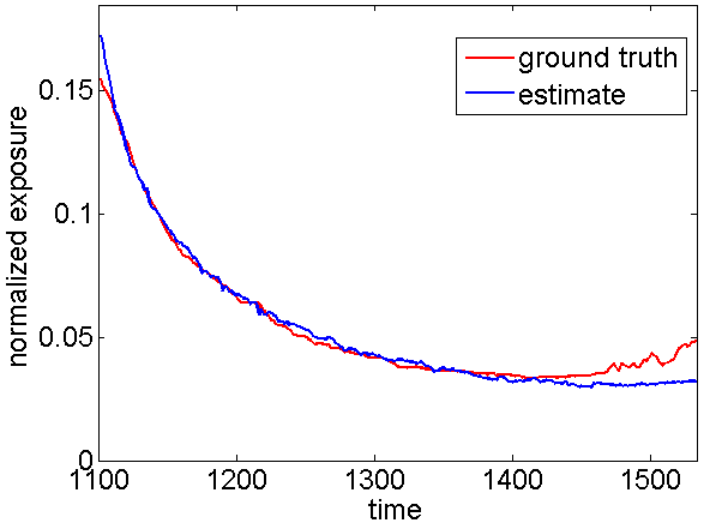


Figure 5.11: Comparison of the estimation with the ground truth exposure

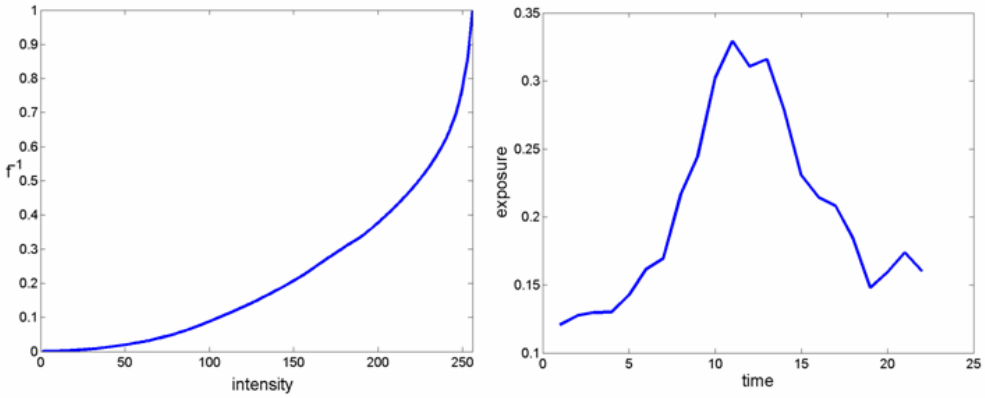


Figure 5.12: Estimated response function (left) and the exposures (right) using dataset introduced in Jacobs et al. (2006) (Figure 5.13)

unknown. We can roughly evaluate the results by comparing the input images and the pixel profiles with the images and the profiles normalized with the estimated exposures as in Figure 5.13. Input profiles tend to stay constant unless affected by shadows. However, after normalizing the images with the estimated exposures, the pixel values vary gradually as expected.

5.5 Conclusion

We have introduced a novel method for computing the radiometric response function of a camera and additionally demonstrated the computation of exposures for outdoor image sequences. This is a challenging problem because the image appearance varies due to the changes in both the exposure of the camera and the lighting conditions. For computing the camera response function, we solved the problem of illumination change by using groups of pixels with a constant behavior towards lighting change where the material property between the pixels in the group remains the same. To compute the exposures, some prior knowledge in either the exposure or the lighting is necessary since different combinations of exposure and lighting result in the same image intensity. We overcome this limitation by modeling the lighting according to the motion of the sun

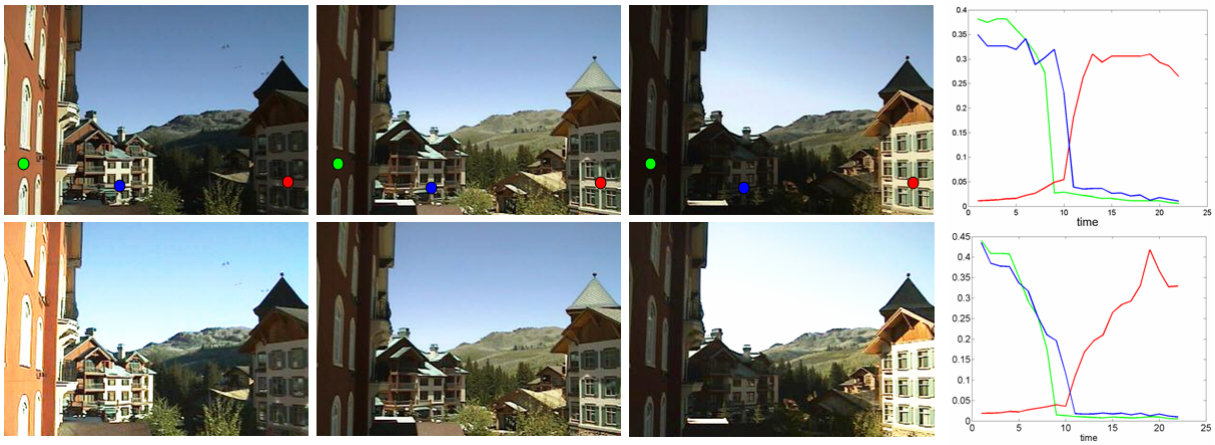


Figure 5.13: (Top) Sample images from one of the dataset introduced in Jacobs et al. (2006) and the pixel profiles of the dotted points. (Bottom) Images and profiles normalized to a fixed exposure. The right side of the figure is to the east.

which turns out to be a sinusoidal function.

Most conventional radiometric calibration approaches use images with different exposures while the illumination stays constant. Hence those algorithms cannot be used in time-lapse sequences as used in this chapter. A couple of methods were presented to deal with the lighting change (Manders et al., 2004; Shafique and Shah, 2004) but both methods are limited to special cases making it difficult to apply to outdoor image sequences. The significance of the method presented in this chapter is that it overcomes the limitations of previous work and can be used for real images of outdoor scenes. We believe that this work can serve as a basis for more exciting outdoor scene analysis applications in computer vision.

Chapter 6

Conclusion

6.1 Summary

In preceding chapters, several algorithms were presented to support the claims made in the thesis statement :

> Given a collection of images of a scene taken under varying conditions, one can compute radiometric properties of the camera (up to some ambiguities) that explain the relationship between the image brightness and the scene radiance as well as the radiometric relationship between multiple images. The radiometric properties include the radiometric response function, exposures, and vignetting.

In Chapter 3, we first introduced an algorithm for computing the radiometric response function, exposures, and vignetting from a set of images taken with freely moving camera. This method advances the state of the art in radiometric calibration by allowing general image sets to be used, while previous methods were limited to images taken with a static camera or a rotating camera. By decoupling the vignetting effect from the radiometric response function, we were able to approach the camera response estimation problem and the vignetting computation problem independently and in a robust way. We evaluated our method with synthetic and real examples. With the

proposed algorithm, we were able to estimate the radiometric properties of the camera accurately even in the presence of large noise and mismatches, whereas other existing methods were not effective against noise and outliers. We applied our method to radiometrically align images for seamless mosaics and 3D model textures. Additionally, we applied the algorithm to create a high dynamic range mosaic that represents the scene radiance better than the conventional image mosaic.

The method presented in Chapter 3 requires correspondences between images, which must be obtained from stereo matching or from a known projective transform between images. In Chapter 4, we tackle the problem of computing the correspondences and the radiometric calibration simultaneously. This method is especially suited for video sequences taken in a high dynamic range environment where the image brightness of a point changes due to the changes in the exposure of the camera. Our method advances the conventional feature tracking algorithm (KLT tracker) which requires the brightness of features to stay constant by unifying the problems of feature tracking and radiometric calibration into a common framework. Our computationally efficient algorithm takes advantage of the structure of the estimation problem which leads to a minimal computational overhead. With our joint estimation, we were able to advance the quality and robustness of the known structure from motion techniques by incorporating the information for 3D camera tracking, the depth from stereo, and radiometric alignment of images for texture-mapping.

The methods in Chapter 3 and Chapter 4 as well as most of the previous work on radiometric calibration are based on the fact that the scene radiance itself is the same for all images in the sequence and the change in the image brightness is due to the change in the exposure only. This means that the illumination condition for all the images is the same. While this condition is valid for many applications where the images are taken in a short period of time, the condition does not hold for images taken over a long period time such as a time-lapse video of an outdoor scene taken over a day. In Chapter 5, we

have introduced a novel radiometric calibration algorithm that solves this problem of illumination change by using groups of pixels with constant behavior towards lighting change where the relationship between the albedo of the pixels in the group remains the same. In addition, we have also presented a method to compute the exposure values by modeling the effect of lighting on images by a sinusoidal function which is computed according to the motion of the sun. The images in the sequence can be normalized using the computed response function and exposures enabling the image sequence to be used for computer vision methods that are based on image variation due to the change in illumination only.

6.2 Future Work

There are various areas that we would like to explore in the future to overcome the limitations as well as to extend our methods.

- **Exponential Ambiguity :** As mentioned several times, the radiometric response function recovered using the algorithms presented in this dissertation is subject to the exponential ambiguity. As stated in Grossberg and Nayar (2003), it is impossible to recover both the response function and the exposures simultaneously without making assumptions on either the response function or the exposures. To resolve this ambiguity problem, we either made an assumption on the exposure by fixing the exposure ratio between an image pair to an arbitrary value (Chapter 3) or on the response function by fixing the the response function to an arbitrary scale (Chapters 4 and 5). This arbitrary fixing of the scale does not affect applications such as radiometric alignment and feature tracking since the change of brightness between images is the same regardless of the scale of the response function and exposures. However, the ambiguity poses problem for applications that require accurate scene radiance measurements such as high dynamic range imaging and

decoupling exposure change from lighting as in Chapter 5. For the future, we would like to extend our methods to find the right scale of the response function automatically. One possibility would be using high-order correlations in the frequency domain such as in Farid (2001). Farid modeled the nonlinearity between the scene radiance and the image intensity with a gamma function ($f(E) = E^{\gamma}$) and estimated the gamma by computing the bispectrum. We could approach the problem of resolving the exponential ambiguity similar to Farid's method since the problem becomes estimating the gamma with the exponential ambiguity.

- **Vignetting :** Vignetting is modeled as circularly symmetric about the image center in Chapter 3. We would like to add more flexibility to our vignetting computation to include nonsymmetric vignetting models as in Litvinov and Schechner (2005b) and also solve the cases where vignetting would not be the same for all images in the sequence. In addition, we are interested in adding vignetting estimation to the method presented in Chapter 4. Since the displacement between frames is small, the vignetting did not affect the joint feature tracking and radiometric calibration framework in Chapter 4. However, feature tracks over multiple frames could be used for computing the vignetting effect.

- **Illumination Change :** While the method presented in Chapter 5 overcame the limitation of conventional methods on images with changes in illumination, we would like to improve several aspects of the proposed algorithm. For the camera response estimation, we would like to improve its robustness against errors in the classification of pixels with same illumination conditions. We would also like to add a step where a set of images optimal for the calibration is automatically selected. Additionally, we plan to enhance the exposure estimation process to take into account the change in indirect lighting as well as the change in lighting due to weather. One of the necessary improvement will be in improving the illumination

model of the sun and the sky.

- **Color and Spectral Analysis :** In this dissertation, color changes were described by the exposure changes in each channel independently. In the future, we would like to extend our method to allow for cross talk between the channels to deal with the correlation between color channels (Agathos and Fisher, 2003; Sunkavalli et al., 2008). Also in this work, spectral analysis including the illumination spectrum, the spectral reflectance of the surface, and the spectral response of the camera were not explicitly dealt with. We plan to extend our methods to include the spectral analysis in the future. This would allow us to extend our radiometric calibration schemes to cases where different types of cameras are used. No current radiometric calibration method can be applied when the camera spectral responses are different. By including the spectral analysis, radiometric relationship between multiple cameras of different type can be computed which would be beneficial for many computer vision applications such as color constancy, commodity photo-collection applications (Snavely et al., 2006; Goesele et al., 2007), and multi-spectral imaging (Park et al., 2007; Schechner and Nayar, 2002).

- **Applications :** There are several applications that we would like to explore in the future. First, we are interested in HDR imaging. In addition to the tone-mapping for displaying HDR images, we are interested in creating HDR video as in Kang et al. (2003) and HDR-textured 3D models. We are also interested in exploring HDR display systems with significant larger dynamic range than the current display systems [1]. We would also like to work on applications based on the method proposed in Chapter 5. Some applications of interest include photometric stereo from time-lapse image sequence of an outdoor scene, tensor representation for outdoor scene images (Vasilescu and Terzopoulos, 2002), and texture alignment and

[1] Brightside DR37-P, http://www.dolby.com/promo/hdr/technology.html

relighting from images taken at different times, as in commodity photo-collections.

Bibliography

Agathos, A. and Fisher, R. (2003). Colour texture fusion of multiple range images. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling*, pages 139–146. 24, 107

Aggarwal, M., Hua, H., and Ahuja, N. (2001). On cosine-fourth and vignetting effects in real lenses. In *Proceedings of IEEE International Conference on Computer Vision*, pages 472–479. 8

Asada, N., Amano, A., and Baba, M. (2001). Photometric calibration of zoom lens systems. In *Proceedings of IEEE International Conference on Pattern Recognition*, pages 186–190. 20

Baker, S., Gross, R., Matthews, I., and Ishikawa, T. (2003). Lucas-kanade 20 years on: A unifying framework: Part 2. Technical Report CMU-RI-TR-03-01, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. 60

Bastuscheck, C. M. (1987). Correction of video camera response using digital techniques. *Journal of Optical Engineering*, 26(12):1257–1262. 20

Beauchesne, E. and Roy, S. (2003). Automatic relighting of overlapping textures of a 3d model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 166–173. 25

Birchfield, S. (1997). Klt: An implementation of the kanade-lucas-tomasi feature tracker. http://www.ces.clemson.edu/ stb/klt/. 59, 71, 73

Brown, M. and Lowe, D. (2003). Recognising panorama. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1218–1225. 24, 27

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of The Franklin Institute*, 310(1):1–26. 14

Burt, P. and Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4):217–236. 24

Candocia, F. M. (2003). Simultaneous homographic and comparametric alignment of multiple exposure-adjusted pictures of the same scene. *IEEE Transactions on Image Processing*, 12(12):1485–1494. 18, 57

Candocia, F. M. and Mandarino, D. (2005). A semiparametric model for accurate camera response function modeling and exposure estimation from comparametric data. *IEEE Transactions on Image Processing*, 14(8):1138–1150. 17

Chen, Y. P. and Mudunuri, B. (1986). An anti-vignetting technique for superwide field of view mosaicked images. *Journal of Imaging Technology*, 12(5):293–295. 20

Debevec, P. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH*, pages 369–378. 16, 34, 40

Farid, H. (2001). Blind inverse gamma correction. *IEEE Transactions on Image Processing*, 10(10):1428–1433. 18, 106

Finlayson, G., Drew, M., and Funt, B. (1994). Color constancy: Generalized diagonal transforms suffice. *Journal of Optical Society of America*, 11(11):3011–3019. 32

Goesele, M., Snavely, N., Curless, B., and H. Hoppe, S. M. S. (2007). Multi-view stereo for community photo collections. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1–8. 107

Goldman, D. and Chen, J. (2005). Vignette and exposure calibration and compensation. In *Proceedings of IEEE International Conference on Computer Vision*, pages 899–906. 21, 26, 28, 36, 37, 42, 43, 44, 45, 46, 47, 48, 49, 54

Grossberg, M. and Nayar, S. (2003). Determining the camera response from images: What is knowable? *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(11):1455–1467. 17, 31, 33, 39, 105

Grossberg, M. and Nayar, S. (2004). Modeling the space of camera response functions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282. 10, 17, 21, 22, 32, 64, 67

Hirschmuller, H. and Scharstein, D. (2007). Evaluation of cost functions for stereo matching. pages 1–8. 57

Horn, B. K. P. (1986). *Robot Vision*. The MIT Press, Cambridge, Mass. 6, 7

Hsu, E., Mertens, T., Paris, S., Avidan, S., and Durand, F. (2008). Light mixture estimation for spatially varying white balance, to appear. *ACM Transaction on Graphics*. 12

Jacobs, N., Roman, N., and Pless, R. (2006). Consistent temporal variations in many outdoor scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. 2, 82, 83, 99, 101, 102

Jacobs, N., Satkin, S., Roman, N., Speyer, R., and Pless, R. (2007). Geolocating static cameras. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1–6. 2, 82, 83

Jia, J. and Tang, C. (2005). Tensor voting for image correction by global and local intensity alignment. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27(1):36–50. 24

Jin, H., Favaro, P., and Soatto, S. (2001). Real-time feature tracking and outlier rejection with changes in illumination. In *Proceedings of IEEE International Conference on Computer Vision*, pages 684–689. 60

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45. 69

Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2003). High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325. 107

Kang, S. B. and Weiss, R. (2000). Can we calibrate a camera using an image of a flat, textureless lambertian surface? In *Proceedings of the European Conference on Computer Vision*, pages 640–653. 20

Kim, S. J., Frahm, J.-M., and Pollefeys, M. (2007). Joint feature tracking and radiometric calibration from auto-exposure video. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1–8. 56

Kim, S. J., Frahm, J.-M., and Pollefeys, M. (2008a). Joint radiometric calibration and feature tracking for an adaptive stereo system. *submitted to Computer Vision and Image Understanding*. 56

Kim, S. J., Frahm, J.-M., and Pollefeys, M. (2008b). Radiometric calibration with illumination change for outdoor scene analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. 86

Kim, S. J. and Pollefeys, M. (2004). Radiometric alignment of image sequences. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 645–651. 25

Kim, S. J. and Pollefeys, M. (2008). Robust radiometric calibration and vignetting correction. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 30(4):562–576. 25

Koppal, S. J. and Narasimhan, S. G. (2006). Clustering appearance for scene analysis. pages 1323 – 1330. 87, 88

Levin, A., Zomet, A., Peleg, S., and Weiss, Y. (2004). Seamless image stitching in the gradient domain. In *Proceedings of the European Conference on Computer Vision*, pages 377–389. 24

Lin, S., Gu, J., Yamazaki, S., and Shum, H. (2004). Radiometric calibration from a single image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 938–945. 19

Lin, S. and Zhang, L. (2005). Determining the radiometric response function from a single grayscale image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 66–73. 19

Litvinov, A. and Schechner, Y. (2005a). Addressing radiometric nonidealities: A unified framework. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 52–59. 21, 26, 28, 39, 42, 43, 44, 45, 46, 47, 49, 54

Litvinov, A. and Schechner, Y. (2005b). Radiometric framework for image mosaicking. *Journal of Optical Society of America*, 22(5):839–848. 21, 26, 28, 41, 48, 106

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the Conference on Artificial Intelligence*, pages 674–659. 55, 58, 71

Manders, C., Aimone, C., and Mann, S. (2004). Camera response function recovery from different illuminations of identical subject matter. In *Proceedings of IEEE International Conference on Image Processing*, pages 24–27. 18, 85, 102

Mann, S. (2000). Comparametric equations with practical applications in quantigraphic image processing. *IEEE Transactions on Image Processing*, 9(8):1389–1406. 30, 56, 57

Mann, S. and Mann, R. (2001). Quantigraphic imaging: Estimating the camera response and exposures from differently exposed images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 842–849. 17

Mann, S. and Picard, R. (1995). On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Procedings of IS&T 46th annual conference*, pages 422–428. 16

Martinkauppi, B. (2002). *Face colour under varying illumination - analysis and applications*. PhD thesis, University of Oulu, Oulu, Finland. 12

Matsushita, Y. and Lin, S. (2007). Radiometric calibration from noise distributon. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1 – 8. 19

Mitsunaga, T. and Nayar, S. (1999). Radiometric self-calibration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 374–380. 17

Narasimhan, S. G., Wang, C., and Nayar, S. K. (2002). All the images of an outdoor scene. In *Proceedings of the European Conference on Computer Vision*, pages 148–162. 82

Pal, C., Szeliski, R., Uyttendaele, M., and Jojic, N. (2004). Probability models for high dynamic range imaging. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 173–180. 17

Park, J., Lee, M., Grossberg, M. D., and Nayar, S. K. (2007). Multispectral imaging using multiplexed illumination. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1–8. 107

Pollefeys, M., Gool, L. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., and Koch, R. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3):207–232. 27, 45, 76

Pollefeys, M., Nister, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S. J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., and Towles, H. (2008). Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision*, 78(2–3):143–167. 77

Pritch, Y., Rav-Acha, A., Gutman, A., and Peleg, S. (2007). Webcam synopsis: Peeking around the world. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1–8. 82

Sawchuk, A. A. (1977). Real-time correction of intensity nonlinearities in imaging systems. *IEEE Transactions on Computers*, 26(1):34–39. 20

Schechner, Y. and Nayar, S. (2002). Generalized mosaicing: Wide field of view multispectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1334–1348. 107

Schechner, Y. and Nayar, S. (2003). Generalized mosaicing: High dynamic range in a wide field of view. *International Journal of Computer Vision*, 53(3):245–267. 20

Shafique, K. and Shah, M. (2004). Estimation of the radiometric response functions of a color camera from differently illuminated images. In *Proceedings of IEEE International Conference on Image Processing*, pages 2339–2342. 18, 85, 102

Shi, J. and Tomasi, C. (1994). Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600. 55, 58, 59, 71

Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3d. *ACM Transactions on Graphics*, 25(3):835–846. 107

Sunkavalli, K., Matusik, W., Pfister, H., and Rusinkiewicz, S. (2006). Factored time-lapse video. *ACM Transactions on Graphics*, 26(3). 95, 96

Sunkavalli, K., Romeiro, F., Matusik, W., Zickler, T., and Pfister, H. (2008). What do color changes reveal about an outdoor scene? In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. 107

Tsin, Y., Ramesh, V., and Kanade, T. (2001). Statistical calibration of the ccd imaging process. In *Proceedings of IEEE International Conference on Computer Vision*, pages 480–487. 17

Uyttendaele, M., Criminisi, A., Kang, S. B., Winder, S., Hartley, R., and Szeliski, R. (2004). Image-based interactive exploration of real-world environments. *IEEE Computer Graphics and Applications*, 24(3):52–63. 20

Vasilescu, M. A. O. and Terzopoulos, D. (2002). Multilinear analysis of image ensembles: Tensorfaces. In *Proceedings of the European Conference on Computer Vision*, pages 447–460. 107

Weiss, Y. (2001). Deriving intrinsic images from image sequences. In *Proceedings of IEEE International Conference on Computer Vision*, pages 68–75. 2, 82, 83

Welch, G. and Bishop, G. (1995). An introduction to the kalman filter. Technical Report TR-95-031, Department of Computer Science, University of North Carolina at Chapel Hill. 69

Yu, W. (2004). Practical anti-vignetting methods for digital cameras. *IEEE Transactions on Consumer Electronics*, 50(4):975–983. 8, 20

Yu, W., Chung, Y., and Soh, J. (2004). Vignetting distortion correction method for high quality digital imaging. In *Proceedings of IEEE International Conference on Pattern Recognition*, pages 666–669. 20

Zach, C., Gallup, D., and Frahm, J.-M. (2008). Fast gain-adaptive klt tracking on the gpu. In *Proceedings of Workshop on Visual Computer Vision on GPUs*, pages 1–7. 76

Zheng, Y., Lin, S., and Kang, S. B. (2006). Single-image vignetting correction. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 461–468. 7, 20