論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Non-Photorealistic Rendering and Informative Region Detection for High Dynamic Range Scenes
著者(和文)	Lu Jiajun
Author(English)	Lu Jiajun
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第9875号, 授与年月日:2015年3月26日, 学位の種別:課程博士, 審査員:廣田　薫,新田　克己,柴田　崇徳,室伏　俊明,小野　功
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第9875号, Conferred date:2015/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Non-Photorealistic Rendering and Informative Region Detection for High Dynamic Range Scenes

Lu Jiajun

Supervisor: Prof. Kaoru Hirota

Doctoral Thesis

Tokyo Institute of Technology

Interdisciplinary Graduate School of Science and Engineering

Department of Computational Intelligence and Systems Science

March 2015

Abstract

In contrast to the conventional quest for realism in the field of rendering, a new quest is growing for creating imagery that is more artistic and easy for communication. The limited dynamic range of normal digital images, however, imposes restrictions on the performance of some Non-Photorealistic Rendering (NPR) methods. An NPR strategy is proposed for rendering High Dynamic Range (HDR) images to mimic the visual perception of human artists and designers. The achromatic brightness (luminance) channel of an HDR image is used to compute gradient values for controlling the placement of elements such as strokes in the rendering process. This modification in the rendering process is presumed to give rise to the occurrence of more accurate strokes, which is rewarding in certain rendering styles, e.g., the painterly rendering. In the proposal, a tone-mapped HDR image of a scene is chosen to provide color information for the rendering process, of which the choice depends on the scene to be rendered. These changes in the information source of the rendering process also enable detailed structures and colors to be preserved in the resulting images, especially for regions that present particularly high or low luminance. Experimental results using a designed pattern show that angles of gradient values obtained from HDR images have a reduction in averaged error of up to 57.5% in comparison to that of conventional digital images. A partial experiment on incorporating HDR images into other NPR styles, such as dithering, shows the wide compatibility of HDR images in providing source information for NPR processes.

Car-mounted camera usually encounters the problem that visual contents in bright region of specific scene are missing due to over-exposure (e.g., the daytime scene of tunnel exit). So, there is a necessity that the Informative Bright Region (IBR) in tunnel exit scene being located in order to find optimum timing for the control process of imaging sensor. An algorithm for estimating location of IBR in tunnel scenes is proposed. First, tunnel lights are marked using the brightness and color tone information of them, and markings of traffic lanes are detected using their geometric properties. Second, a confidence map, which represents the possibility of IBR existence, is calculated by combining the geometric information of both tunnel lights and traffic lane markings. Third, the most possible location of IBR is determined by searching the confidence map of the current frame, and a re-locating process is carried out by using already detected IBR locations in the previous frames, if needed. Finally, a refinement of the IBR location is done to compensate for the view angle of car-mounted camera by using the geometric properties of the over-exposed region. To evaluate the performance of the proposed algorithm, a dataset consisting of three typical tunnel scenes on general roads is created. It is evaluated by comparing

the estimated location of IBR with that annotated by multiple human observers on the self-built tunnel scene video dataset, and running time of the algorithm on each frame achieves 10ms. The locations of IBR given by the algorithm have averaged error of 4.14 pixels out of 640 pixels on X-axis and 4.23 pixels out of 480 pixels on Y-axis in comparison with the IBR labeled by the human observers. The algorithm aims at providing control timing of imaging sensor on a low cost platform like vehicle driving recorder to enhance the captured visual contents in over-exposed regions.

The high dynamic range of daytime tunnel scene makes it difficult to capture scene information with conventional car-mounted camera, of which the captured dynamic range is limited by the quantization level of each individual pixel. Based on the observation that the same quantity in the low portion of quantized bits may provide much more details than that in the high portion, a method for visually enhancing the IBR by triggering exposure compensation on the imaging device (e.g. camera-mounted camera) in HDR tunnel scenes is proposed. In brief, the exposure of imaging device is adjusted based on the change of the detected IBR in a tunnel scene, and the captured image sequence is compensated using gamma correction. Experiment is carried out by porting the IBR location detection algorithm on single-board computer with a camera module. The IBR location detection algorithm runs during the video capturing process. Exposure compensation through EV adjustment is triggered by monitoring the size change of IBR. The captured video sequence is thus influenced by the adjusted camera parameter (EV bias). The experimental results show that the tunnel exit in a strong luminance condition is made visible after this vision enhancement process.

Acknowledgements

This doctoral dissertation is the best reflection of my study and life in Tokyo Institute of Technology. And the experience of research, whether pleasant or not, has melted into my blood vessels. First, I'd like give my most sincere thanks to my supervisor Prof. Kaoru Hirota. His supervision is always casting light of wisdom on this long journey. Prof. Fangyan Dong is like the mother figure who always gives me her best kindness and support. Ms. Harumi Hoshino, the secretary of Hirota lab, is the reliable lady who supports all the lab members. The discussions with my lab mates are so enlightening that I will never forget those experience. They are Tianyu Li, Luefeng Chen, Fei Yan, Chehung Lin, Atsushi Shibata, Kazuhiro Ohnishi, Toshihiro Akamatsu, and so on. During more than three years of study in Japan, my dear parents are caring and missing me for every minute of their day. They are my most honest supporters, and our hearts are always together. During every day of my study in Japan, I feel grateful for the financial support from MEXT (Ministry of Education, Culture, Sports, Science and Technology) and for every Japanese citizen. I have so many appreciations to give, and surely I will contribute my power to add a small piece to happiness of humanity.

Contents

1 Introduction **1**

2 Imaging in High Dynamic Range Scenes **9**

2.1 High Dynamic Range Image Composition 10

2.2 Tone Mapping for High Dynamic Range Images 13

2.3 Hardware for High Dynamic Range Imaging 17

2.4 Chapter Summary 20

3 Non-Photorealistic Rendering for High Dynamic Range Scenes **21**

3.1 Gradient-related Non-Photorealistic Rendering 22

3.2 Gradient of High Dynamic Range Images in Non-Photorealistic Rendering 24

3.3 Experiments on Gradient Evaluation and Rendering Results 27

3.3.1 Assessment by Error in Gradient Computation of HDR Images . . . 27

3.3.2 Rendering Results for HDR Images 33

3.4 Chapter Summary 39

4 Informative Bright Region Detection in High Dynamic Range Scenes **42**

4.1 Response of Imaging Device due to Luminance in Tunnel Scenes 43

4.2 Locating Informative Bright Region using Lighting and Traffic Lane Cues in Tunnel Scenes 45

4.2.1 Lighting and Traffic Lane Cues for Locating IBR 45

4.2.2 Confidence Map Computation 48

4.2.3 IBR Re-locating and View Angle Compensation 51
4.3 Experiments on Locating Informative Bright Regions on General Roads . . 53
4.3.1 Tunnel Video Dataset and Labeling Tool 53
4.3.2 Evaluation of IBR Locating Accuracy and Computation Cost . . . 54
4.4 Chapter Summary 59

5 Vision Enhancement of Informative Bright Region in High Dynamic Range Scenes 62
5.1 Vision Enhancement of Informative Bright Region on Imaging Device . . . 64
5.1.1 Exposure Control of Imaging Device for Informative Bright Region Enhancement 64
5.1.2 Dark Region Enhancement by Gamma Correction 67
5.2 Vision Enhancement of Informative Bright Region by Liquid Crystal Glasses 76
5.3 Experiments of Enhancing Tunnel Scenes on General Roads 78
5.4 Chapter Summary 81

6 Conclusion 84

Bibliography 87

Related Publications 91

List of Figures

1.1 A simplified flowchart showing NPR strategy for HDR scenes. 3
1.2 A Simplified flowchart showing the IBR detection algorithm. 5
1.3 The processing of visual enhancement of HDR tunnel scene using IBR detection. 7
1.4 The structure of the thesis. 8

2.1 A "cup" scene used to recover the camera response curve (Canon^{TM} EOS 300D). 13
2.2 The recovered camera response curve using the image set shown in Figure 2.1. (a) The camera response curve for red channel. (b) The camera response curve for green channel. (c) The camera response curve for blue channel. 14
2.3 A "corner" scene with a window used to reconstruct the HDR radiance map (HDR image). 15
2.4 The reconstructed HDR radiance map (HDR image) in log scale using the image set shown in Figure 2.3. 16
2.5 The log-scaled base layer and the log-scaled detail layer of the HDR scene radiance map in Figure 2.4 calculated by fast bilateral filter. (a) The base layer of Figure 2.4. (b) The detail layer of Figure 2.4. (c) The compressed base layer of Figure 2.4 by a factor of 3.5. (d) The unchanged detail layer of Figure 2.4. 17

2.6 The original HDR scene radiance map (Figure 2.4) in log scale compared with the one whose base layer is compressed by a factor of 3.5. (a) The original scene radiance map shown in Figure 2.4. (b) The scene radiance map whose base layer is compressed using fast bilateral filter by a factor of 3.5. 18

2.7 The linearly-scaled base layer before and after compression together with the linearly-scaled detail layer of the HDR scene radiance map in Figure 2.4 calculated by fast bilateral filter. (a) The uncompressed base layer of Figure 2.4 in linear scale. (b) The detail layer of Figure 2.4 in linear scale. (c) The compressed base layer (by a factor of 3.5 in log scale) of Figure 2.4 in linear scale. (d) The unchanged detail layer of Figure 2.4 in linear scale. 19

2.8 The tone-mapped HDR image by fast bilateral filter (the "corner" scene shown in Figure 2.3). 19

3.1 Rendering result of Lena in a painterly style. 23

3.2 Rendering result of Lena in a decorative mosaic style. 24

3.3 Rendering process for HDR images. 26

3.4 A 3D plot for the scene radiance function $I_F(x, y)$ for which each value is given by (3.7). 28

3.5 Response curve for generating the virtual digital images. 29

3.6 The generated virtual images with their shutter time. 30

3.7 Angular maps of gradient values computed using the scene radiance function, the constructed HDR image, and one of the generated virtual images (shutter time: $0.6s$). (a) $\theta_F(m, n)$. (b) $\theta_{HDR}(m, n)$. (c) $\theta_{NOR}(m, n)$. 31

3.8 A plot for $\Delta\theta_{HDR}$ and $\Delta\theta_{NOR}$. (a) Errors in computing the gradient value of the constructed HDR image. (b) Errors in computing the gradient values of a generated image (shutter time: $0.6s$) from image set I. 32

3.9 Averaged error in gradient computation of the HDR image ($\overline{\Delta\theta_{HDR}}$) and the generated images ($\overline{\Delta\theta_{NOR}}$). For all of the generated images in image set I, $\overline{\Delta\theta_{NOR}}$ values are computed. 32

3.10 Image sets used for constructing HDR images 33
3.11 The response curve of the digital camera used in this work. 34
3.12 Tone-mapped HDR images for three scenes of the image sets shown in Figure 3.10. 35
3.13 Rendering results of HDR images and normal images using painterly rendering. (a), (c), (e) Rendering results for HDR images of scenes in Figure 3.10. (b), (d), (f) Rendering results for conventional digital images of the scenes. 36
3.14 Rendering results for HDR images and conventional digital images using the decorative mosaic style. (a), (c), (e) Rendering results for the HDR images. (b), (d), (f) Rendering results for the conventional digital images. . 37
3.15 First attended regions (green border) and saliency maps of the rendering results. (a), (b), (c) Rendering results for HDR images with first attended regions. (d), (e), (f) Saliency maps for the rendered HDR images. (g), (h), (i) Rendering results for conventional digital images with first attended regions. (j), (k), (l) Saliency maps for rendered conventional digital images. (results of "SaliencyToolbox" by Walther and Koch [18][19]). 39
3.16 Rendering results for an HDR image and a conventional digital image in the dithering style. (a) Rendering results for an HDR image of the third scene in Figure 3.10. (b) Rendering results for a conventional digital image of the third scene in Figure 3.10. 40

4.1 Image sequence of a daytime tunnel scene captured by a car-mounted camera. (a) A frame captured far from the tunnel exit. (b) A frame captured close to the tunnel exit. 44
4.2 A flowchart of the locating process of IBR using lighting and traffic lane cues. 46
4.3 Light pixel marked in tunnel scenes (by the center of the cross mark) along with their RGB values. 47

4.4 Light mask computed from two tunnel scenes. (a) Tunnel scene with orange lights ($THRES = 230$). (b) Light mask of the tunnel scene (a). (c) Tunnel scene with white lights. (d) Light mask of the tunnel scene (c). 48
4.5 Combined image of light mask and pavement part for edge detection. (a) Original image. (b) Combined image. (c) Result of detected edges. 49
4.6 Line segments detected on Figure 4.5(c) and shown in the original frame. . 49
4.7 Superimposing process of weight belts for computing confidence map. . . . 51
4.8 Confidence map computed from traffic lane and lighting cues in Figure 4.6. 51
4.9 A frame with the highest value in its confidence map (Figure 4.8) marked as a cross. 52
4.10 A frame with its LOC_{MAX}, $LOC2_{MAX}$, and IBR_n positions marked as circle, triangle, and rectangle shapes. 53
4.11 Three typical tunnel scenes in the self-built dataset. 55
4.12 The window form of the designed labeling tool with the marker. 56
4.13 The examples of confidence map for the three tunnel scenes in Figure 4.11. (a)(c)(e) The original frames for calculating confidence map. (b)(d)(f) The obtained confidence map. 57
4.14 The calculated IBR locations using the algorithm for the three tunnel scenes in Figure 4.11. (a)(c)(e) The original frames for calculating IBR location. (b)(d)(f) The obtained IBR locations in the corresponding frames. 58
4.15 The location errors of the algorithm compared with the labeled data (cross: before view angle compensation, circle: after view angle compensation). (a) Errors occurred the scene in Figure 4.11(a). (b) Errors occurred in the scene in Figure 4.11(b). (c) Errors occurred in the scene in Figure 4.11(c). 60
4.16 Processing speed of the algorithm in the three tunnel scenes. (a) Processing speed in the scene of Figure 4.11(a). (b) Processing speed in the scene of Figure 4.11(b). (c) Processing speed in the scene of Figure 4.11(c). 61

5.1 An abstracted control circuit for imaging device. 65

5.2 Video frames captured by a car-mounted camera when driving out from a tunnel in daytime. 66
5.3 IBR as a factor for exposure control in image devices. 67
5.4 Three tunnel scenes with the searching area (80×80) for over-exposed pixels centered in IBR location. 68
5.5 Size change of over-exposed region in practical daytime tunnel scenes shown in Figure 5.4. 69
5.6 Examples of applying the triggering strategy (searching area set to 80×80, $A_{trigger}$ set to 2500 and 3500) to three tunnel scenes in Figure 5.4.(a)(d)(g) The size change of over-exposed region with two overlaid triggers. (b)(e)(h) The frame correspond to the trigger with $A_{trigger}$ set to 2500. (c)(f)(i) The frame correspond to the trigger with $A_{trigger}$ set to 3500. 70
5.7 An image set of daytime tunnel exit showing influence of EV adjustment by controlling shutter duration alone ($f/5.6$, ISO:800). 71
5.8 Enhancement of dark region in a captured still image by gamma correction ($\gamma = 0.7$, 0.6, 0.5, 0.4, and 0.3). 72
5.9 The change of average absolute difference with respect to value of γ. (a) The original frame before gamma correction. (b) The historical reference frame. (c) Average absolute difference for varying γ value. 74
5.10 The gamma-corrected frame using the optimal γ value from Figure 5.9(c) ($\gamma = 0.45$) together with the histogram. (a) The original frame before gamma correction. (b) The histogram of the original frame before gamma correction. (c) The historical reference frame. (d) The histogram of the historical reference frame. (e) The gamma-corrected frame using the optimal γ value. (f) The histogram of the optimal gamma-corrected frame. 75
5.11 A color checker quantized with 8-bit and 2-bit. (a) Quantized image of 8-bit. (b) Quantized image of 2-bit. 76

5.12 A color checker quantized with 2-bit after adding uniformly distributed noise. (a) Quantized image of 2-bit with added noise. (b) Recovered image by averaging the pixels in each block. 77
5.13 The vision enhancement process of IBR on imaging device. 78
5.14 The light transmittance of the liquid crystal glasses controlled by the applied voltage. 79
5.15 The hardware connection of single-board computer and liquid crystal glasses. 79
5.16 The single-board computer (Raspberry Pi Model B) and its camera module fixed on a plastic box. 80
5.17 The relation between the IBR locating algorithm and function modules of the camera. 80
5.18 The time duration for encoding each frame (on the output port of the encoder component) and for IBR detection and exposure compensation triggering (on the preview port of the camera component). 81
5.19 Two frames captured before and after the triggering of exposure compensation (+4 EV) in two typical tunnel scenes.(a)(c) Before exposure compensation. (b)(d) After exposure compensation. 82
5.20 The state change of the liquid crystal glasses together with the IBR detection results for two video sequences of daytime tunnel scenes. 83

List of Tables

3.1 Parameters for scene radiance function. 28

3.2 Shutter time used in generating virtual images. 28

4.1 Averaged location error of the IBR detection algorithm (in pixel). 56

4.2 Averaged processing speed of the algorithm for one frame. 59

Chapter 1

Introduction

Vision is considered as one of the most important sources for human beings in the process of collecting the environmental information. From ancient times, efforts have been made to record and reproduce the scenes in certain forms (e.g., line sketch, Chinese blush painting, and oil painting), of which these styles are reflections of human vision through certain materials. The progress of modern chemistry enabled the mass production of light-sensitive materials, which paves the road to modern photography technologies. The dodging and burning techniques, which are well applied in film enlargement process by photographers, have shown well-defined concept of processing. However, modern digital image processing is more recognized as a branch of computer science.

In contrast to the conventional quest for realism in the field of rendering, a new quest is growing for creating imagery that is more artistic and easy for communication [1]. Rendering methods developed for enabling a variety of expressive styles are referred to as NPR (Non-Photorealistic Rendering). Typical NPR styles are usually inspired by artistic works, e.g., oil painting [2] and pen-and-ink illustration [3]. In NPR tasks, three dimensional models, vector fields, or simply digital images serve as input information. Among these types of input data, digital images of real scenes are usually utilized as references for generating NPR results [1]. The limited dynamic range of normal digital images, however, imposes restrictions on the performance of some NPR methods. With the development of image processing technologies, HDR (High Dynamic Range) images have attracted wide attention [4][5][6][7][8]. The development of HDR imaging technologies, method for reconstructing HDR image, and tone-mapping of HDR image are introduced in Chapter 2.

Limitations on the performance of NPR methods resulting from low dynamic range of

digital images are overcome when HDR images are incorporated into the NPR process. Some researchers have shown an interest in developing NPR methods for HDR images. Colbert et al., for example, proposed two kinds of brushes for virtual painting on unmapped HDR imagery [9]. Vergne et al. proposed a pipeline for the stylized rendering of HDR environment-mapped objects [10]. Smith et al. showed that HDR images provide better results in edge detection and segmentation in NPR [11].

In Chapter 3, attention is paid to gradient values of HDR images in developing the NPR strategy. The proposal mainly focuses on blending HDR images into conventional gradient-related NPR methods such as the painterly rendering proposed by Hertzmann [2] and the decorative mosaic rendering developed by Hausner [12]. The achromatic brightness (luminance) channel of an HDR image is used to compute gradient values for controlling the placement of elements such as strokes in the rendering process. This modification in the rendering process [2] is assumed to give rise to the occurrence of more accurate strokes, which is rewarding in certain rendering styles, e.g., the painterly rendering. For an NPR process, rendering results may be greatly affected by the color information used. In the proposal, a tone-mapped HDR image or a well-exposed digital image of a scene is chosen to provide color information for the rendering process, of which the choice depends on the scene to be rendered. These changes in the information source of the rendering process also enable detailed structures and colors to be preserved in the resulting images, especially for regions that present particularly high or low luminance. The proposed NPR strategy is depicted in Figure 1.1.

In order to evaluate the proposal, two experiments are carried out, and the desktop computer used for the experiments has a 2.66GHz CPU and 2GB of memory. In the first experiment, averaged angular errors in gradient computation for both a set of generated digital images and a constructed HDR image are figured out by simulating a photo capture process in which the radiance map of a virtual scene is a superposition of three Gaussian functions whose gradient values are easily obtainable in the closed form. An image dataset composed of three scenes is created using a digital single-lens reflex camera for the second experiment. For each scene in the image dataset, a set of images is photographed at different shutter timings. In the second experiment, intuitive rendering results are given to show visual improvements when the proposal is applied to typical NPR styles, e.g., painterly rendering [2] and decorative mosaic rendering [12]. A bottom-up

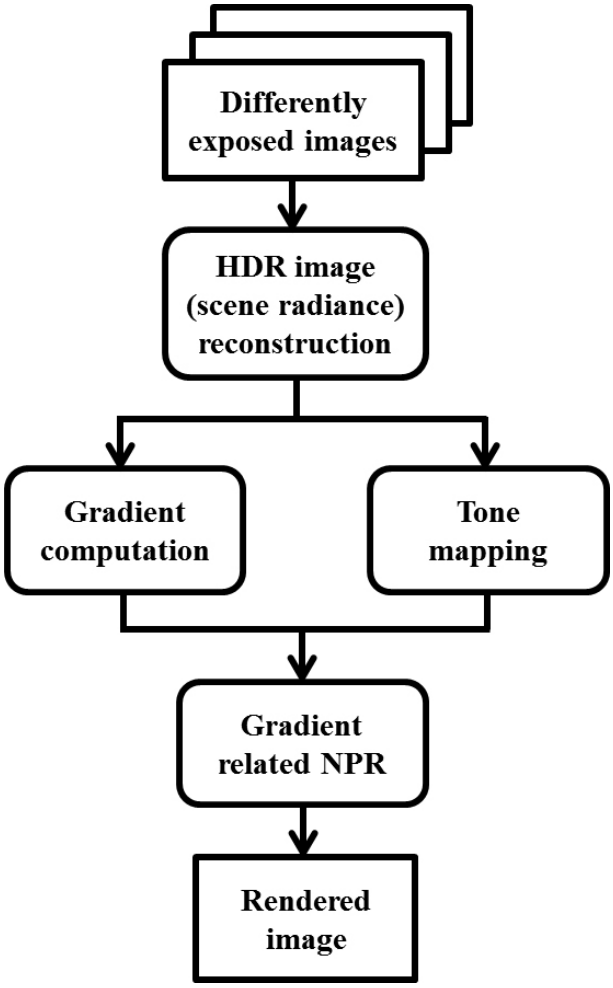


Figure 1.1: *A simplified flowchart showing NPR strategy for HDR scenes.*

visual attention model is then applied to evaluate these rendering results. The attended regions in rendering results show the applicability of the proposed method for scenes with rapidly changing light conditions.

Visual cues like traffic lanes [13][14] and other traffic signs [15][16] are recognized from image sequence captured by car-mounted camera in autonomous driving systems [17]. Car-mounted camera usually encounters the problem that visual contents in bright region of specific scene are missing due to over-exposure (e.g., the daytime scene of tunnel exit). To capture such a wide dynamic range scene, a variety of HDR image composition methods [5][6][4][18] and hardware [19][20] have been developed. Hardware and computation costs, however, limit the large promotion of these well-designed methods. For a daytime tunnel scene in which the region outside tunnels is over-exposed, shortening the shutter time or lowering the gain of Charge Coupled Device (CCD) imaging sensor at the right time enables acquisition of the over-exposed region (i.e., IBR: Informative Bright Region).

Also, the incidence rate of traffic accidents in tunnel scene tends to increase near the tunnel exits [21]. So, there is a necessity that the IBR in tunnel exit scene being located in order to find optimum timing for the control process.

An algorithm for estimating location of IBR in tunnel scenes is proposed in Chapter 4. In a typical tunnel scene, regularly laid lights on tunnel roof and road marking of traffic lane are preferable cues for the estimation process. First, tunnel lights are marked using the brightness and color tone information of them, and markings of traffic lanes are detected using their geometric properties. Second, a confidence map, which represents the possibility of IBR existence, is calculated by combining the geometric information of both tunnel lights and traffic lane markings. Third, the most possible location of IBR is determined by searching the confidence map of the current frame, and a re-locating process is carried out by using already detected IBR locations in the previous frames, if needed. Finally, a refinement of the IBR location is done to compensate for the view angle of car-mounted camera by using the geometric properties of the over-exposed region. The IBR detection algorithm is further depicted by a simplified flowchart as Figure 1.2.

Compared with the tunnels on expressways, of which the lighting conditions are well designed, the algorithm is designed to deal with more challenging tunnel scenes on local roads with comparatively irregular lighting conditions. It takes advantage of visual cues which are closely related with IBR and not severely affected by luminance condition outside the tunnel. So the algorithm locates IBR under the condition that the visual information inside IBR is lost. Lighting and traffic lane cues are extracted using low cost image processing techniques (e.g., Canny edge detector), and the confidence map computation is done in limited region of the whole frame to achieve the real-time processing demand.

To evaluate the performance of the proposed algorithm, a dataset consisting of three typical tunnel scenes on local roads is created. The video clips in the dataset are recorded by a camcorder mounted on a vehicle when it is passing through tunnels in the daytime. The first experiment is designed to make a comparison between the estimated locations of IBR given by the algorithm with those marked by four human observers. The testing platform used for experiments is a PC with a 3.4GHz CPU and 8GB of memory. The video clips used in experiment are normalized into the same size (640 pixels in width and 480 pixels in height). For each video clip, the human observers are requested to mark the

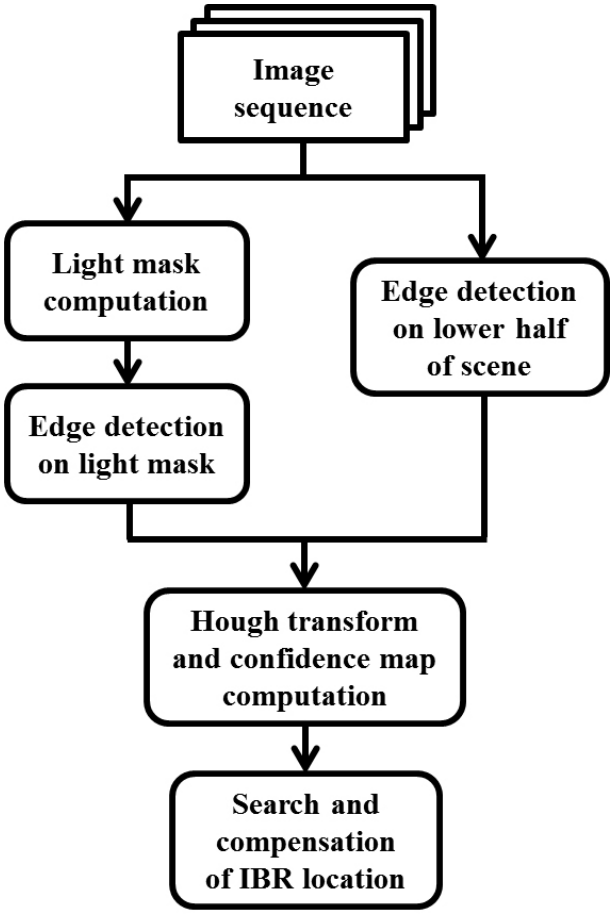


Figure 1.2: *A Simplified flowchart showing the IBR detection algorithm.*

potential IBR in 30 frames on their best efforts, and the algorithm is run on the same video for providing an estimation of the IBR. The second experiment is targeted at testing processing speed of the algorithm on the same platform. The averaged processing time is figured out by recording the time consumed in processing each frame of the videos in the dataset.

The video stream output of car-mounted camera is usually influenced by the rapidly changing luminance condition on general roads. Tunnel scene is a typical one with changing luminance condition. The high dynamic range of tunnel scene makes it different to capture scene information with conventional car-mounted camera, of which the captured dynamic range is limited by the quantization level of each individual pixel. For capturing an HDR scene in its full dynamic range, typical methods such as scene radiance map reconstruction using multiple exposure have been developed by Debevec et al. [5]. Extension of this multiple exposure method from single HDR image to HDR video has also been reported [19]. However, specially designed imaging device, whose shutter

speed is possible for fast adjustment, is necessary. Also, the high computation cost for post-processing of this HDR video capturing method limits its application in low-cost hardware like driving recorder. Mertens et al. [22] proposed a technique for fusing a set of differently exposed images into a high quality one. This image fusion technique skips the reconstruction process of HDR image (scene radiance map), and the tone mapping step is also not needed accordingly. For capturing video in an HDR tunnel scene using this technique, highly controllable imaging device and additional computation resource for image registration are still imperative as Kang's method does[19]. Tocci et al. [20] introduced a optical structure, which generates differently exposed images with beam spitters, for capturing HDR video. Instead of using single imaging sensor, at least three imaging sensor is needed for constructing the optical structure. Based on the observation that the same quantity in the low portion of quantized bits may provide much more details than that in the high portion, a method for adjusting the exposure of the video capture device (e.g. camera-mounted camera) in HDR tunnel scenes is proposed in Chapter 5.

In brief, the exposure of video capture device is adjusted based on the change of the detected IBR in a tunnel scene, and the captured image sequence is compensated using adaptive gamma correction. There are mainly four steps in this proposal. First, the location of IBR (over-exposed tunnel exit) is detected using the algorithm introduced in Chapter 4. Second, size of over-exposed region around the IBR location is figured out and recorded for each frame. Third, a certain level of exposure compensation is triggered when the size of over-exposed region exceeds a predefined threshold. Finally, the captured frame under exposure compensation is enhanced using gamma correction.

Forcing less exposure during video capture makes the dynamic range more adapted to the over-exposed exit region of the tunnel scene. Considering the prerequisite that computation power of hardware is limited, the dark region caused by the forced exposure compensation is enhanced by adaptive gamma correction in each pixel. By selecting a suitable gamma value, which is mainly decided referring to light condition inside and outside the tunnel, both the strongly illuminated objects observed through tunnel exit and the weakly illuminated region inside the tunnel are recorded and presented in a clearly perceptible manner. The proposed process for visual enhancement of HDR tunnel scene using IBR detection is shown in Figure 1.3.

Experiment is carried out by porting the IBR location detection algorithm on single-

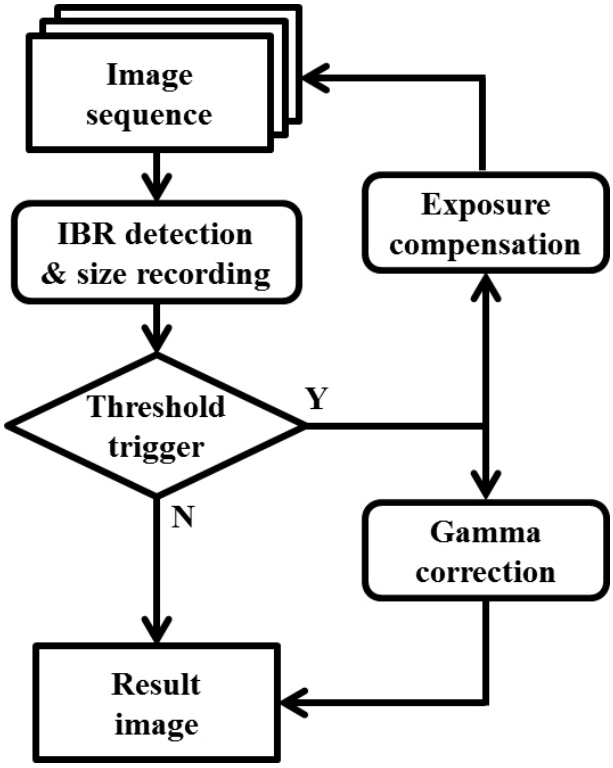


Figure 1.3: *The processing of visual enhancement of HDR tunnel scene using IBR detection.*

board computer with a camera module. The IBR location detection algorithm runs during the video capturing process. Camera parameter related with exposure compensation is adjusted according to area ratio of the IBR. And the captured video sequence is influenced by the adjusted camera parameter and is saved in storage media after an encoding process.

The structure of the thesis is briefly summarized in Figure 1.4.

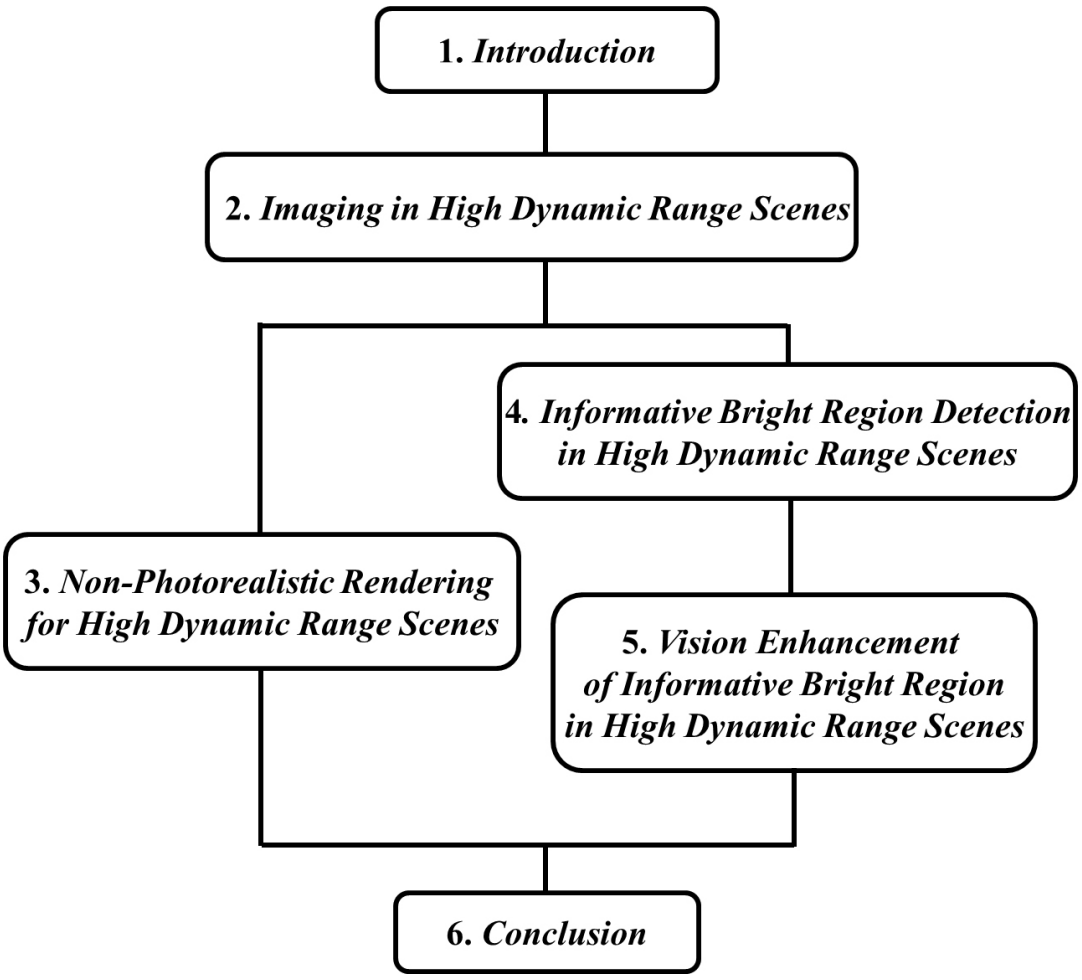


Figure 1.4: *The structure of the thesis.*

Chapter 2

Imaging in High Dynamic Range Scenes

Reproducing scenes with High Dynamic Range (HDR) has become a challenging topic for artists, photographers, and computer vision scientists. Great effects have been put on seeking methods for reproducing scenes with spatially varying luminance conditions in a more natural way. From a different point of view, the reproduced works (paintings, photos, and digital images) have to be visually similar to human vision system.

Painters have developed various means of artistic expression for HDR scenes. Leonardo da Vinci's "Lady with an Ermine", Gerrit van Honthorst's "The Childhood of Christ", and Rembrandt Harmenszoon van Rijn's "The Night Watch" all give the examples of painters rendering for HDR scenes [23]. According to McCann [23], the range of light depicted in these paintings exceeds the physical limit of dynamic range that can be produced by oil-painting. These painters observed or even just imagined the response of human vision system and used the painting techniques they developed to reproduce luminance of scenes.

From the 19th century, the development of fundamental industry supports the wide use of photography. Photographers also faced the same problem that painters have faced hundreds of years before. Robinson's well-known work "Fading Away", which captures an HDR scene by a composition of several differently exposed negative films, showed a typical solution for that issue [23]. Later, Mees studied this method of capturing HDR scenes systematically [23][24], and he also argued that HDR scenes can be reproduced by a combination of low-slope negative films and high-slope prints [23][25].

In the designing process of digital cameras, their response curves are adjusted to be an approximation of the film cameras. And the outputs of digital cameras are limited due to the fact of its quantized values (usually 8-bit for one pixel in each channel). Hence, it is natural that the multiple exposure technique is inherited from film cameras to digital

cameras for capturing HDR scenes.

2.1 High Dynamic Range Image Composition

For a developed negative film which has captured an HDR scene, dodging and burning techniques are used to produce a print in which details of the scene are enhanced [23][26]. When dodging is carried out, the dark area is slightly lightened due to the negative acting fact of the print. When burning is carried out, the corresponding area is made darker by the increased exposure on the print [23].

Due to the existence of fixed camera response curve for each digital camera, Debevec and Malik [5] proposed a method for recovering the scene radiance by compositing a set of digital images with different exposure. This method is well expected because it provides a simple way of finding the non-linear camera response curve. The mapping from CCD sensor output to pixel value of digital image is one of the reasons causing this non-linearity [5].

The recovering process of the scene radiance function proposed by Debevec et al. [5] involves two steps . The first step is to find the camera response curve of the digital camera. The second step is to construct the scene radiance (in other words: High Dynamic Range image) using the obtained the camera response curve. According to [5], the relationship among irradiance, shutter time, and quantized pixel value was formulated as

$$g(Z_{ij}) = \ln E_i + \ln \Delta t_j, \tag{2.1}$$

where $g = \ln f^{-1}$ is the log inverse function of the camera response function, Z_{ij} denotes the pixel value at location i in digital image j, E_i is the irradiance function for location i, and Δt_j represents the shutter time for j-th digital image. Among them, the Z_{ij} and Δt_j are unknowns. For recovering the camera response curve, namely g, Debevec et al. [5] designed the following objective function

$$\mathcal{O} = \sum_{i=1}^{N}\sum_{j=1}^{P}\{w(Z_{ij})[g(Z_{ij}) - \ln E_i - \ln \Delta t_j]\}^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1}[w(z)g^{''}(z)]^2, \tag{2.2}$$

where N denotes the number of pixels from each image, P denotes the total number of digital images used for recovering the camera response curve, $w(Z)$ represents the weight

function related with the pixel value Z, $g''(z)$ is the second derivative of log inverse function of camera response curve. Since the pixel values are quantized as discrete values, Debevec [5] used $g''(z) = g(z-1)-2g(z)+g(z+1)$. The first part of the objective function makes sure the relation between exposure $E_i\Delta t_j$ and quantized pixel value Z_{ij} is satisfied, and the second part is a term which suggests a more smooth camera response curve by putting constraints on the second derivative of $g(z)$. It is easy to understand that λ takes the balance between more accurate fitting and smoothness assumption. Debevec [5] also suggested that the a simple but well-worked choice of weight function has a hat shape

$$w(z) = \begin{cases} z - Z_{min}, & \text{for } z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z, & \text{for } z > \frac{1}{2}(Z_{min} + Z_{max}) \end{cases}, \tag{2.3}$$

where z is the pixel value, and Z_{min} and Z_{max} are the lower and upper bound of the pixel value. The idea behind weight function $w(z)$ is that the smoothness of the obtained camera response curve is possible to be enhanced and the pixel values related with the middle range of the response curve are well fitted.

Due to quadratic nature of the objective function (2.2), the minimization process is to solve a least square problem, in which the solution is robustly given by Singular Value Decomposition (SVD) method [5]. As mentioned by Debevec, N (the number of pixels selected from each digital image) and P (the number of used digital images) have to be suitably selected for ensuring both an over-determined system and computational efficiency since SVD is applied.

When $g(z)$ (log inverse function of the camera response curve) is obtained by solving the over-determined system, the HDR radiance map (HDR image) can be reconstructed using digital images captured with varying but known exposure time [5]. Debevec suggested that all the pixel values captured under varying exposure should be used to recover the HDR radiance map [5] as

$$\ln E_i = \frac{\sum_{j=1}^{P} w(Z_{ij})(g(Z_{ij}) - \ln \Delta t_j)}{\sum_{j=1}^{P} w(Z_{ij})}, \tag{2.4}$$

where E_i denotes the radiance value at location i, $w(z)$ is the weight function used for putting emphasis on the pixel values in the middle range of the inverse camera response curve. Note that once the camera response curve is computed by (2.3), it is possible to use

the response curve for scene radiance recovery task of other scenes. When a combination of digital images with varying exposure is used for recovering the HDR scene radiance, the noise in the scene radiance value also can be reduced [5].

An example for recovering the HDR scene is given using two sets of images with multiple exposure. Figure 2.1 shows the set of the digital images with multiple exposure for computing the camera response curve. The color of the image set is well balanced, which ensures more accurate recovery of the response curve. The digital camera used for capturing the images is Canon^{TM} EOS 300D, and the response curves for all the color channels are recovered separately. The recovered response curves for the RGB color channels are depicted in Figure 2.2. The result shows that the three color channels share very close response properties.

The reconstruction of the HDR scene radiance is presented using the image set in Figure 2.3. The selected scene is a "corner" with a window located in the middle of scene. When the exposure of one single digital image is adjusted for the window region, the scenery outside the windows is well observed while the indoor region has a dark appearance. On the contrary, if the exposure is adjusted to the indoor region, the scenery outside the window is missing due saturation caused by over-exposure. As is mentioned, the different sets of images are used for the camera response curve recovery process and the reconstruction process of HDR scene radiance (HDR image) respectively.

The reconstructed scene radiance map of the "corner" scene (Figure 2.3) is shown in Figure 2.4 in log scale for its radiance values. The window region has the highest radiance level inside the scene, which mainly due to the bright daytime luminance condition outside. Note that one strip region on the wall also shows comparatively high radiance value, which is caused by the reflectance of light from the window region. McCann [23][27] argued that the dynamic range captured by both film camera and digital camera are limited by veiling glare, which is related with both optical architecture of camera and specific scene. The recovered HDR scene radiance map itself has potential applications, such as providing better HDR texture for rendering tasks on graphics hardware [28], working as luminance data acquisition system [29], and capturing the HDR scene with optimized noise [30].

The reconstructed HDR scene radiance map, whether grayscale or color (multi-channel), is usually unsuitable for being displayed on conventional displays. Although special hardware [31] like a combination of Digital Light Processing (DLP) projector and Liquid

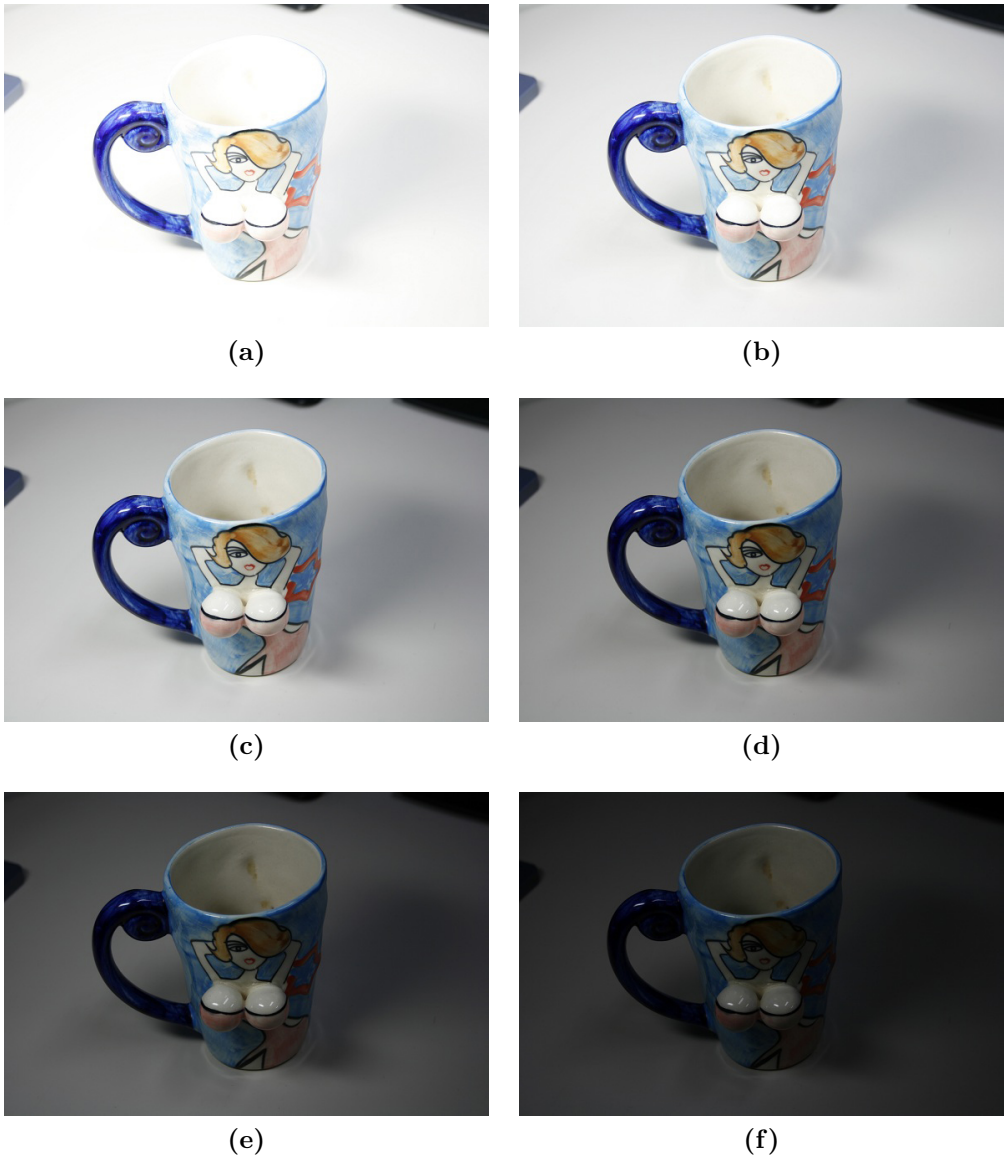


Figure 2.1: *A "cup" scene used to recover the camera response curve (Canon^{TM} EOS 300D).*

Crystal Display (LCD) panel has been proposed for reproducing the HDR scene radiance map. Some certain dynamic range compression methods are necessary for viewing the obtained HDR scene radiance map on conventional displays. Note that some recent LCD displays are equipped with Light-Emitting Diode (LED) backlights, which has the potential for displaying HDR scene radiance maps in nature.

2.2 Tone Mapping for High Dynamic Range Images

For viewing HDR images (HDR scene radiance map) on conventional displays, the dynamic range of HDR images has to be compressed. By carrying out tone mapping on

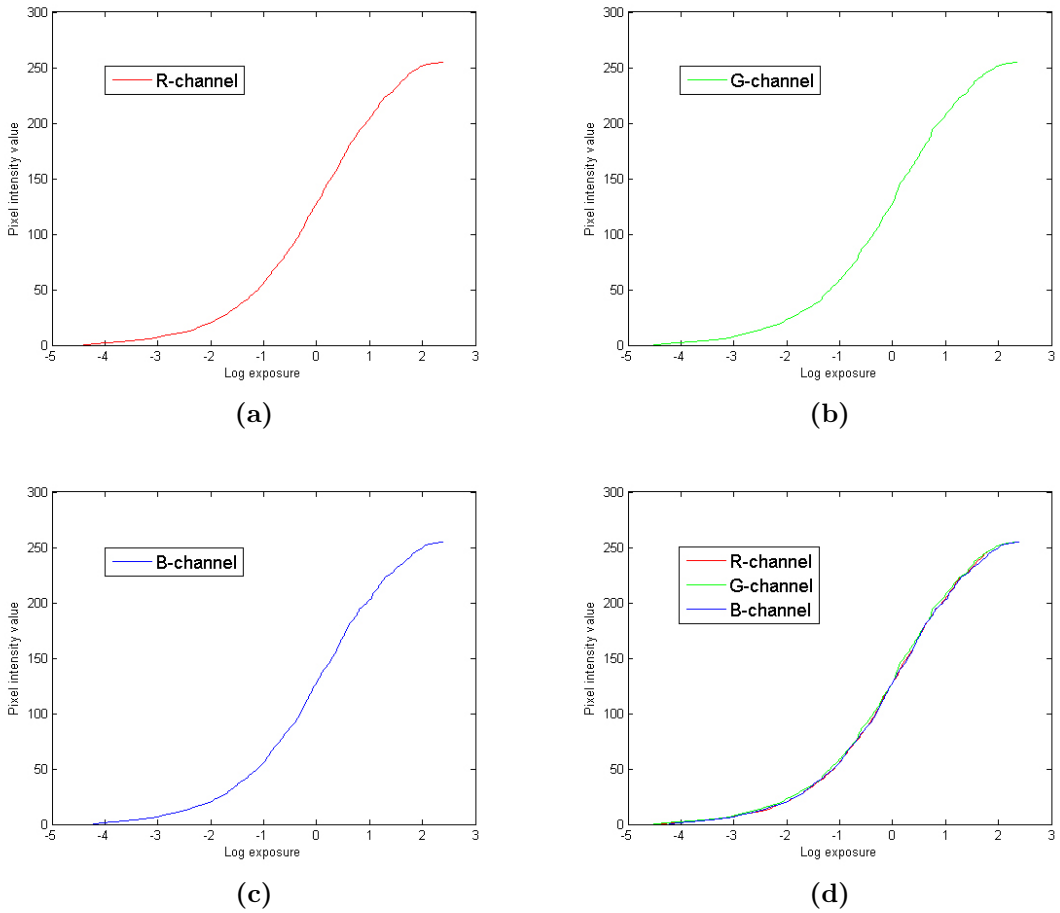


Figure 2.2: *The recovered camera response curve using the image set shown in Figure 2.1. (a) The camera response curve for red channel. (b) The camera response curve for green channel. (c) The camera response curve for blue channel.*

HDR image, the contrast of the scene radiance is greatly reduced so that it matches the physical limit of dynamic range on the medium (e.g. printed paper and LCD display). When designing the algorithms of tone mapping, the most important prerequisite is that the tone mapped HDR image provides a similar visual effect as the real HDR scene.

Land and McCann carried out thorough research on human's perception of color [32]. They revealed that the sensation of color is closely related with object reflectance and developed the Retinex theory. Based on the Retinex theory, researches have been carried on rendering of HDR images. The patent by Frankle and McCann shows the detailed solution for applying the Retinex theory on HDR image and the way for representing a scene on medium with physically limited dynamic range [33]. Meylan et al. [34] proposed a Retinex-based adaptive filter for rendering HDR images, of which the shape of the filter follows the high contrast edge of the image. One of the merits for apply such an adaptive

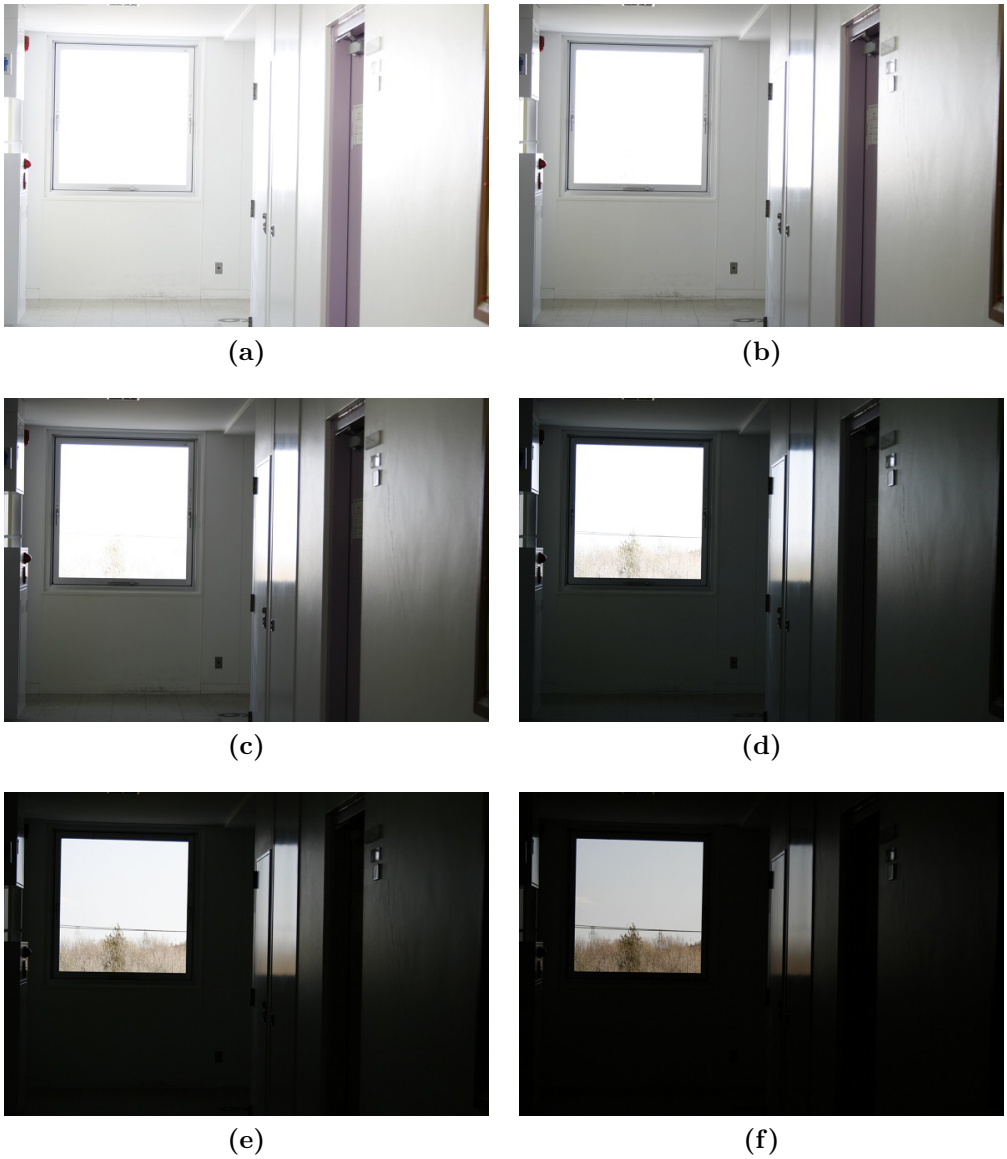
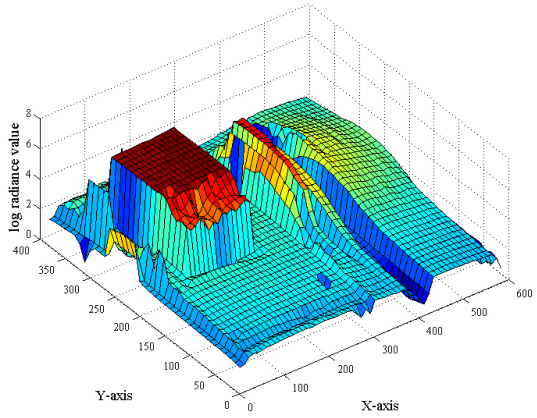


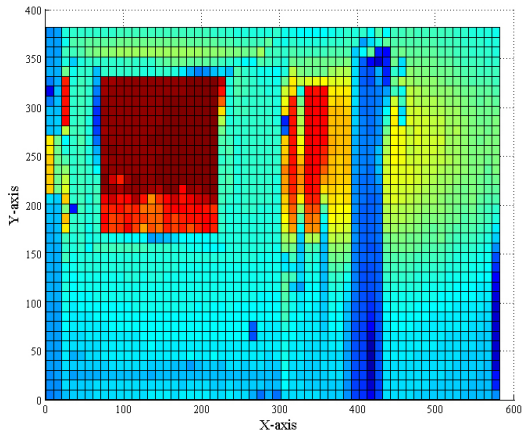
Figure 2.3: *A "corner" scene with a window used to reconstruct the HDR radiance map (HDR image).*

filter is that halo artifact is possible to be reduced. Zhang et al. [4] addressed the issue of asymmetric illumination estimation caused by sequential translation operation in various directions.

In an HDR image, the gradient on locations between bright and dark region shows comparatively large magnitude. Based on this observation, Fattal et al. [8] proposed a rendering method for HDR image on the gradient domain. The magnitude of large gradient on luminance image is compressed, while the small gradient is preserved. By solving Poisson equation on the adjusted gradient field, a new luminance image is obtained [8].



(a)



(b)

Figure 2.4: *The reconstructed HDR radiance map (HDR image) in log scale using the image set shown in Figure 2.3.*

Fast bilateral filter [6] is a widely accepted rendering approach for HDR images. As argued by the proposers of this method (Durand et al.), the major characteristic of the fast bilateral filter is that the contrast of HDR image is reduced while the local details are preserved. The fast bilateral filter separates the HDR images into two layers, a base layer and a detail layer (usually in log scale). The base layer is possible to be compressed for reducing the dynamic range of HDR image, while the detail layer is unchanged for the purpose of detail preservation. A bilateral filter is formed by multiplying a 2-D spatial Gaussian function with a 1-D intensity influence function, which causes the non-linearity of the filter [6]. An example for applying the fast bilateral filtering on an HDR image is given using the corner scene shown in Figure 2.3. The base layer and the detail layer divided in log scale by the fast bilateral filtering is shown in Figure 2.5. And the radiance

map of the compressed HDR image is shown in Figure 2.6 together with the original scene radiance map. in Figure 2.7, the base layer is shown in linear scale before and after compression together with the detail layer. And the final rendering result of the HDR image is given in Figure 2.8.

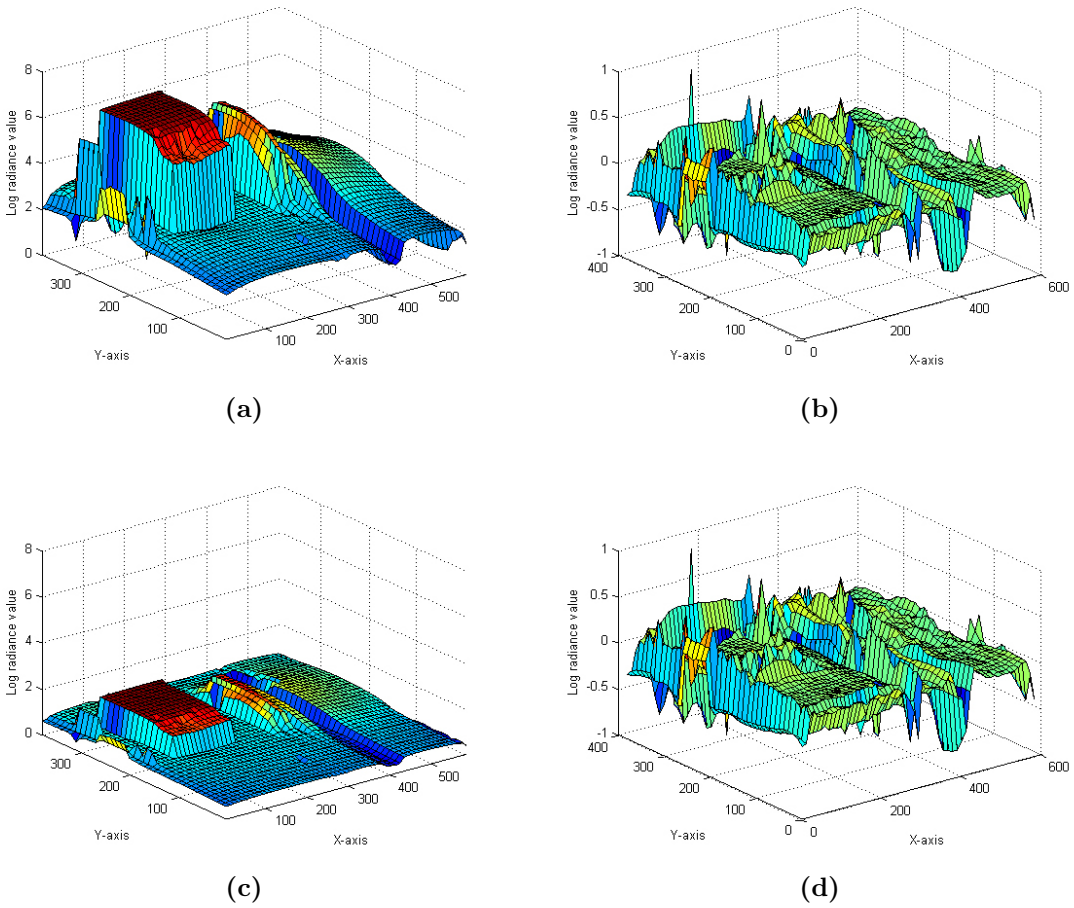
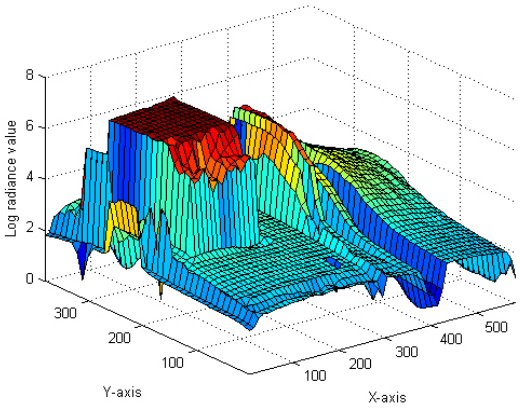


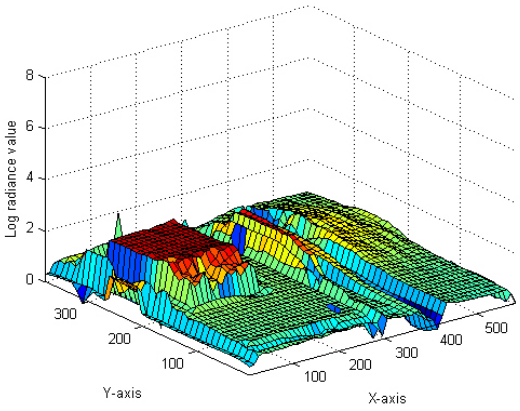
Figure 2.5: *The log-scaled base layer and the log-scaled detail layer of the HDR scene radiance map in Figure 2.4 calculated by fast bilateral filter. (a) The base layer of Figure 2.4. (b) The detail layer of Figure 2.4. (c) The compressed base layer of Figure 2.4 by a factor of 3.5. (d) The unchanged detail layer of Figure 2.4.*

2.3 Hardware for High Dynamic Range Imaging

Motivated by the growing need for recording varying HDR scenes, hardware have been developed. One type of hardware follows the methodology proposed by Debevec [5], namely composting an HDR image with a set of digital images photographed under different exposure durations [19][20]. The other type of hardware uses specially designed image sensors which intrinsically captures scene radiance with a wider dynamic range [35].



(a)



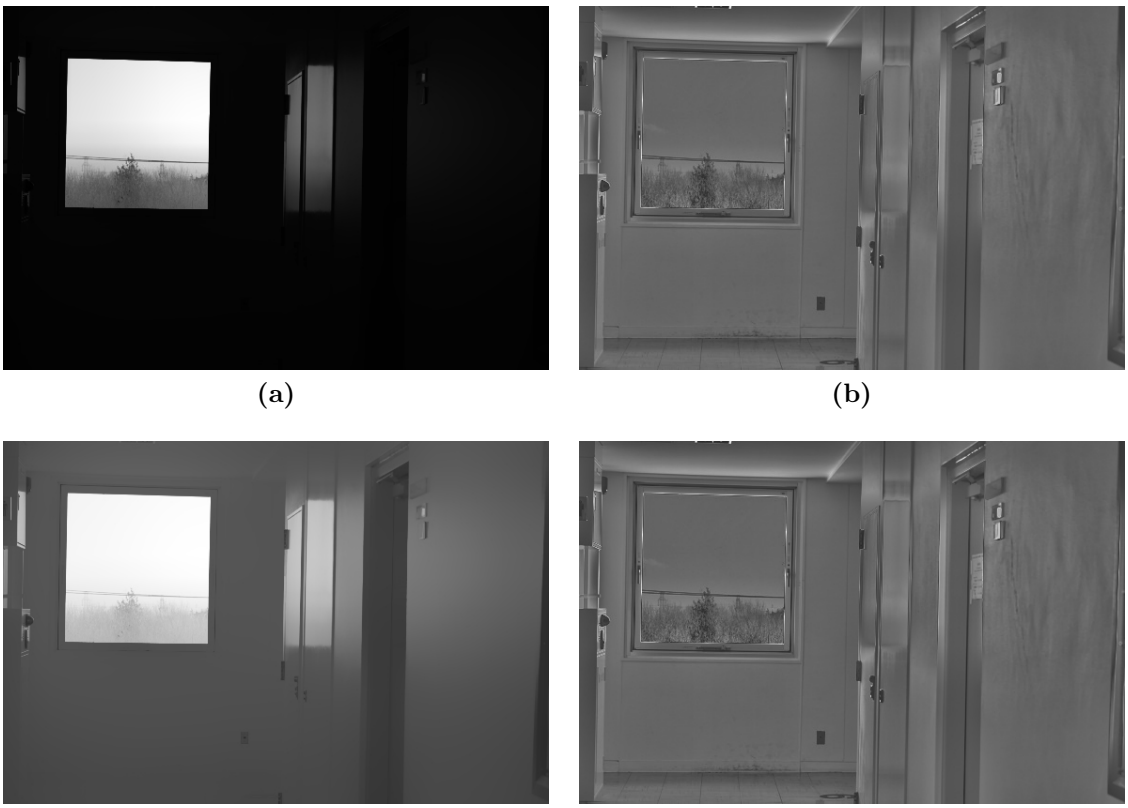
(b)

Figure 2.6: *The original HDR scene radiance map (Figure 2.4) in log scale compared with the one whose base layer is compressed by a factor of 3.5. (a) The original scene radiance map shown in Figure 2.4. (b) The scene radiance map whose base layer is compressed using fast bilateral filter by a factor of 3.5.*

Kang et al. [19] developed a method for capturing HDR video with iterating short and long shutter durations. Due to the moving nature of the selected driving scene, registration becomes a critical issue on this work.

Tocci et al. [20] designed an optical architecture for capturing HDR scene. The beam splitters in the architecture allow the capture of the HDR scene with multiple exposures at the same time, which ensures an easy registration process. The light efficiency of this architecture is improved by reusing the optical path. At least three imaging sensors are necessary for constructing the optical architecture.

For biomedical applications in which the imaging sensor is designed to mimic the response of human vision system, CMOS image sensors with logarithmic response property are designed for dealing with the dynamic range of daily scenes [35].



(a) **(b)**

(c) **(d)**

Figure 2.7: *The linearly-scaled base layer before and after compression together with the linearly-scaled detail layer of the HDR scene radiance map in Figure 2.4 calculated by fast bilateral filter. (a) The uncompressed base layer of Figure 2.4 in linear scale. (b) The detail layer of Figure 2.4 in linear scale. (c) The compressed base layer (by a factor of 3.5 in log scale) of Figure 2.4 in linear scale. (d) The unchanged detail layer of Figure 2.4 in linear scale.*

Figure 2.8: *The tone-mapped HDR image by fast bilateral filter (the "corner" scene shown in Figure 2.3).*

2.4 Chapter Summary

Methods for recording and reproducing HDR scenes have been studied for hundreds of years by painters, photographers and computer vision researchers. HDR image is usually reconstructed from a set of images with different exposure durations. There are two steps for HDR image reconstruction, calculating camera response curve and compositing HDR image with image set of multiple exposures. Tone mapping methods (e.g. Retinex model, gradient domain compression, and fast bilateral filter) are well-designed tools for displaying HDR images on mediums with physically limited dynamic range. Under different scenarios, researchers have designed varying types of specialized hardware for capturing targeted HDR scenes.

Chapter 3

Non-Photorealistic Rendering for High Dynamic Range Scenes

In contrast to the conventional quest for realism in the field of rendering, a new quest is growing for creating imagery that is more artistic and easy for communication [1]. Rendering methods developed for enabling a variety of expressive styles are referred to as NPR. Typical NPR styles are usually inspired by artistic works, e.g., oil painting [2] and pen-and-ink illustration [3]. In NPR tasks, three dimensional models, vector fields, or simply digital images serve as input information. Among these types of input data, digital images of real scenes are usually utilized as references for generating NPR results [1]. The limited dynamic range of normal digital images, however, imposes restrictions on the performance of some NPR methods. With the development of image processing technologies, HDR images have attracted wide attention [4][5][6][7][8]. Limitations on the performance of NPR methods resulting from low dynamic range of digital images are overcome when HDR images are incorporated into the NPR process. Some researchers have shown an interest in developing NPR methods for HDR images. Colbert et al., for example, proposed two kinds of brushes for virtual painting on unmapped HDR imagery [9]. Vergne et al. proposed a pipeline for the stylized rendering of HDR environment-mapped objects [10]. Smith et al. showed that HDR images provide better results in edge detection and segmentation in NPR [11]. Accordingly, attention is paid to gradient values of HDR images in developing the NPR method.

The proposal mainly focuses on blending HDR images into conventional stroke-based NPR methods such as the painterly rendering proposed by Hertzmann [2] and the decorative mosaic rendering developed by Hausner [12]. The achromatic brightness (luminance)

channel of an HDR image is used to compute gradient values for controlling the placement of elements such as strokes in the rendering process. This modification in the stroke-based rendering process [2] is presumed to give rise to the occurrence of more accurate strokes, which is rewarding in the painterly rendering style. For an NPR process, rendering results may be greatly affected by the color information used. In the proposal, a tone-mapped HDR image or a well-exposed digital image of a scene is chosen to provide color information for the rendering process, of which the choice depends on the scene to be rendered. These changes in the information source of the rendering process also enable detailed structures and colors to be preserved in the resulting images, especially for regions that present particularly high or low luminance.

In order to evaluate the proposal, two experiments are carried out, and the desktop computer used for the experiments has a 2.66GHz CPU and 2GB memory. In the first experiment, averaged angular errors in gradient computation for both a set of generated digital images and a constructed HDR image are figured out by simulating a photo capture process in which the radiance map of a virtual scene is a superposition of three Gaussian functions whose gradient values are easily obtainable in the closed form. An image dataset composed of three scenes is created using a digital single-lens reflex camera for the second experiment. For each scene in the image dataset, a set of images is photographed at different shutter timings. In the second experiment, intuitive rendering results are given to show visual improvements when the proposal is applied to typical NPR styles, e.g., painterly rendering [2] and decorative mosaic rendering [12]. A bottom-up visual attention model is then applied to evaluate these rendering results. The attended regions in rendering results show the applicability of the proposed method for scenes with rapidly changing light conditions.

3.1 Gradient-related Non-Photorealistic Rendering

Painterly rendering [2] is one typical NPR method for which a rendering result is shown in Figure 3.1. The rendering process for this style imitates the placement of strokes on the canvas as is done by skilled human painters. The path for guiding the placement of each stroke consists of a set of control points. Relative positions for sequential control points are determined by gradient values of the realistic digital image [2]. This methodology

for computing this guidance path of curved strokes is inherited, but images used for computing gradient are replaced by the radiance maps of HDR images, which is due to the improved accuracy of gradient value computation.

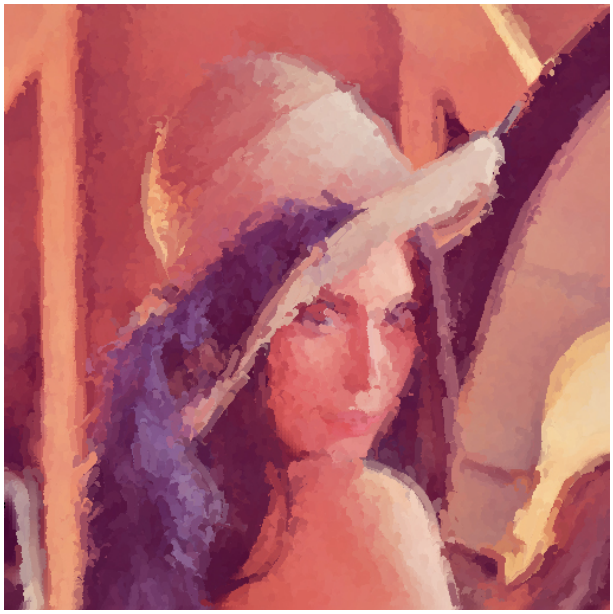


Figure 3.1: *Rendering result of Lena in a painterly style.*

An intuitive explanation for the improved precision for the gradient computation of HDR images is that the gap between discrete gray levels in conventional digital images are made smooth if an HDR image is used for computing the gradient. The quantitative evaluation of the accuracy of gradient computation is given in Section 3.2. The decorative mosaic style is another well-known NPR method [12], of which the rendering result of a conventional Lena image is shown in Figure 3.2.

In the rendering process of the decorative mosaic method, the orientation of the elements (mosaics) is determined by a predefined vector field which is replaced by the gradient value of the HDR images. The orientation of these mosaics in the rendering result also implies structural information of the scene. Improved precision in gradient computation using HDR images thus also improves the structure the rendering result. One key idea behind the proposal is that the gradient value computed using HDR images improves rendering results by more precise placement of the elements for typical NPR methods such as painterly rendering [2] and decorative mosaic rendering [12].

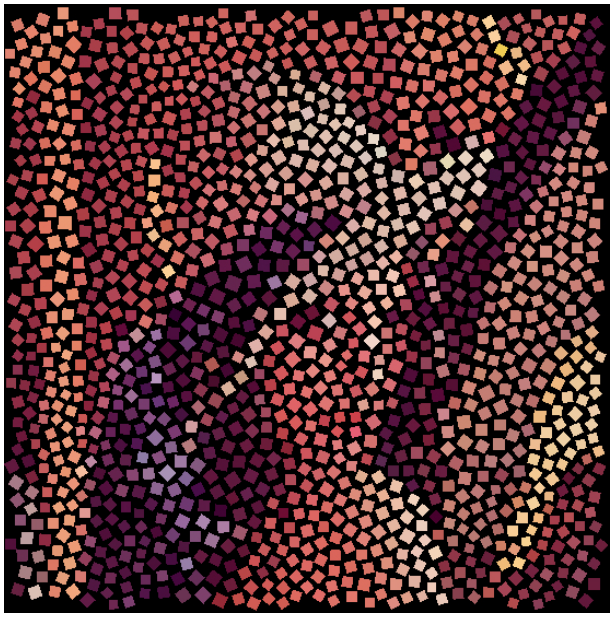


Figure 3.2: *Rendering result of Lena in a decorative mosaic style.*

3.2 Gradient of High Dynamic Range Images in Non-Photorealistic Rendering

It is usually difficult to quantitatively evaluate the performance of NPR algorithms [36]. As is mentioned in Section 3.1., gradient values play a crucial role in guiding the placement of strokes in the painterly rendering process [2] and mosaics in the decorative mosaic style [12]. So, instead of rendered images being directly evaluated, errors occurring in the gradient computation of HDR images are estimated.

It is difficult to obtain a precise radiance map for any real scene, so the precise gradient value of the radiance map is also difficult to obtain, which becomes an obstacle in evaluating the accuracy of the gradient computation of HDR images. So, instead of using digital images shot in a real scene, a set of virtual images is generated by simulating the imaging process of a typical digital camera. The light intensity (radiance map) of the virtual scene $I_F(x, y)$ is given in its closed form. The only tunable parameter in the simulation process is the shutter timing. Using a set of predefined shutter timings, virtual image set $I = \{I_1(m, n), ..., I_k(m, n)\}$ is generated. HDR image $I_{HDR}(m, n)$ is constructed from generated image set I using the reconstruction method proposed by Debevec et al. [5].

The angle of the precise gradient value in fixed location (m, n) of the virtual scene is

computed using

$$\theta_F(m,n) = tan^{-}1\left(\frac{\partial I_F(x,y)}{\partial y}\Big/\frac{\partial I_F(x,y)}{\partial x}\right). \tag{3.1}$$

In order to compute the angle of the gradient value at location (m,n) in an HDR image, a Sobel operator is utilized in accordance with the gradient computation method used in [2]. The derivative value of $I_{HDR}(m,n)$ in the horizontal direction is computed as

$$Gx_{HDR}(m,n) = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} I_{HDR}(m,n). \tag{3.2}$$

Similarly, the derivative value in the vertical direction is calculated using

$$Gy_{HDR}(m,n) = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} I_{HDR}(m,n). \tag{3.3}$$

The formula used for computing the angle of the gradient value of $I_{HDR}(m,n)$ is then given as

$$\theta_{HDR}(m,n) = tan^{-1}\left(\frac{Gy_{HDR}(m,n)}{Gx_{HDR}(m,n)}\right). \tag{3.4}$$

The gradient value of any generated image $I_{NOR}(m,n)$ in image set I is computed as

$$\theta_{NOR}(m,n) = tan^{-1}\left(\frac{Gy_{NOR}(m,n)}{Gx_{NOR}(m,n)}\right). \tag{3.5}$$

Then, error occurring in computing the gradient value of $I_{HDR}(m,n)$ is measured by determining the averaged absolute difference between θ_{HDR} and $\theta_F(m,n)$ as

$$\overline{\Delta\theta_{HDR}} = \frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M}\left|\theta_F(m,n) - \theta_{HDR}(m,n)\right|, \tag{3.6}$$

where M and N denote the number of pixels in each row and column for the constructed HDR image. To evaluate error in the gradient computation of $I_{NOR}(m,n)$, averaged absolute difference $\overline{\Delta\theta_{NOR}}$ is calculated in the same way as $\overline{\Delta\theta_{HDR}}$. More details of this evaluation process are given in Section 3.3.

The flowchart for the proposed rendering process of HDR images is given in Figure 3.3.

The basic structure of this method comes from the painterly rendering method proposed by Hertzmann [2].

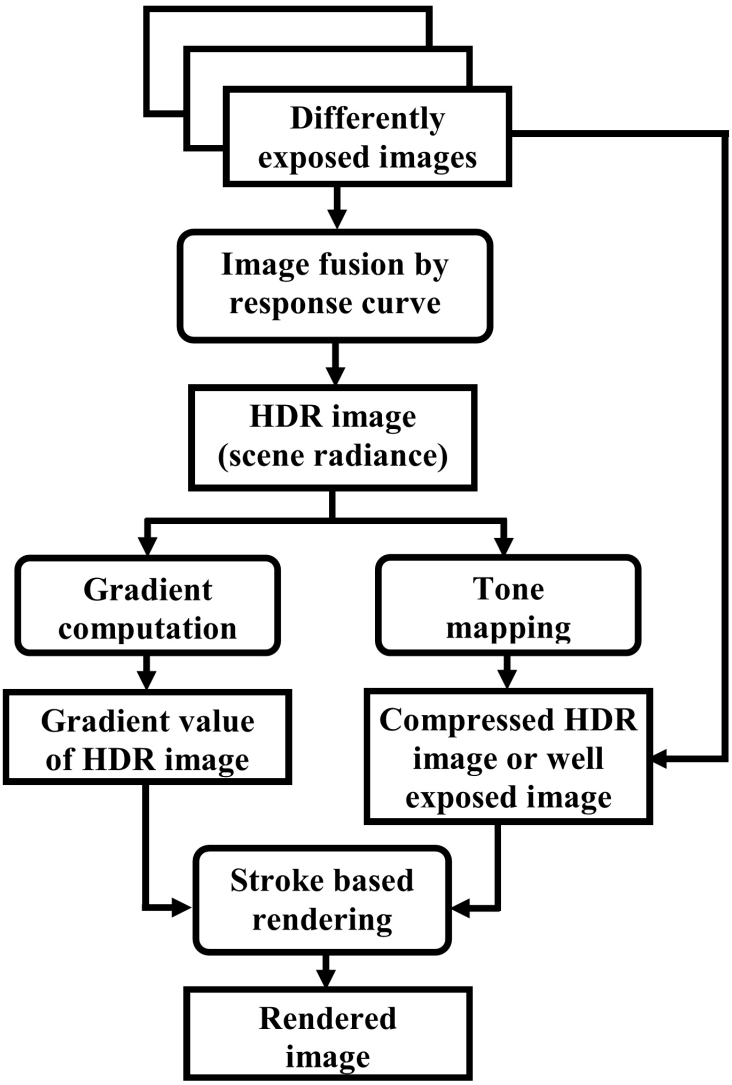


Figure 3.3: *Rendering process for HDR images.*

The first step is to construct an HDR image through an image fusion process using the method proposed by Debevec et al. [5]. Gradient values and a tone-mapped image of the constructed HDR image are then computed separately. Gradient values of the HDR image are calculated to guide the placement of elements in the rendering process. The tone-mapped image is used to provide color information for rendering. Well established tone mapping algorithms [6][8] are utilized to compress the HDR image. Note that a well-exposed image of a scene can also be used to provide color information instead of a tone-

mapped HDR image depending on the illumination of the scene. Specifically, tone-mapped HDR images are good providers of color information in the NPR process for scenes having greatly changing luminance in different regions. For scenes with comparatively uniform luminance over all regions, both well-exposed images and tone-mapped HDR images can be used to provide color information based on the preference of the user. Combining gradient values and color information from the tone-mapped image, a rendered image is achieved by following stroke-based NPR processes, e.g., painterly rendering [2] and decorative mosaic [12].

3.3 Experiments on Gradient Evaluation and Rendering Results

3.3.1 Assessment by Error in Gradient Computation of HDR Images

Gradient values of HDR images play an important role in guiding the placement of elements in the proposed NPR method. In the first experiment, errors in computing the gradient of HDR images are estimated.

Images used for constructing the HDR image are generated by simulating the imaging process of a digital camera as mentioned in Section 3.2. Scene radiance function $I_F(x, y)$ used for generating virtual images is the superposition of three two-dimensional Gaussian functions

$$I_F(x, y) = \alpha_1 g_1(x, y) + \alpha_2 g_2(x, y) + \alpha_3 g_3(x, y), \tag{3.7}$$

where α_1, α_2, and α_3 are scaling factors for each of the three Gaussian functions, for which the peaks and standard deviations are listed in Table 3.1. Scene radiance function $I_F(x, y)$ is plotted in Figure 3.4. One of the advantages for using Gaussian functions is that the closed form expression for their gradient values is obtained easily after applying a transformation.

In this experiment, a set of virtual images is shot by simulating the process of digital image capturing. Shutter time is tuned for generating individual photographs. The

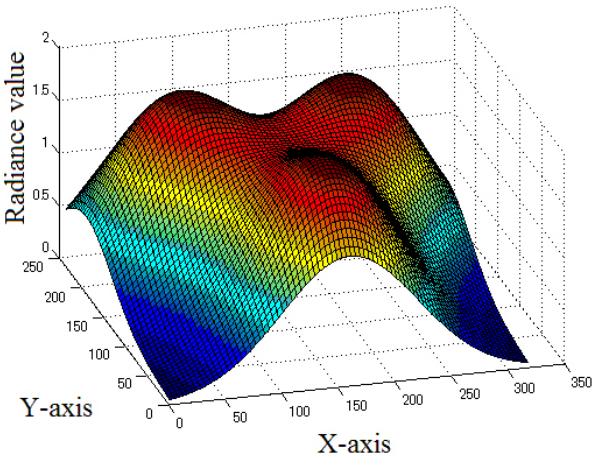


Figure 3.4: *A 3D plot for the scene radiance function $I_F(x, y)$ for which each value is given by (3.7).*

Table 3.1: *Parameters for scene radiance function.*

Gaussian Function	Gaussian Peak	Standard Deviation
g_1	$(160, 50)$	60.0
g_2	$(80, 190)$	60.0
g_3	$(240, 190)$	60.0

response function for the imaging sensor in this simulation is given as

$$f(ex) = \begin{cases} 255, & \text{if } ex \geq 3.0 \\ round(37.13 * ln(1 + 10ex) + 127.5), & \text{if } 0 \leq ex < 3.0 \\ round(127.5 - 37.13 * ln(1 - 10ex)), & \text{if } -3.0 \leq ex < 0 \\ 0, & \text{if } ex < -3.0 \end{cases}, \tag{3.8}$$

where ex is log exposure of the radiance value. A figure for the response curve used in this experiment is demonstrated in Figure 3.5.

Shutter time values used in this experiment are listed in Table 3.2. The virtual images generated are quantized in 8 bits for each pixel, which follows the most widely used 256 intensity (gray) levels for conventional digital images. Samples from the generated image set I are given in Figure 3.6 along with shutter times. An HDR image is then

Table 3.2: *Shutter time used in generating virtual images.*

Image No.	1	2	3	4	5	6	7
Shutter time (s)	0.6	0.8	1.0	1.2	1.5	2.0	2.5

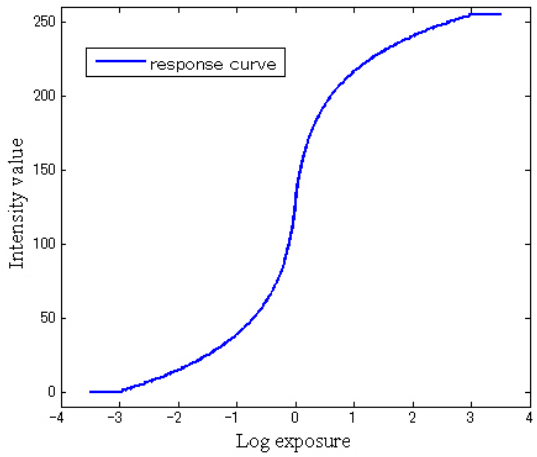


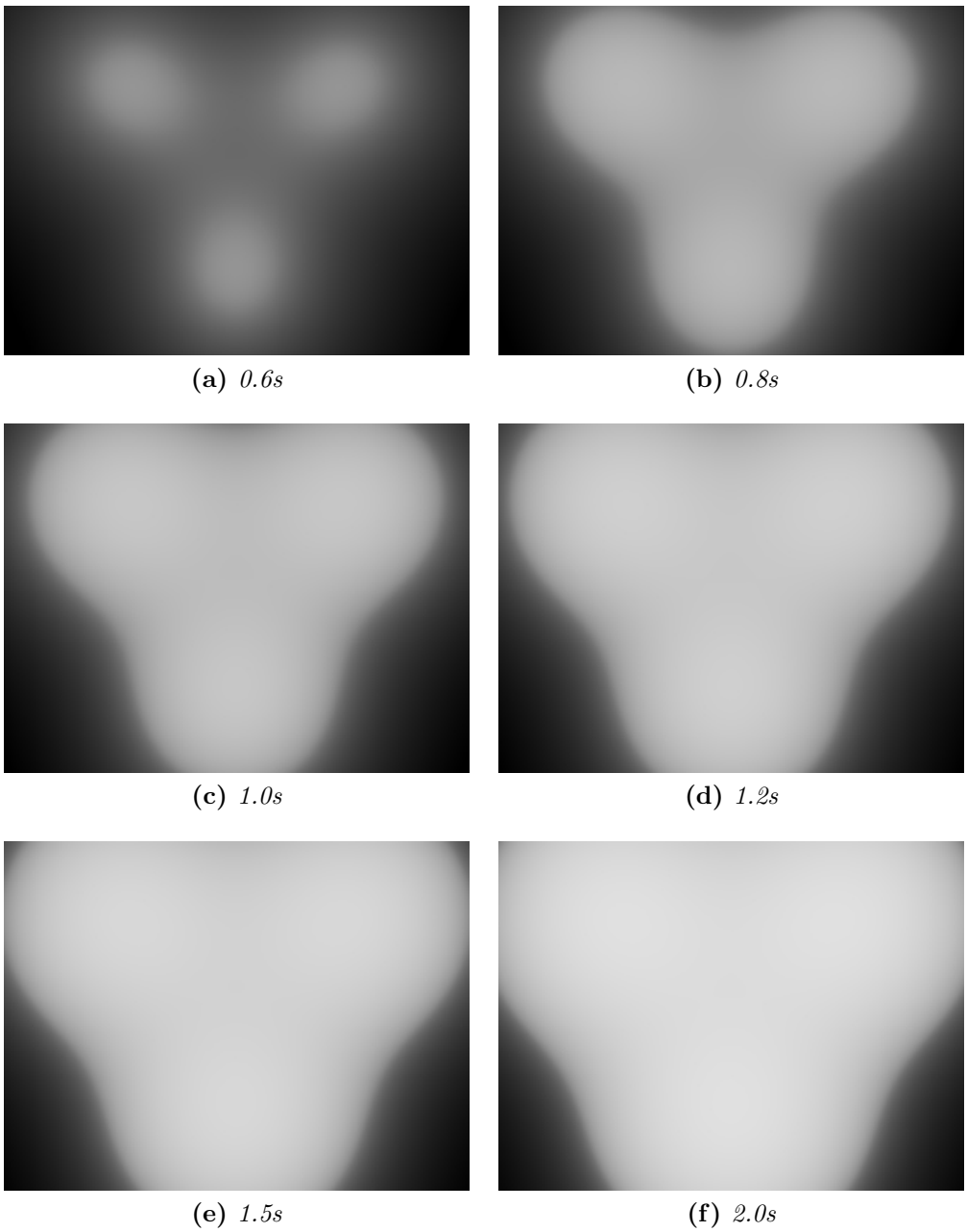
Figure 3.5: *Response curve for generating the virtual digital images.*

constructed from this set of differently exposed images using the response curve obtained by minimizing objective function (2.2) [5]. The obtained response curve shows a consistent shape compared to the predefined virtual response curve shown in Figure 3.5. After an HDR image is constructed from the generated image set I using (2.4), $\theta_F(m,n)$, $\theta_{NOR}(m,n)$, and $\theta_{HDR}(m,n)$ are determined using (3.1), (3.4), and (3.5) as shown in Figure 3.7. To avoid numerical errors in the division operation, a small additional value (0.00001) is added to the divider in gradient computation.

The angular map of gradient values for the scene radiance function shown in Figure 3.7(a) appears very smooth because the closed form expression of the gradient values is available using 3.1. The angular map of gradient values of the constructed HDR image shown in Figure 3.7(b) has a shape similar to that of the scene radiance function, which improves accuracy in gradient computation.

The angular map of gradient values of a generated virtual 8-bit image shown in Figure 3.7(c) has a coarse surface, mainly due to limited quantization levels used in the simulated image capture process. Even though the angular map in Figure 3.7(c) has a coarse shape, this is the best result among the angular maps obtained from the generated virtual images shown in Figure 3.6, which is mainly due to an appropriate choice of shutter time, namely $0.6s$, in terms of the radiance function and the assumed virtual response curve for this experiment.

For each pixel, $\Delta\theta_{HDR}(m,n) = |\theta_F(m,n) - \theta_{HDR}(m,n)|$ and $\Delta\theta_{NOR}(m,n) = |\theta_F(m,n) - \theta_{NOR}(m,n)|$ are computed to measure error in gradient computation. If an entry of abso-



(a) *0.6s* **(b)** *0.8s*

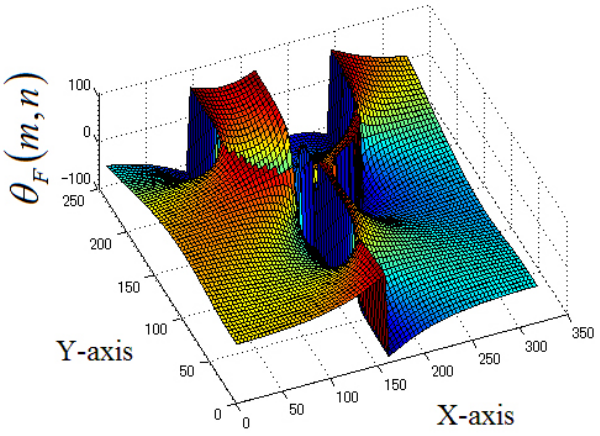
(c) *1.0s* **(d)** *1.2s*

(e) *1.5s* **(f)** *2.0s*

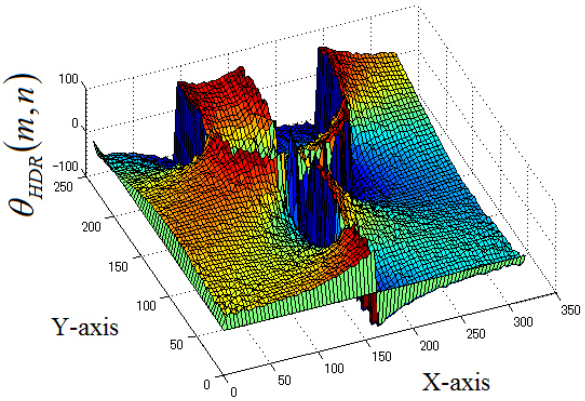
Figure 3.6: *The generated virtual images with their shutter time.*

lute angle difference $\Delta\theta_{HDR}$ or $\Delta\theta_{NOR}$ is larger than 90°, then its value is replaced with $180° - \Delta\theta_{HDR}$ or $180° - \Delta\theta_{NOR}$, which is a reasonable operation due to the arc tangent function used for computing the angle of gradient values. Plots for $\Delta\theta_{HDR}$ and $\Delta\theta_{NOR}$ are shown in Figure 3.8.

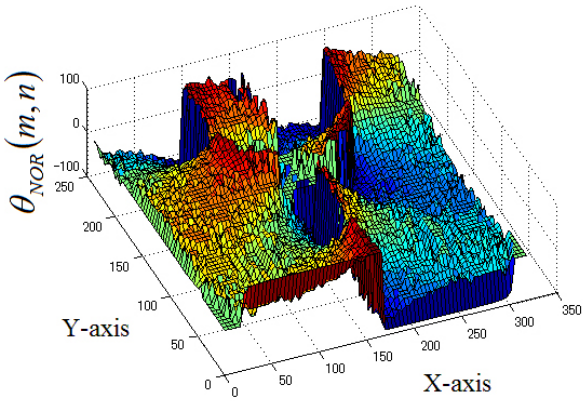
There exist a few locations with values higher than the vast majority due to the fact that the derivative in the horizontal direction approaches zero at these locations, which results in a rapid change in their gradient values. The $\overline{\Delta\theta_{HDR}}$ and $\overline{\Delta\theta_{NOR}}$ are calculated using (3.6) to quantitatively evaluate the performance in terms of the averaged error



(a)



(b)



(c)

Figure 3.7: *Angular maps of gradient values computed using the scene radiance function, the constructed HDR image, and one of the generated virtual images (shutter time:* $0.6s$*). (a)* $\theta_F(m,n)$*. (b)* $\theta_{HDR}(m,n)$*. (c)* $\theta_{NOR}(m,n)$*.*

in gradient estimation of HDR images, for which the results are shown in Figure 3.9. Compared to the lowest value of $\overline{\Delta\theta_{NOR}}$ (shutter time: $0.6s$) in the generated image set

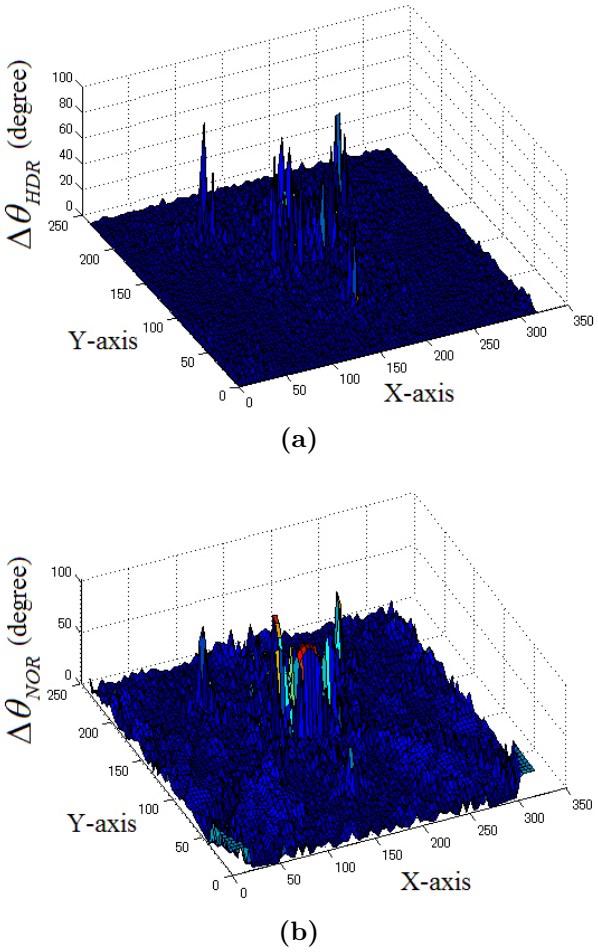


Figure 3.8: *A plot for $\Delta\theta_{HDR}$ and $\Delta\theta_{NOR}$. (a) Errors in computing the gradient value of the constructed HDR image. (b) Errors in computing the gradient values of a generated image (shutter time: 0.6s) from image set I.*

I, $\overline{\Delta\theta_{HDR}}$ achieves a reduction of error by 57.5%.

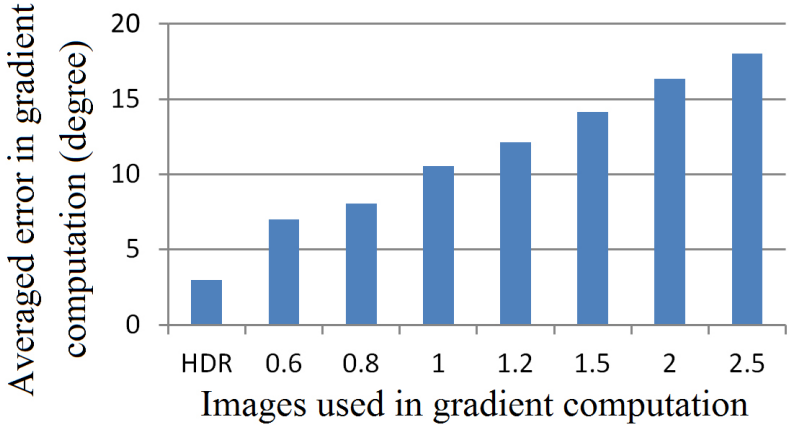


Figure 3.9: *Averaged error in gradient computation of the HDR image ($\overline{\Delta\theta_{HDR}}$) and the generated images ($\overline{\Delta\theta_{NOR}}$). For all of the generated images in image set I, $\overline{\Delta\theta_{NOR}}$ values are computed.*

3.3.2 Rendering Results for HDR Images

In the second experiment, rendering results after applying the proposal to the painterly rendering [2] and the decorative mosaic [12] are demonstrated. Note that when constructing the modules for the final rendering step, the main ideas for both methods [2][12] are followed, but some changes are also made in the programming step.

An image dataset of several different scenes is built as depicted in Figure 3.10. Note that only a portion of the images for each scene is demonstrated. As the scene in the third row of Figure 3.10 shows, the luminance condition changes rapidly inside and outside of the window, and this also results in a high dynamic range. When this image dataset is built, strongly lit or daytime scenes are chosen so that HDR imaging techniques can be applied without disrupting the appearance of such scenes.

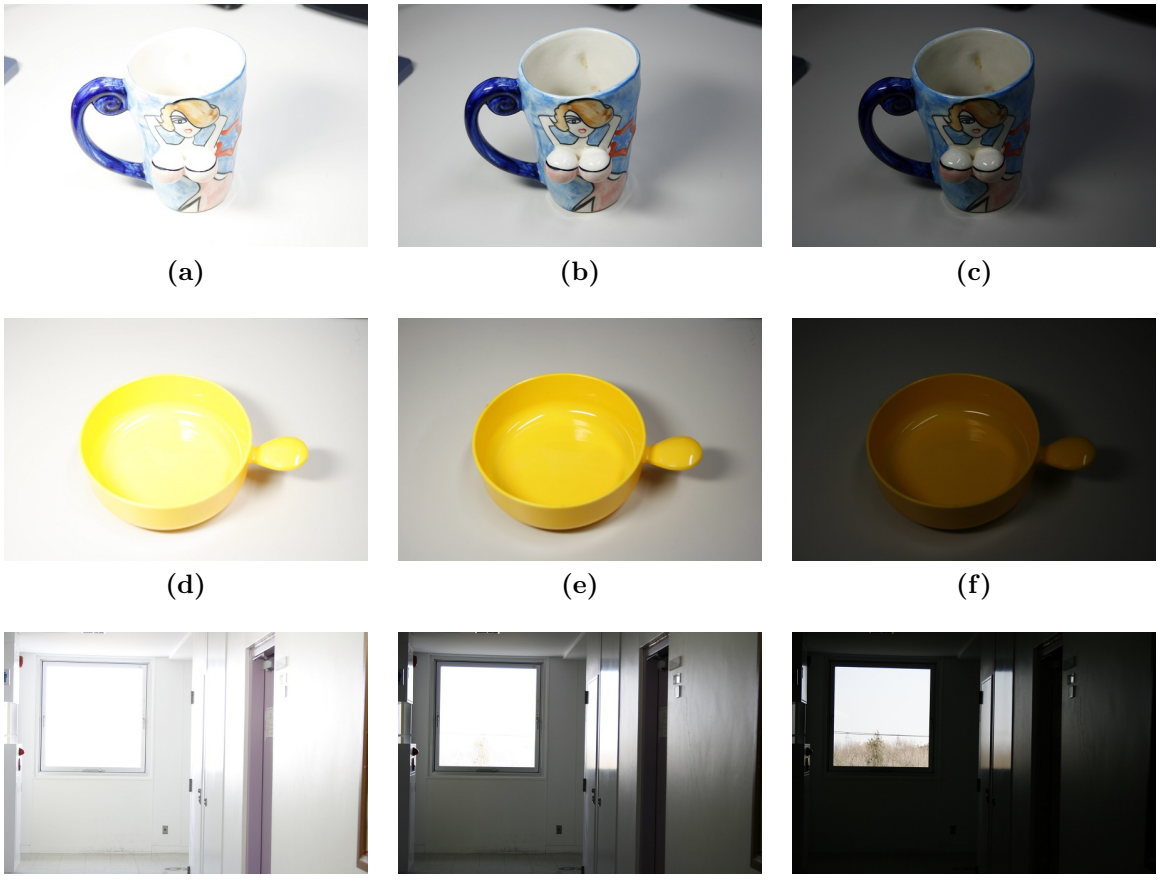


Figure 3.10: *Image sets used for constructing HDR images*

The response curve of the digital camera used in this work is computed using the cup scene (the first row) of the image set partially shown in Figure 3.10, and the obtained response curve is demonstrated in Figure 3.11. The color tone of the cup scene is

well balanced, which means that each color channel of the images accounts for a similar percentage of the scene.

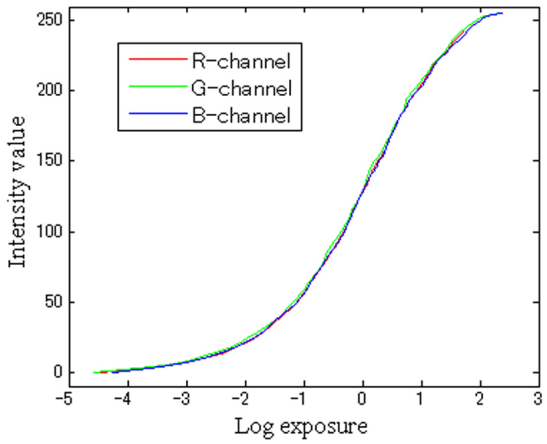


Figure 3.11: *The response curve of the digital camera used in this work.*

As is mentioned in Section 3.2, tone-mapped HDR images are used to provide color information for the rendering process in the proposal. Tone-mapped HDR images of the scenes in Figure 3.10 are demonstrated in Figure 3.12. Note that the total brightness of tone-mapped HDR images shown in Figure 3.12(a) and Figure 3.12(b) are comparatively dark, mainly due to bilateral filtering [6] used to compress the contrast of these HDR images. In this experiment, contrast compression in the tone mapping process is done under the log scale, which is mainly due to the good avoidance of halo effects. The result of the corner scene shown in Figure 3.12(c) has good visual effects both in the region with high luminance (inside the window) and in the region with low luminance (outside the window), which explains the detail preserving property of the proposal well.

Applying the proposal to painterly rendering [2], rendered results for three scenes in Figure 3.10 are depicted in Figure 3.13. For the results in Figure 3.13(a), Figure 3.13(c), and Figure 3.13(e), color information in the rendering process of the HDR images is provided by tone-mapped HDR images using bilateral filtering [6]. For the rendering results in the Figure 3.13(b), Figure 3.13(d), and Figure 3.13(f), conventional 8-bit color digital images provide color information. Comparing the results shown in Figure 3.13(a) and Figure 3.13(b), the holder of the cup rendered from the HDR image demonstrates an improvement in its structure and color. In the region of the cup rabbet, the separation of the cup and the background is clearly depicted in Figure 3.13(a). This good depiction

(a)

(b)

(c)

Figure 3.12: *Tone-mapped HDR images for three scenes of the image sets shown in Figure 3.10.*

of structural and color information of the original scene is also explained by weak scene brightness, but details of this are beyond the scope of this work.

For the scene in Figure 3.13(e) and Figure 3.13(f), rendering results for the HDR image of the scene preserve more details, especially in the window region, which is closer to human vision. The color of Figure 3.13(e) is more vivid than that of Figure 3.13(f).

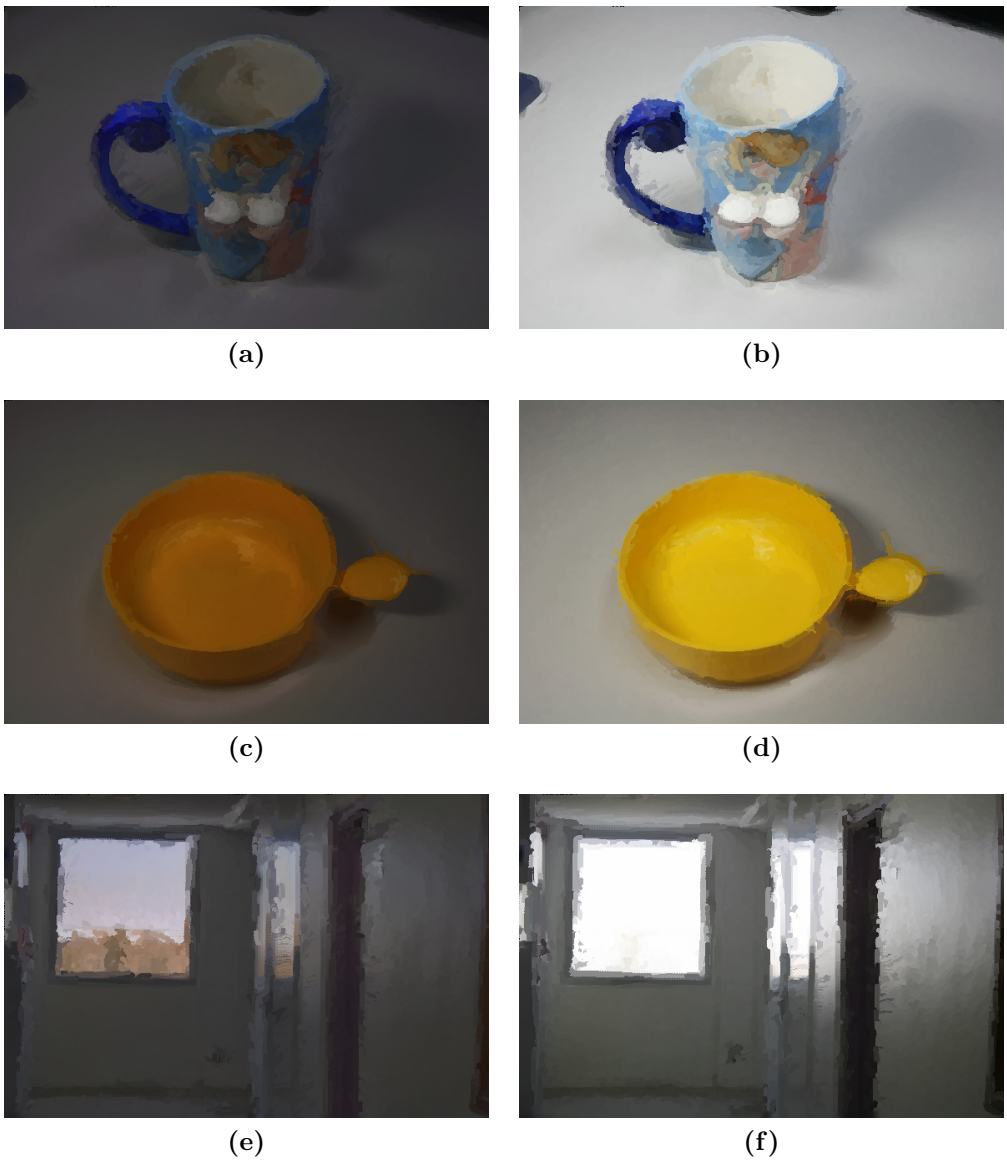


Figure 3.13: *Rendering results of HDR images and normal images using painterly rendering. (a), (c), (e) Rendering results for HDR images of scenes in Figure 3.10. (b), (d), (f) Rendering results for conventional digital images of the scenes.*

Rendering results of the HDR images seem to be darker than those of the conventional 8bit color digital images, which is mainly due to tone mapping methods applied in this experiment.

For the decorative mosaic proposed by Hausner [12], the proposal is applicable to preserve more details in structural and color information of scenes. Results for rendering HDR images of the scenes in Figure 3.10 using the decorative mosaic method are shown in Figure 3.14 compared to those for the conventional digital images.

Orientations of the mosaics in Figure 3.14(a), (c), and (e) are determined using gradient values of the HDR images, and colors of these mosaics are given by tone-mapped

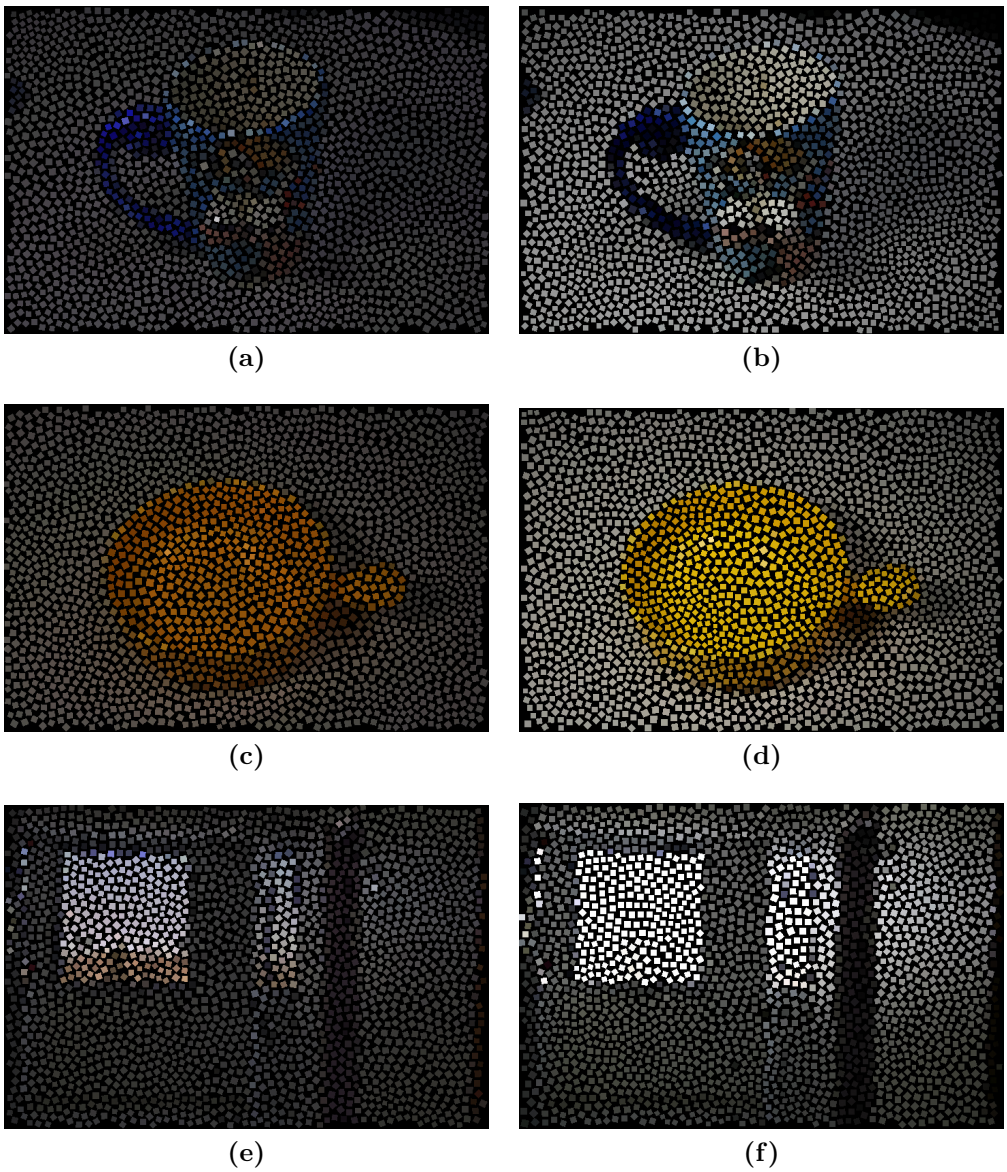


Figure 3.14: *Rendering results for HDR images and conventional digital images using the decorative mosaic style. (a), (c), (e) Rendering results for the HDR images. (b), (d), (f) Rendering results for the conventional digital images.*

HDR images. Orientations and colors of the mosaics in Figure 3.14(b), (d), and (f) are determined using conventional digital images of the scenes. The rendering result for the conventional digital image shown in Figure 3.14(b) has a slightly distorted shape; while the rendering result for the HDR image shown in Figure 3.14(a) preserves the structure of the cartoon cup. The mosaics formed in Figure 3.14(e) convey more color information for the scene than those in Figure 3.14(f), and orientations of mosaics in Figure 3.14(e) are also better arranged for conveying structural information of the scene. For rendering results shown in Figure 3.14(c) and (d), it is difficult to judge which result is better, so the

final performance of the proposal is greatly influenced by the scene to be rendered. If the luminance condition changes rapidly in different regions of the scene, then the proposal is sure to improve the rendering results.

For human observers, it is difficult to set universal standards for assessing rendered images. Instead of letting human observers assess rendering results, an experiment using the bottom-up attention model [37] is implemented to show the most salient regions ("Saliency Toolbox" by Walther & Koch [38][39]) in rendering results for both HDR images and conventional digital images. The first attended region is obtained from the saliency map of each rendering result and shown in Figure 3.15. Note that the saliency map here only shows a plausible model of the bottom-up human visual attention. Different first attended regions are observed in Figure 3.15(a) and (g) for the same cup scene. Analogous results are also obtained from Figure 3.15(c) and (i). The two comparisons show that when HDR images are chosen as the information source for the rendering process, the attention of human observers is guided to newly-formed salient regions of the scene. For the bowl scene, first attended regions of the rendered HDR image (Figure 3.15(b)) and the rendered conventional digital image (Figure 3.15(h)) share almost the same region. An explanation for such results given by bottom-up visual attention model can be that HDR images carry more information for the scenes.

For other NPR styles such as dithering [1], tone-mapped HDR images can also be used as a good information source for the rendering process. The rendering results in the dithering style for both an HDR image and a conventional digital image are shown in Figure 3.16, where the scene used in Figure 3.16 is displayed in the third row in Figure 3.10.

Improvement in element placement and preservation of details are the two main benefits when applying the proposed rendering method for HDR images. Apart from the rendering styles mentioned here, this idea can be further extended to varying NPR methods, which may provide a new choice for those who want to demonstrate a certain scene impressively. Improved accuracy in gradient computation also suggests that HDR images are a good source for other image processing tasks closely related to the gradient values. Evaluation by the visual attention model shows that the proposal is a potential tool for guiding human visual attention to newly formed informative regions in scenes. In scenes having greatly varying light strengths in different regions, the proposal gives

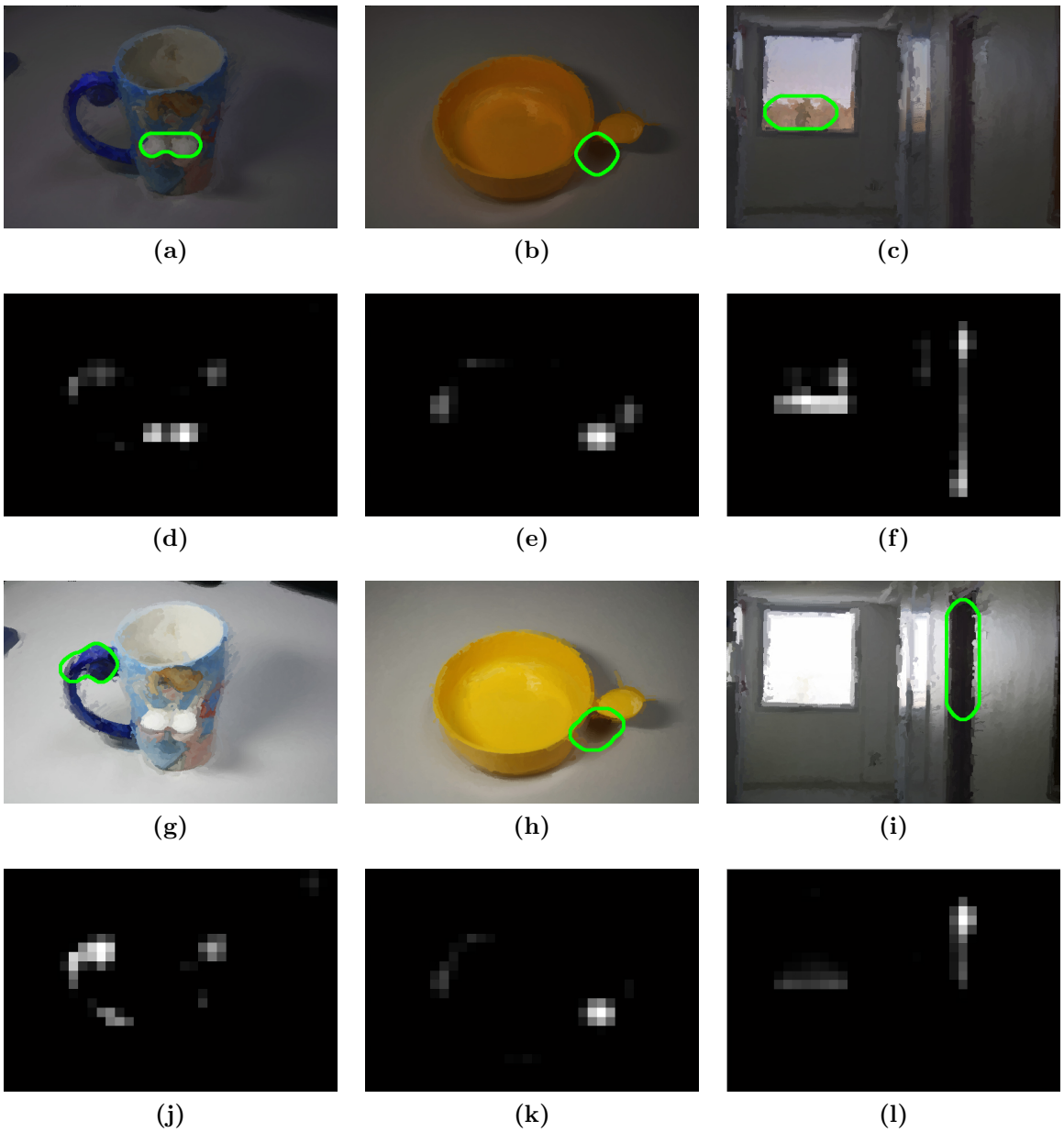
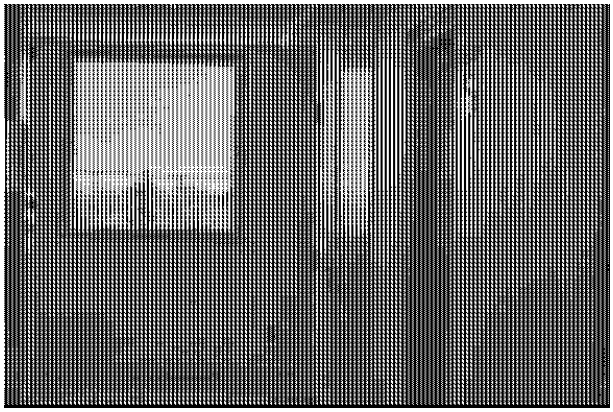


Figure 3.15: *First attended regions (green border) and saliency maps of the rendering results. (a), (b), (c) Rendering results for HDR images with first attended regions. (d), (e), (f) Saliency maps for the rendered HDR images. (g), (h), (i) Rendering results for conventional digital images with first attended regions. (j), (k), (l) Saliency maps for rendered conventional digital images. (results of "SaliencyToolbox" by Walther and Koch [18][19]).*

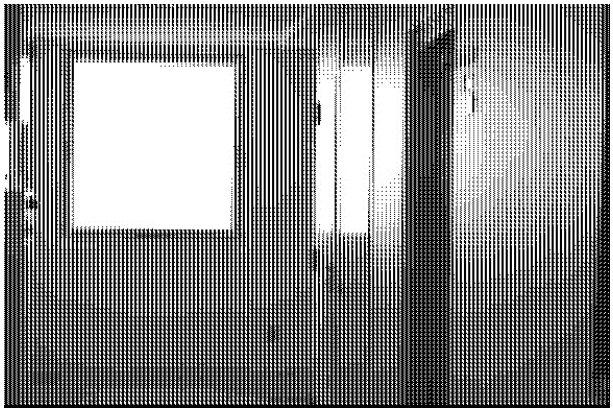
human observers a more comprehensive depiction of such scenes.

3.4 Chapter Summary

Experimental results obtained from simulating the photo capture process show that an HDR image reduces averaged angular error of gradient values by 57.5% for a designed



(a)



(b)

Figure 3.16: *Rendering results for an HDR image and a conventional digital image in the dithering style. (a) Rendering results for an HDR image of the third scene in Figure 3.10. (b) Rendering results for a conventional digital image of the third scene in Figure 3.10.*

Gaussian pattern compared to that of the generated conventional 8-bit digital images. Preservation of details in both over-exposed and under-exposed regions of the scene is easily observed in rendered images. This detail preservation property of HDR images in NPR processes is crucial to reflecting human perception in visual information. Structures of objects in some scenes are well maintained when the proposal is applied to decorative mosaic rendering; while for conventional 8-bit digital images, there are some distortions in shapes of the rendering results. Rendering results for some HDR images appear dark compared to those of corresponding conventional digital images, which is mainly due to the tone mapping method used for compressing the dynamic range of HDR images. The final performance of the proposal is greatly influenced by the scene to be rendered. If the luminance condition changes rapidly in different regions of the scene to be rendered, then the proposal surely improves the rendering results. Furthermore, the evaluation experiment using the visual attention model shows that new salient regions are formed in

rendering results for HDR images.

For designers and artists who seek approaches to more informative illustrations with newly formed salient regions that attract visual attention, potential rendering tools can be extended from the proposal. How gradient values of HDR images may influence other features such as colors and textures in an NPR process are, moreover, interesting topics to be studied further.

Chapter 4

Informative Bright Region Detection in High Dynamic Range Scenes

Visual cues like traffic lanes [13][14] and other traffic signs [15][16] are recognized from image sequence captured by car-mounted camera in autonomous driving systems [17]. Car-mounted camera usually encounters the problem that visual contents in bright region of specific scene are missing due to over-exposure (e.g., the daytime scene of tunnel exit). To capture such a wide dynamic range scene, a variety of High Dynamic Range (HDR) image composition methods [5][6][4][18] and hardware [19][20] have been developed. Hardware and computation costs, however, limit the large promotion of these well-designed methods. For a daytime tunnel scene in which the region outside tunnels is over-exposed, shortening the shutter time or lowering the gain of Charge Coupled Device (CCD) image sensor at the right time enables acquisition of the over-exposed region (i.e., IBR: Informative Bright Region). Also, the incidence rate of traffic accidents in tunnel scene tends to increase near the tunnel exits [21]. So, there is a necessity that the IBR in tunnel exit scene being located in order to find optimum timing for the control process. An algorithm for estimating location of IBR in tunnel scenes is proposed. In a typical tunnel scene, regularly laid lights on tunnel roof and road marking of traffic lane are preferable cues for the estimation process. First, tunnel lights are marked using the brightness and color tone information of them, and markings of traffic lanes are detected using their geometric properties. Second, a confidence map, which represents the possibility of IBR existence, is calculated by combining the geometric information of both tunnel lights and traffic lane markings. Third, the most possible location of IBR is determined by searching the confidence map of the current frame, and a re-locating process is carried out by using

already detected IBR locations in the previous frames, if needed. Finally, a refinement of the IBR location is done to compensate for the view angle of car-mounted camera by using the geometric properties of the over-exposed region. Compared with the tunnels on expressways, of which the lighting conditions are well designed, the algorithm is designed to deal with more challenging tunnel scenes on local roads with comparatively irregular lighting conditions. It takes advantage of visual cues which are closely related with IBR and not severely affected by luminance condition outside the tunnel. So the algorithm locates IBR under the condition that the visual information inside IBR is lost. Lighting and traffic lane cues are extracted using low cost image processing techniques (e.g., Canny edge detector), and the confidence map computation is done in limited region of the whole frame to achieve the real-time processing demand. To evaluate the performance of the proposed algorithm, a dataset consisting of three typical tunnel scenes on local roads is created. The video clips in the dataset are recorded by a camcorder mounted on a vehicle when it is passing through tunnels in the daytime. The first experiment is designed to make a comparison between the estimated locations of IBR given by the algorithm with those marked by four human observers. The testing platform used for experiments is a PC with a 3.4GHz CPU and 8GB of memory. The video clips used in experiment are normalized into the same size (640 pixels in width and 480 pixels in height). For each video clip, the human observers are requested to mark the potential IBR in 30 frames on their best efforts, and the algorithm is run on the same video for providing an estimation of the IBR. The second experiment is targeted at testing processing speed of the algorithm on the same platform. The averaged processing time is figured out by recording the time consumed in processing each frame of the videos in the dataset.

4.1 Response of Imaging Device due to Luminance in Tunnel Scenes

To record a specific scene as image sequence, shutter speed and gain value of imaging devices are controlled following some well-designed rules. In general, signal strength inside an image sensor is measured and compared with certain threshold for capturing the corresponding scene luminance.

In a typical daytime tunnel scene, luminance outside the tunnel is much stronger than

that inside the tunnel. In image sequence recorded by car-mounted camera on a car passing through a tunnel at daytime, the IBR (i.e., the region of tunnel exit) is over-exposed due to the limited dynamic range of image sensor as shown in Figure 4.1(a). When the car is getting closer to the tunnel exit, shutter time and gain of the image sensor is gradually adapted to the luminance condition outside of the tunnel as shown in Figure 4.1(b).

(a)

(b)

Figure 4.1: *Image sequence of a daytime tunnel scene captured by a car-mounted camera. (a) A frame captured far from the tunnel exit. (b) A frame captured close to the tunnel exit.*

For typical driving assistance tasks like lane detection and vehicle detection, there is a necessity that image sensor being adapted to the luminance condition outside the tunnel at an earlier stage so that information in IBR becomes visible for these tasks, which is challenging for the shutter and gain control mechanism of image sensors. The high incidence of traffic accidents within 100m from the tunnel exit [21] also suggests an earlier control of image sensor. To control shutter time and gain of the image sensor at

an earlier stage in a tunnel scene, the IBR should be located. An algorithm is designed for locating the IBR in tunnel scenes with low computational cost. The details of the proposed algorithm are covered in Section 4.2.

4.2 Locating Informative Bright Region using Lighting and Traffic Lane Cues in Tunnel Scenes

4.2.1 Lighting and Traffic Lane Cues for Locating IBR

The IBR in a daytime tunnel scene is usually over-exposed, which provides a distinct cue for locating the region. However, tunnel lights and headlights of vehicles share the similar property with IBR, which makes its locating process more challenging. Tunnel scenes have comparatively more regular structures than other traffic scenes, in which lines determined by the marking of traffic lanes and tunnel lights usually pass through IBR. So, the algorithm is designed for locating IBR in tunnel scenes by combining traffic lane and lighting cues. A flowchart of the locating process is depicted in Figure 4.2.

In rough, a confidence map is computed based on the properties of detected tunnel lights and traffic lanes, and the position with the highest value in the confidence map is selected as candidate location of IBR.

As a pre-processing step, the tunnel lights have to be detected in each frame of image sequence. The color tone of tunnel lights is determined by their types (e.g., sodium lamp or LED). Two tunnel scenes, in which color tone of the lights differs, are presented in Figure 4.3 with one light pixel marked in company with its RGB value.

The color tone of the light in Figure 4.3(a) has a comparatively lower intensity in one of color channels, while the light pixel in Figure 4.3(b) has a more balanced tone in three color channels. So, the light mask is designed as

$$L(x) = \begin{cases} 1, & \text{if } \sum_{n=1}^{3} Mask_n(x) \geq 2 \\ 0, & \text{otherwise} \end{cases}, \tag{4.1}$$

$$Mask_n(x) = \begin{cases} 1, & \text{if } C_n(x) > THRES \\ 0, & \text{if } C_n(x) \leq THRES \end{cases}, \tag{4.2}$$

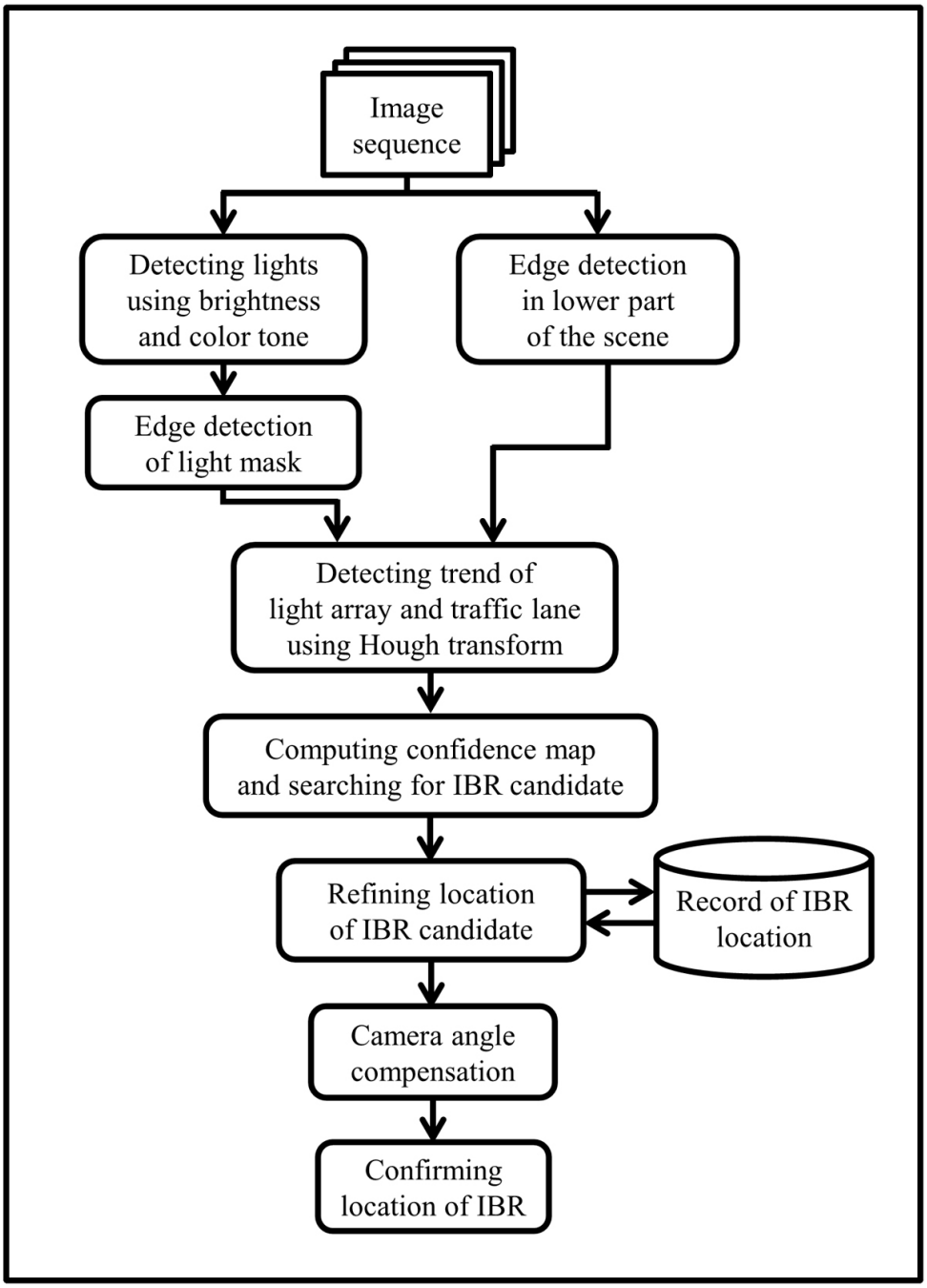
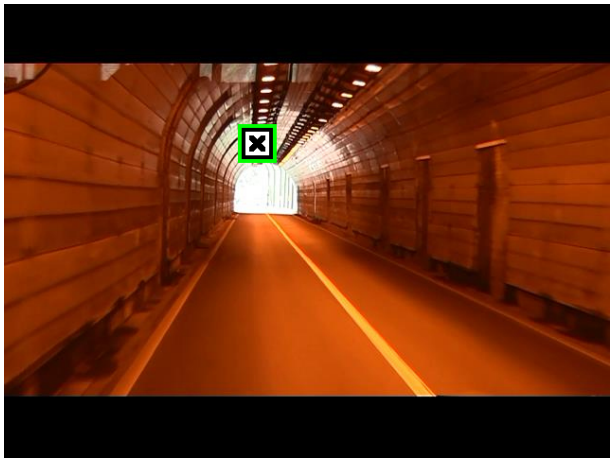


Figure 4.2: *A flowchart of the locating process of IBR using lighting and traffic lane cues.*

where $C_n(x)$ denotes the intensity value at position x in color channel n ($n = 1, 2, 3$) of the frame, and $THRES$ is a predefined threshold value. The obtained light mask of the two scenes in Figure 4.3 is depicted in Figure 4.4. The value of $THRES$ is set up by trial and error depending on lighting of tunnel scenes. In the computed light mask, not only the tunnel lights but also the tunnel exit is depicted.

The light mask and the pavement part (containing traffic lane information) of the



(a) *[R, G, B] = [255, 252, 206]*

(b) *[R, G, B] = [249, 246, 253]*

Figure 4.3: *Light pixel marked in tunnel scenes (by the center of the cross mark) along with their RGB values.*

corresponding frame are combined in one image for edge detection. The border between the light mask and the pavement part is adjusted according to the location of the IBR in previous frames (manually set in the first frame). Canny edge detector is chosen due to its reliability under various luminance conditions. The combining process and edge detection result is shown in Figure 4.5.

A combined image of the light mask and the bottom part is shown in Figure 4.5(b), and the output of the edge detector is shown in Figure 4.5(c). This combination process of light mask and the pavement part not only contributes to the computation cost of the whole algorithm but also gets the dominant information from each region (lights on the upper part and traffic lanes on the lower part) involved in the computing process of the confidence map.

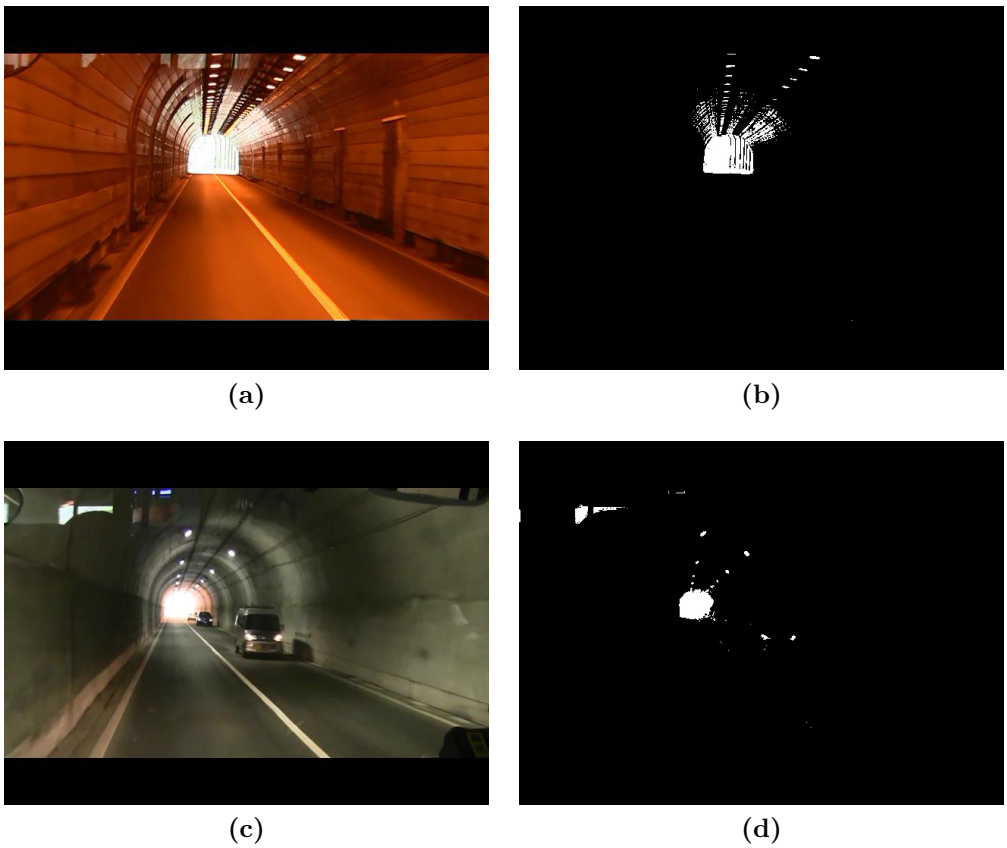
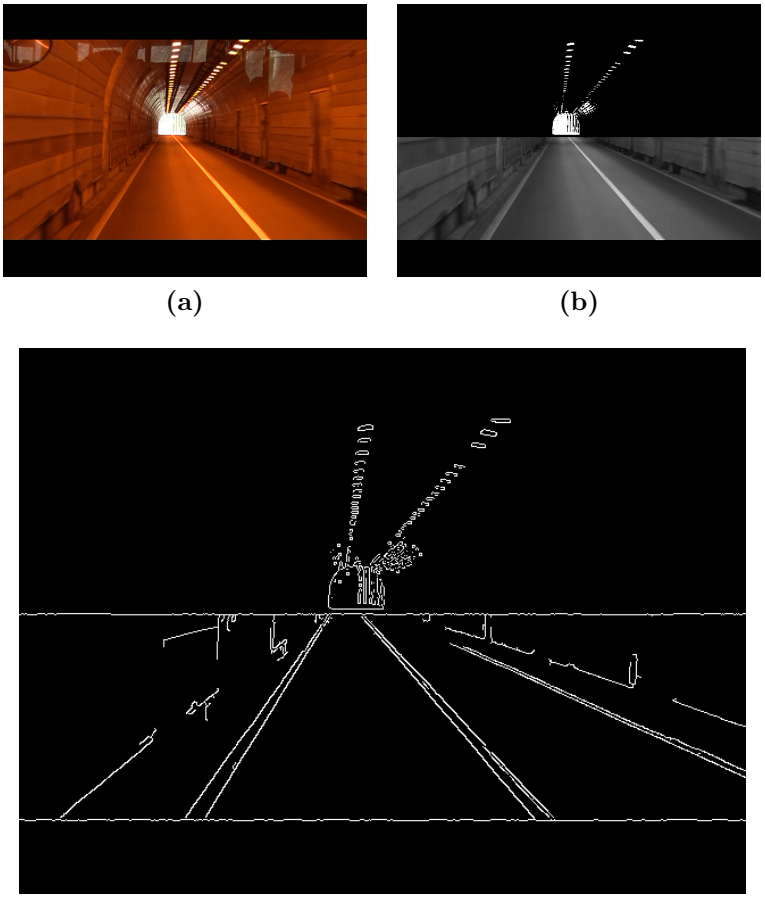


Figure 4.4: *Light mask computed from two tunnel scenes. (a) Tunnel scene with orange lights ($THRES = 230$). (b) Light mask of the tunnel scene (a). (c) Tunnel scene with white lights. (d) Light mask of the tunnel scene (c).*

4.2.2 Confidence Map Computation

Confidence map is defined as a two-dimensional array, of which each element shows the possibility of IBR existence. And indices of each element in the array represent a two-dimensional position in the corresponding tunnel image. The cues of lighting and traffic lane in tunnel have to be investigated as a whole in order to locate the IBR. Hence, a confidence map is computed by superimposing some weight belts whose properties are determined by the geometric attributes of traffic lanes and tunnel lights. Before computing the confidence map, the geometric attributes of traffic lanes and tunnel light arrays are figured out by applying a probabilistic Hough transform [40] to the edge detection result of the combined image. The line segments detected in Figure 4.5(c) is depicted on Figure 4.6. Note that the boundary of visible region in the frame and boarder between light mask and pavement part are filtered using their slope properties.

The confidence map is computed in a Region of Interest (ROI) for each frame in order to lower the computation cost. The ROI in the first frame is manually set by referring



(a) (b)

(c)

Figure 4.5: *Combined image of light mask and pavement part for edge detection. (a) Original image. (b) Combined image. (c) Result of detected edges.*

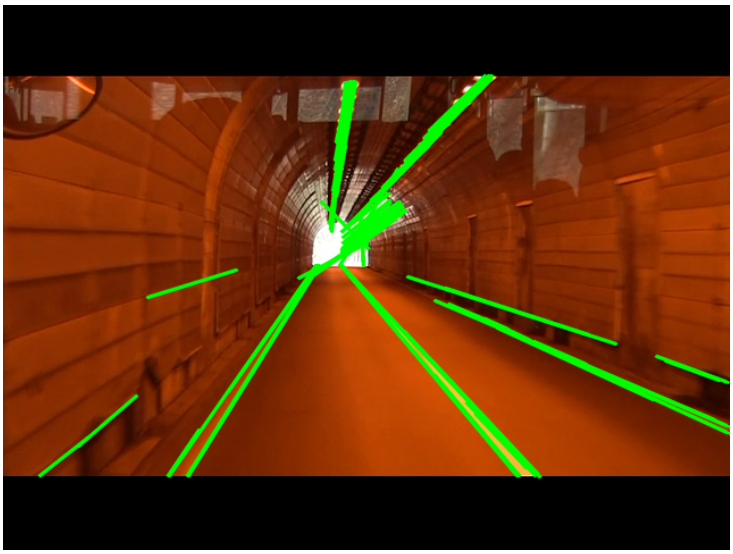


Figure 4.6: *Line segments detected on Figure 4.5(c) and shown in the original frame.*

to camera angle. From the second frame, the ROI is adjusted by IBR location in the previous frame. For each detected line segment LS_k $\{(x_0, y_0), (x_1, y_1)\}$ in the edge map, a corresponding weight belt $WB_k(x, y)$ is added into the confidence map $CM(x, y)$. The position and orientation of the weight belt $WB_k(x, y)$ follows those of the line segment as

$$WB_k(x,y) = \begin{cases} \alpha, & \text{if } \frac{|x-b_k y-c_k|}{\sqrt{1+b_k^2}} < \frac{BR_{WB}}{2}, \quad (x,y) \in ROI_{CM} \\ 0, & \text{otherwise} \end{cases}, \tag{4.3}$$

$$b_k = \frac{x_1 - x_0}{y_1 - y_0}, \tag{4.4}$$

$$c_k = x_0 - b_k y_0, \tag{4.5}$$

where α is a constant in all weight belts, b_k and c_k are slope and intercept of line segment LS_k, BR_{WB} and ROI_{CM} are the predefined breadth of weight belt and ROI used in confidence computation. The confidence map $CM(x, y)$ is computed by accumulating these weight belts $WB_k(x, y)$ as

$$CM(x,y) = \sum_{k=1}^{K} WB_k(x,y) + CM_{init}(x,y), \tag{4.6}$$

$$CM_{init}(x,y) = \beta \cdot \exp\left(-\frac{(y - Y_{ROI})^2}{2\sigma^2}\right), \tag{4.7}$$

where K is the number of line segments, $CM_{init}(x, y)$ is initial confidence, β is a constant for adjusting the influence of the initial confidence, Y_{ROI} is the ROI midpoint on Y-axis, and σ is used to control the shape of the Gaussian function. A sketch of superimposing weight belts on a confidence map is shown in Figure 4.7.

The confidence map computed from the traffic lane and lighting cues in Figure 4.6 is shown in Figure 4.8. The computation process of the confidence map is designed as a voting in which each line segment votes for a region that is represented by its corresponding weight belt. The computation cost of confidence map is influenced by two factors: the amount of the line segments and the width of the weight belt. When the number of the line segments used for computation is limited to a certain amount, the total cost for this step is controlled.

After the confidence map is figured out, a search is carried out on the confidence map to locate the position LOC_{MAX} with the highest confidence value. The position with

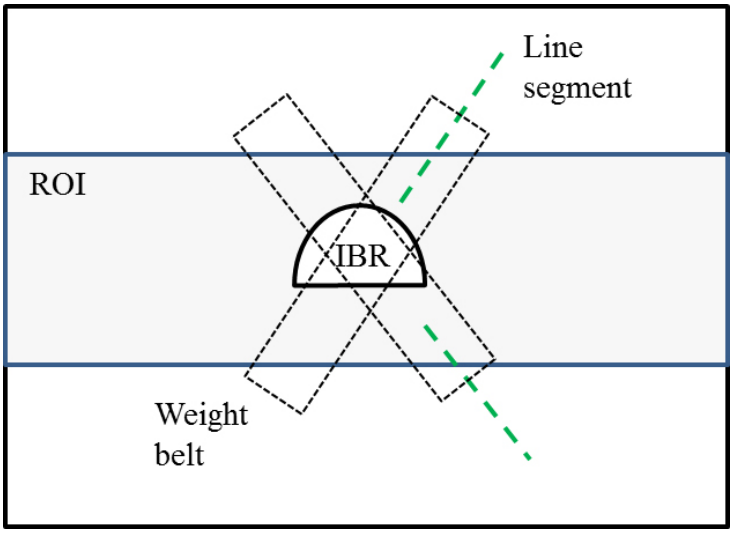


Figure 4.7: *Superimposing process of weight belts for computing confidence map.*

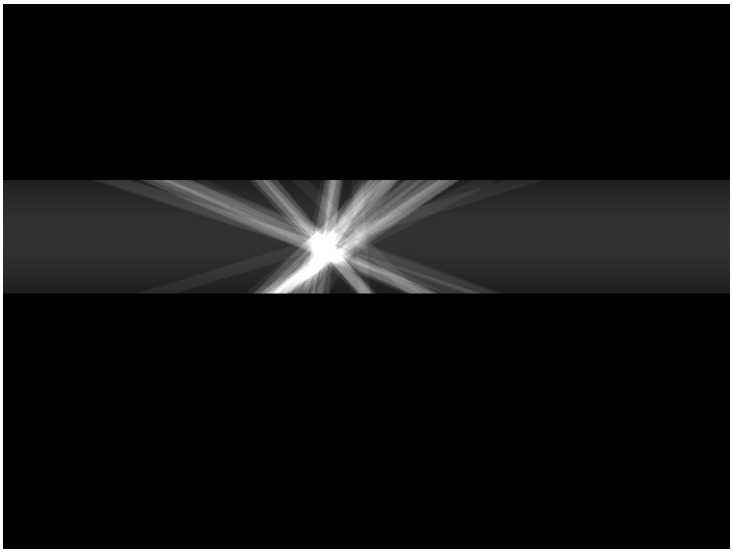


Figure 4.8: *Confidence map computed from traffic lane and lighting cues in Figure 4.6.*

the highest confidence value in Figure 4.8 is shown in Figure 4.9. The located position LOC_{MAX} is the candidate of IBR, and further processing steps (re-locating and view angle compensation) are discussed in Section 4.2.3.

4.2.3 IBR Re-locating and View Angle Compensation

Instead of the targeted visual cues, some other cues like extra tunnel structures, boundaries near tunnel exit, and passing vehicles also affect computation of the confidence map. The position LOC_{MAX} with highest value in the confidence map is occasionally directed to a deviated position due to the influence of these undesirable components. In such

Figure 4.9: *A frame with the highest value in its confidence map (Figure 4.8) marked as a cross.*

a case, the position of LOC_{MAX} is re-located by a second search in a narrower region determined by the IBR locations in the previous N frames. Note that the IBR locations in first N frames are directly set as LOC_{MAX} without a re-locating process.

The re-locating process involves two steps:

(1) A judgment is made about whether re-locating is needed in current frame by calculating the Euclidean distance between LOC_{MAX} and recorded IBR location IBR_n in each of the previous N frames. If the distance between LOC_{MAX} and IBR_n exceeds a pre-defined threshold in more than M frames out of the previous N frames ($M < N$), a re-locating process of LOC_{MAX} is needed in the current frame.

(2) The averaged IBR location IBR_{MEAN} is calculated in the previous N frames. A search for maximum position $LOC2_{MAX}$ is carried out in a limited region (e.g., 40×40) centered at IBR_{MEAN}.

View angle compensation is usually needed due to varying positions and angles of the car-mounted cameras. For the purpose of fast computation, the view angle compensation is done by averaging the locations of the over-exposed pixels along a horizontal line passing LOC_{MAX} or $LOC2_{MAX}$ (in the case that LOC_{MAX} is re-located). The IBR position IBR_n of frame n is finally determined as the averaged location of these over-exposed pixels. Note that view angle compensation is carried out from the first frame of the video. The re-locating and view angle compensation process is illustrated in Figure 4.10,

where the circle shows LOC_{MAX} (the position with the highest value in confidence map), the triangle shows $LOC2_{MAX}$ (re-located position of LOC_{MAX}), and the rectangle shows IBR_n (location of IBR after view angle compensation).

Figure 4.10: *A frame with its LOC_{MAX}, $LOC2_{MAX}$, and IBR_n positions marked as circle, triangle, and rectangle shapes.*

4.3 Experiments on Locating Informative Bright Regions on General Roads

4.3.1 Tunnel Video Dataset and Labeling Tool

To evaluate the algorithm, a dataset which contains three typical tunnel scenes is created. Instead of selecting the expressway tunnel scenes in which the luminance condition is well designed, more challenging tunnel scenes on local roads are chosen for building the dataset. Tunnel scenes with different lighting conditions, which cover varying cases of tunnel scenes on Japanese local roads, are recorded for validating the algorithms tolerance to varying tunnel lighting as shown in Figure 4.11. The array number of lights on tunnel ceiling varies in the recorded tunnel scenes. The tunnel lights in Figure 4.11(a) and (c) share the similar orange color tone, while the lights in Figure 4.11(b) shows white color tone. The density of tunnel lights is also different in the recorded tunnel scenes as shown in Figure 4.11(b) and (c). The tunnel scene videos in the dataset are normalized

into the same size (640 pixels in width and 480 pixels in height) and the same frame rate ($29.97FPS$) for the experiments.

labeling tool is also designed for human observers to mark the IBRs in the tunnel scene dataset. The window form of the labeling tool and the marker is shown in Figure 4.12. In the window form of the labeling tool, the mouse cursor is replaced with an original marker which is composed of two nested squares and one cross. When a human observer identifies location of IBR in one frame and left-clicks the mouse, the location of IBR is recorded in a log file. After the IBR in one frame is labeled, the location is marked as a square for the human observer to check it.

4.3.2 Evaluation of IBR Locating Accuracy and Computation Cost

The testing platform used for the evaluation is a PC equipped with 3.4GHz CPU and 8GB memory. The code is programmed with C++ and compiled with Visual Studio 2010 in the release mode.

Four human observers are requested to label three typical scenes shown in Figure 4.11 using the labeling tool shown in Figure 4.12. In each of the three video clips, 30 frames are labeled with an interval of 5 frames. In each labeled frame, the IBR locations assigned by the four human observers are averaged for relieving each observers labeling mistake and used as the baseline for evaluation. The mean labeled position of IBRs is firstly compared with the location computed by the algorithm until the re-locating step is done (view angle compensation not applied yet). When running the algorithm, the value of $THRES$ for light mask computation is set to 230 for all the three tunnel scenes in the dataset. The obtained confidence map examples for the three tunnel scenes in Figure 4.11 are presented in Figure 4.13. Note that the ROI for computing confidence map is automatically adapted on the vertical direction to the specific frame. This adaptation process of the ROI is guided by the averaged IBR locations in $N(N = 10)$ previous frames. And the calculated IBR locations using the algorithm for the frames in Figure 4.13 are shown in Figure 4.14. The errors in locations given by algorithm are shown in Figure 4.15 as cross markers for each of the three tunnel scenes in Figure 4.11. The view angle of the car-mounted camera makes the errors on X-axis becoming larger. After view angle compensation, the errors on X-axis are decreased to a lower level and shown as circle markers in Figure 4.15. There

(a)

(b)

(c)

Figure 4.11: *Three typical tunnel scenes in the self-built dataset.*

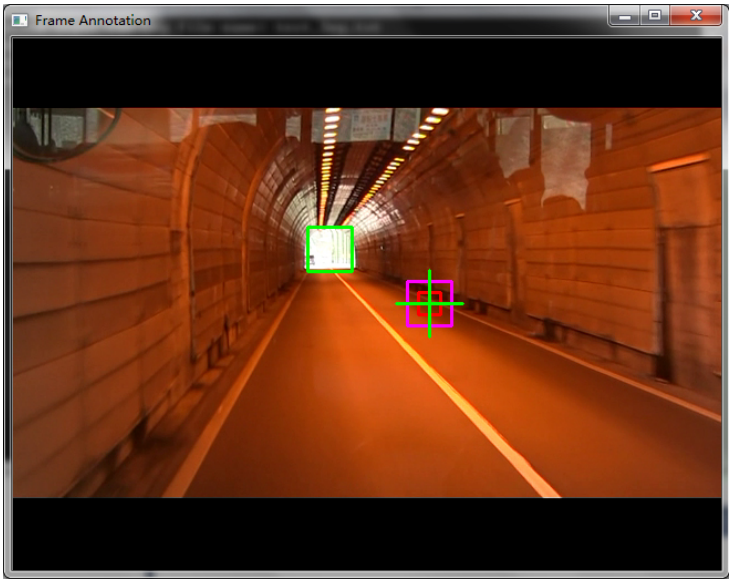


Figure 4.12: *The window form of the designed labeling tool with the marker.*

Table 4.1: *Averaged location error of the IBR detection algorithm (in pixel).*

		Video No.1	Video No.2	Video No.3	In total
Without Compensation	X-axis	19.49	13.98	15.43	16.30
	Y-axis	4.62	3.08	4.98	4.23
With Compensation	X-axis	4.13	4.28	4.03	4.14
	Y-axis	4.62	3.08	4.98	4.23

is one error point on the top left corner of Figure 4.15(c) that is not well compensated, of which the error originates from the searching step of highest value in confidence map. The computed locations without view angle compensation tend to distribute on the left side of the compensated locations. This is due to the fact that the camera is mounted on a vehicle running on the left lane of the road when the dataset is built.

The averaged error of the locations given by algorithm is summarized in Table 4.1. When view angle compensation is carried out in the algorithm, the averaged error of location on X-axis is reduced to the same level as that on Y-axis. Note that the errors are calculated in their absolute value. After view angle compensation, the location error on X-axis is reduced by 74.6%. The unchanged location error on Y-axis after view angle compensation is owing to the fact that this process is designed for compensating the location horizontally.

The processing time for the first 200 frames in each of the three tunnel scenes is presented in Figure 4.16. The averaged processing speed for the three tunnel scenes and their average is summarized in Table 4.2. The averaged processing time of one frame

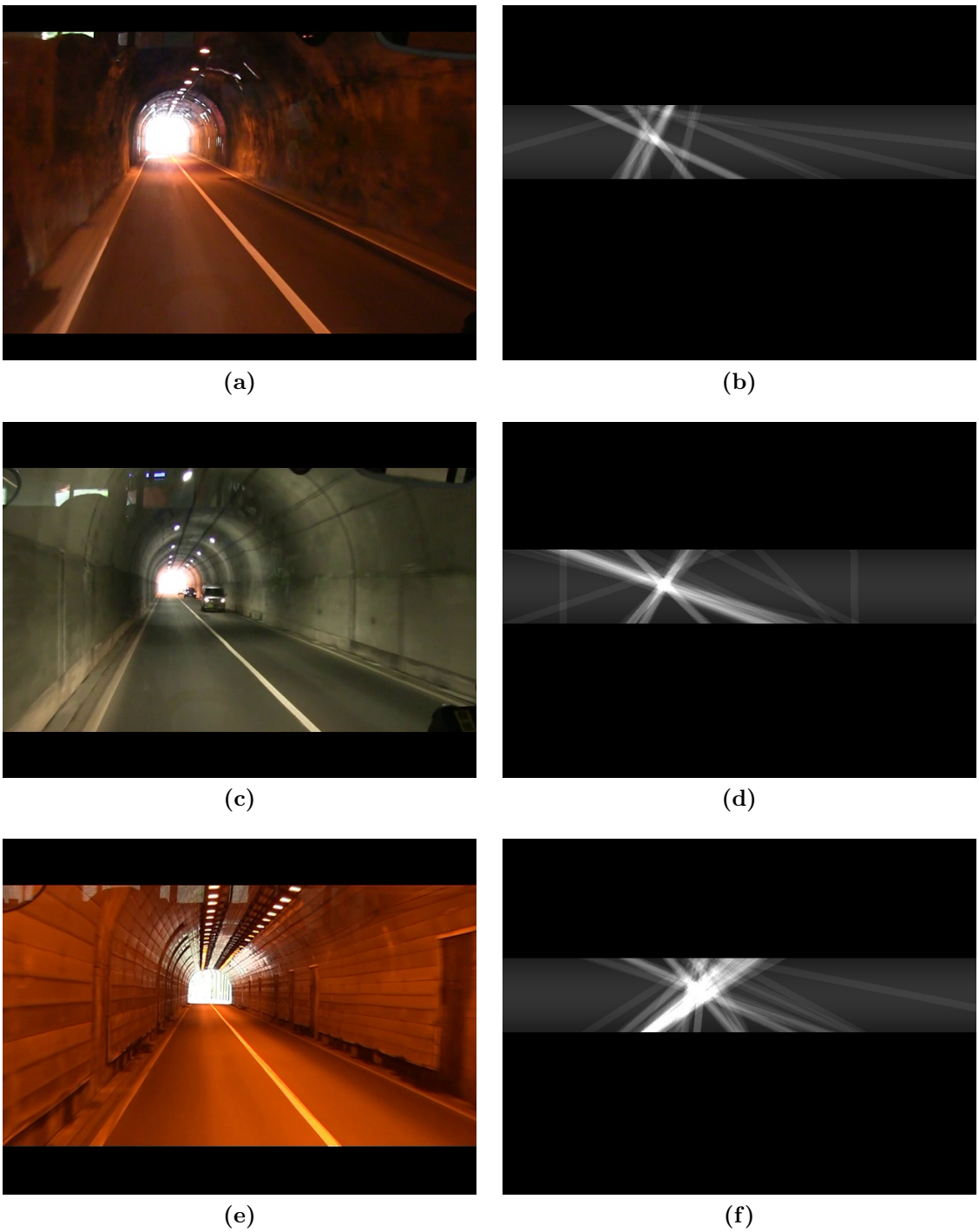


Figure 4.13: *The examples of confidence map for the three tunnel scenes in Figure 4.11. (a)(c)(e) The original frames for calculating confidence map. (b)(d)(f) The obtained confidence map.*

(with a frame size of 640×480) is around 10ms, which achieves the real-time processing demand. A further insight into the algorithm shows that the most time consuming steps of the algorithm are edge detection, Hough transform for line segment searching, and confidence map computation. The edge detection and Hough transform steps have the

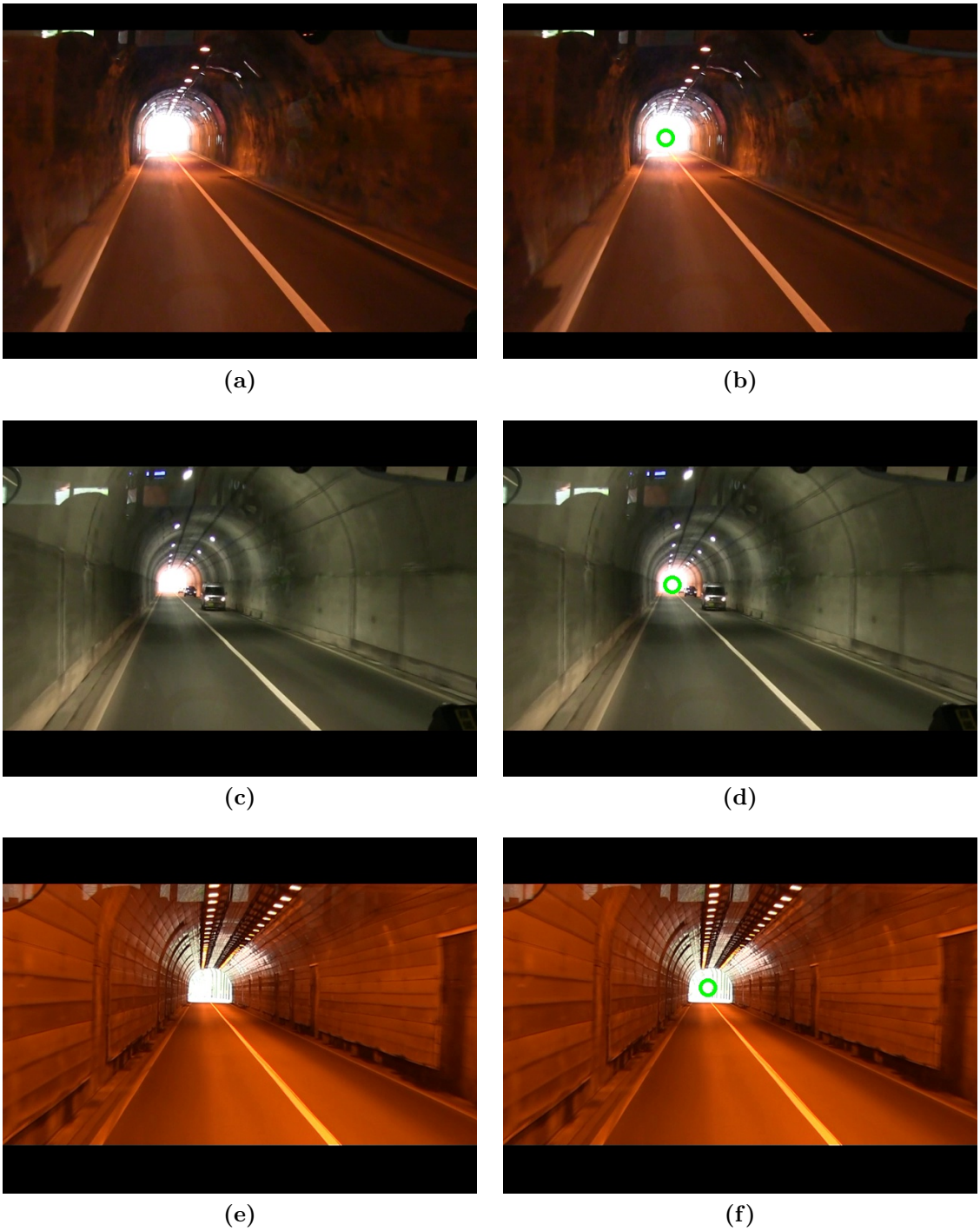


Figure 4.14: *The calculated IBR locations using the algorithm for the three tunnel scenes in Figure 4.11. (a)(c)(e) The original frames for calculating IBR location. (b)(d)(f) The obtained IBR locations in the corresponding frames.*

potential to be further accelerated using Graphics Processing Unit (GPU).

Table 4.2: *Averaged processing speed of the algorithm for one frame.*

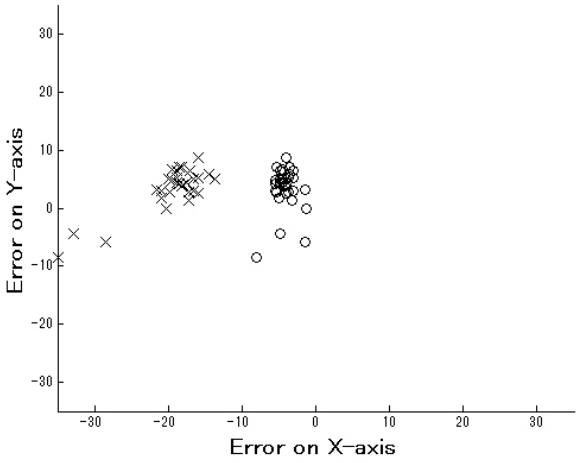
	Video No.1	Video No.2	Video No.3	In total
Processing time (ms)	9.74	9.95	10.44	10.04

4.4 Chapter Summary

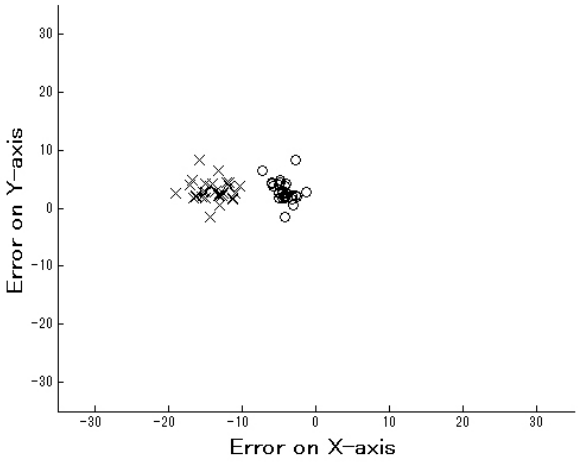
The experimental results using the tunnel scene dataset show that the locations of IBRs given by the algorithm have averaged error of 4.14 pixels out of 640 pixels on X-axis and 4.23 pixels out of 480 pixels on Y-axis in comparison with the IBRs labeled by the human observers. For varying tunnel scenes with different structures and lighting conditions on local roads, the averaged location error ranges from 4.03 to 4.28 pixels on X-axis and from 3.08 to 4.98 pixels on Y-axis. The view angle compensation step makes the location error on X-axis reduce to the same level of that on Y-axis by 74.6%. The average processing speed of the algorithm on one frame (640×480) is around 10ms for varying tunnel scenes in the dataset.

The accuracy (within 1% on each axis) of the algorithm is appropriate for image sensor control due to the fact that shutter time is supposed to be made shorten when the IBR covers a certain portion of whole frame. The processing speed (around 10ms per frame) of algorithm is also able to keep pace with the frame rate (usually 30 frames per second) of conventional cameras.

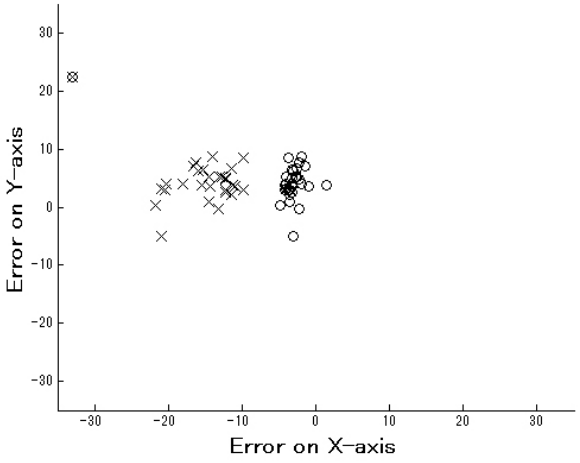
The low computation cost of the algorithm supports its application in devices with limited computing power like vehicle driving recorder. The algorithm aims at providing good timing for shutter and gain control of image sensor inside these devices so that the critical visual information inside the IBR are made visible at an earlier stage, which contributes to earlier record of over-exposed region of tunnel exits and enhancement of driving safety.



(a)

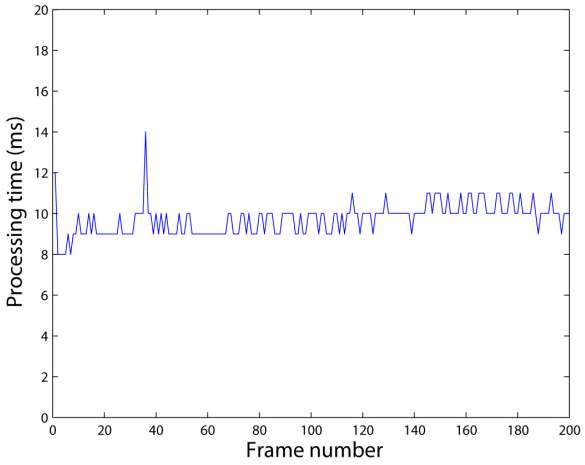


(b)

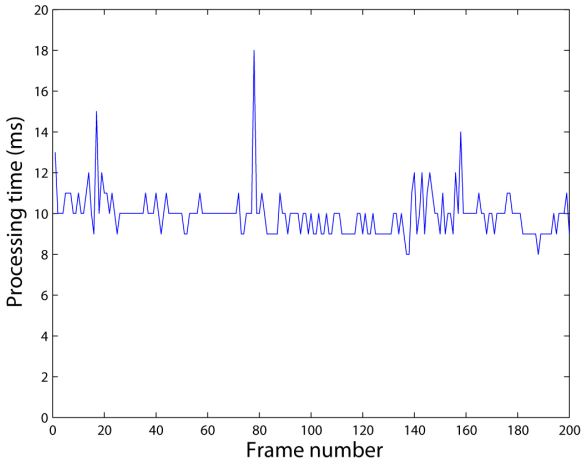


(c)

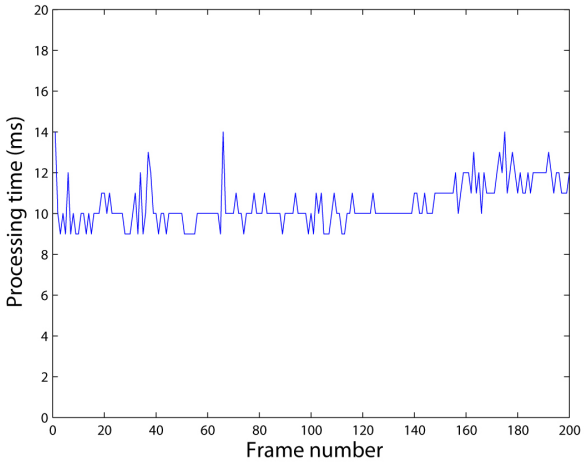
Figure 4.15: *The location errors of the algorithm compared with the labeled data (cross: before view angle compensation, circle: after view angle compensation). (a) Errors occurred the scene in Figure 4.11(a). (b) Errors occurred in the scene in Figure 4.11(b). (c) Errors occurred in the scene in Figure 4.11(c).*



(a)



(b)



(c)

Figure 4.16: *Processing speed of the algorithm in the three tunnel scenes. (a) Processing speed in the scene of Figure 4.11(a). (b) Processing speed in the scene of Figure 4.11(b). (c) Processing speed in the scene of Figure 4.11(c).*

Chapter 5

Vision Enhancement of Informative Bright Region in High Dynamic Range Scenes

The video stream output of car mounted camera is usually influenced by the rapidly changing luminance condition on general roads. Tunnel scene is a typical one with changing luminance condition. The high dynamic range of tunnel scene makes it different to capture scene information with conventional car mounted camera, of which the captured dynamic range is limited by the quantization level of each individual pixel. For capturing an HDR scene in its full dynamic range, typical methods such as scene radiance map reconstruction using multiple exposure have been developed by Debevec et al. [5]. Extension of this multiple exposure method from single HDR image to HDR video has also been reported [19]. However, specially designed imaging device, whose shutter speed is possible for fast adjustment, is necessary. Also, the high computation cost for post-processing of this HDR video capturing method limits its application in low-cost hardware like driving recorder. Mertens et al. [22] and Song et al. [41] proposed a technique for fusing a set of differently exposed images into a high quality one. This image fusion technique skips the reconstruction process of HDR image (scene radiance map), and the tone mapping step is also not needed accordingly. For capturing video in an HDR tunnel scene using this technique, highly controllable imaging device and additional computation resource for image registration are still imperative as Kang's method does[19]. Tocci et al. [20] introduced an optical structure, which generates differently exposed images with beam splitters, for capturing HDR video. Instead of using single image sensor, at least three image sensor is needed for constructing the optical structure. Based on the observation that the same quantity in the low portion of quantized bits may provide much more de-

tails than that in the high portion, a method is proposed for visually enhancing the IBR by triggering exposure compensation on the video capture device (e.g. camera-mounted camera) in HDR tunnel scenes.

In brief, the exposure of video capture device is adjusted based on the change of the detected IBR in a tunnel scene, and the captured image sequence is compensated using gamma correction. There are mainly four steps in this proposal. First, the location of IBR (over-exposed tunnel exit) is detected using the algorithm introduced in Chapter 4. Second, size of over-exposed region around the IBR location is figured out and recorded for each frame. Third, a certain level of exposure compensation is triggered when the size of over-exposed region exceeds a predefined threshold. Finally, the captured frame under exposure compensation is enhanced using gamma correction. Another method is also given for vision enhancement of IBR using liquid crystal glasses. The transmittance of the glasses is changed from high to low when the size of over-exposed region exceeds a predefined threshold.

Forcing less exposure during video capturing process makes the dynamic range be more adapted to the over-exposed exit region of the tunnel scene. Considering the prerequisite that computation power of hardware is limited, the dark region caused by the forced exposure compensation is enhanced by gamma correction in each pixel. By selecting a suitable gamma value, which is mainly decided by referring to a historical frame before exposure compensation, both the strongly illuminated objects observed through tunnel exit and the weakly illuminated region inside the tunnel are recorded and presented in a clearly perceptible manner.

Experiment is carried out by porting the IBR location detection algorithm on single-board computer with a camera module. The IBR location detection algorithm runs during the video capturing process. Exposure compensation through EV adjustment is triggered by monitoring the size change of IBR. The captured video sequence is thus influenced by the adjusted camera parameter (EV bias) and is saved in storage media after encoding process.

5.1 Vision Enhancement of Informative Bright Region on Imaging Device

5.1.1 Exposure Control of Imaging Device for Informative Bright Region Enhancement

The amount of light energy reaching per unit area on an image sensor surface is mainly determined by three factors: aperture size, shutter duration, and luminance condition [42]. For a specific scene, its luminance condition is possible to be considered as unchanged during a short period. So, shutter duration and aperture size are adjusted manually or automatically in order to properly capture the scene. The combination of aperture size and shutter duration is widely expected as exposure value (EV) [43], and is numerically defined as

$$\begin{aligned} EV &= \log_2 \frac{N_{aperture}^2}{t_{shutter}} \\ &= 2\log_2 N_{aperture} - \log_2 t_{shutter}, \end{aligned} \tag{5.1}$$

where $N_{aperture}$ is f-number (i.e. relative aperture size), and $t_{shutter}$ is the shutter duration. For instance, a shutter duration of $1s$ and a relative aperture size of $f/1.0$ result in a EV value of 0 [43]. A larger EV value causes less light energy being casted on unit area of image sensor surface, and vice versa. For embedded camera module on electronic device (e.g. car driving recorder, mobile phone), its f-number ($N_{aperture}$) is usually a fixed value, which compromises on the need for mass production. In terms of these devices, shutter duration $t_{shutter}$ becomes the main controllable factor for adapting to scenes with different luminance conditions. Note that signal gain in image sensor is also an adjustable parameter for scene luminance adaptation. Inside conventional imaging device (CCD or CMOS), electronic shutter is applied due to its low cost and easiness for control. An abstracted controlling circuit for imaging device is shown in Figure 5.1. The light is passing through the lens system and the aperture, and then converted into electrical signal by the CCD image sensor. Before the analog-to-digital (A-D) conversion step, the strength of the output signal by the image sensor is amplified by introducing a certain level of gain. The combination of shutter time control and gain control determines the exposure of each frame. And the exposure is usually adjusted by referring to the weighted average luminance of a specific scene. The sequential images in Figure 5.2 show the exposure

adjustment of a car-mounted camera from a position deep inside the tunnel to a position close to the tunnel exit. The tunnel exit regions in Figure 5.2(a) and Figure 5.2(b) are over-exposed due to the gap between the luminance inside the tunnel and that outside the tunnel. Due to the importance of visual information inside IBR in daytime tunnel scenes, it is advisable to take the location of IBR as a factor for exposure control as shown in Figure 5.3.

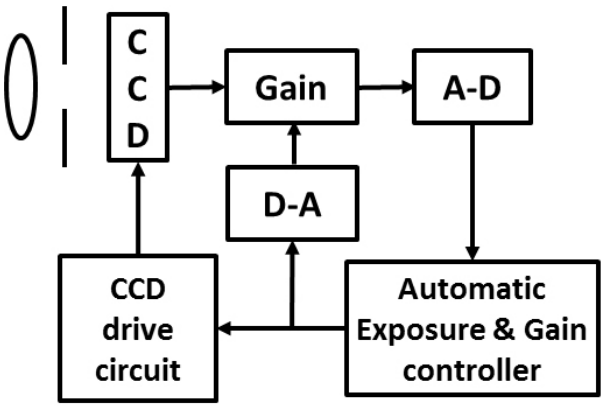


Figure 5.1: *An abstracted control circuit for imaging device.*

When the video camera mounted on a vehicle is getting close to the tunnel exit, the number of the over-exposed pixels in the IBR is increasing. The IBR detection algorithm provides the position of searching area for monitoring size change of the over-exposed region. The searching area ($L \times L$) used for counting pixels in the over-exposed region has to be set to a proper value so that the whole process is monitored until the video camera reaches a spot for triggering exposure compensation. Three tunnel scenes with the searching area centered on the IBR location are shown in Figure 5.4. And the change of over-exposed pixels is presented in Figure 5.5 for these three scenes. Size change of over-exposed region in practical daytime tunnel scenes is influenced by several factors, e.g., accuracy of the IBR detection algorithm, vehicles near the tunnel exit, vibration of the car-mounted video camera, and so on. Besides these factors, the over-exposed region becomes suitably exposed when the car-mounted video camera is moving to a spot near the tunnel exit. This is the primary cause for the decreasing trend of over-exposed region in Figure 5.5(c).

The change of the over-exposed pixels in all the three tunnel scenes of Figure 5.5 has a growing trend, which provides a timing reference for setting the camera parameters. Considering the nature that various factors (e.g., vehicles in the tunnel and vibration

(a)

(b)

(c)

Figure 5.2: *Video frames captured by a car-mounted camera when driving out from a tunnel in daytime.*

during driving) may give rise to a sudden size change of the over-exposed region, the trigger of exposure compensation for the imaging device is designed as that over-exposed pixel number is greater than a threshold $A_{trigger}$ in M out of N sequential video frames.

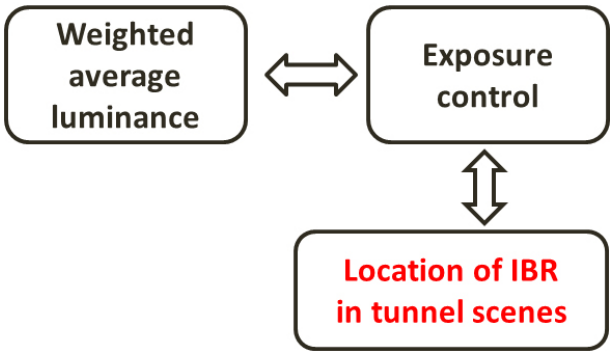


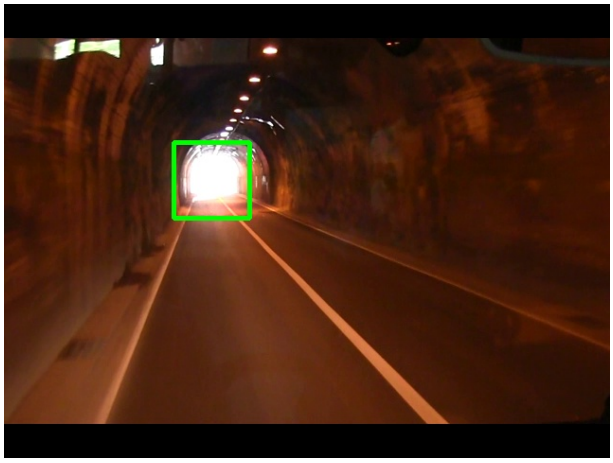
Figure 5.3: *IBR as a factor for exposure control in image devices.*

Examples of applying this triggering strategy to the three tunnel scenes are shown in Figure 5.6. For each tunnel scene, the threshold $A_{trigger}$ is set to two different values (2500, 3500), and the scene frames correspond to the triggers are also given. As shown in this example, a larger threshold value ($A_{trigger}$) leads to a trigger closer to the tunnel exit, and vice versa.

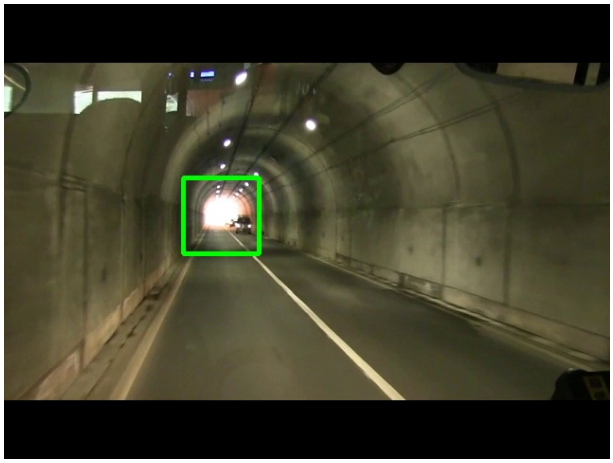
When the timing for exposure compensation is determined by the triggering strategy in a daytime tunnel scene, a larger EV (exposure value) is selected so that less light is casted on unit area of the image sensor. According to (5.1), increasing one stop on EV leads to the casted light being reduced by half. An image set of a daytime tunnel exit is presented in Figure 5.7 for showing the effect of EV adjustment. Note that this image set is captured in still format by manually setting the camera parameters. As additional stops are added on EV (less light to be casted on image sensor), the over-exposed tunnel exit region becomes visible. A certain level of exposure compensation functionality (adjustment by EV) is usually available in modern image sensors, which backs up its applicability on low cost hardware.

5.1.2 Dark Region Enhancement by Gamma Correction

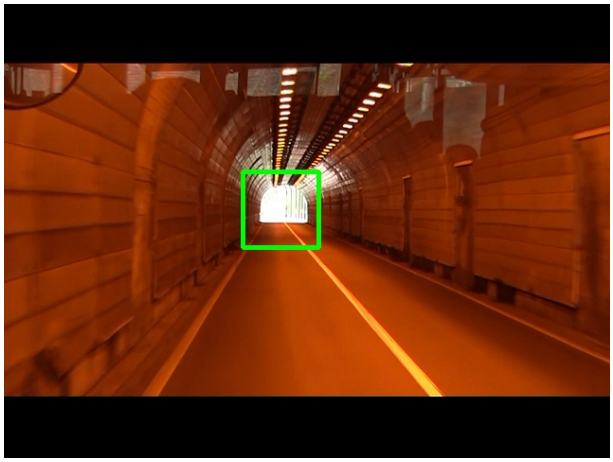
When the exposure compensation is triggered by the proposed strategy, the dark region of the scene is represented with less quantization level. The simplest strategy for enhancing the dark region is to make the region brighter by amplifying its strength. However, the bright region of the scene has to be processed separately, which causes visually discontinuous boundaries between the bright region and the dark region. For processing on platform with high computation power (e.g., desktop computer), various well-designed



(a)



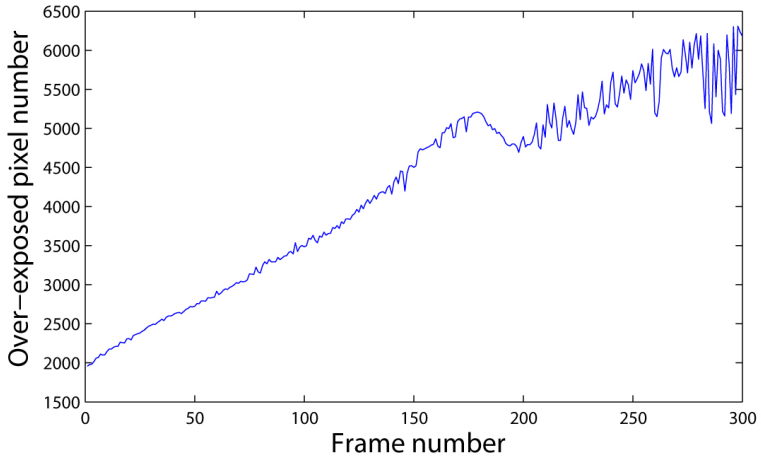
(b)



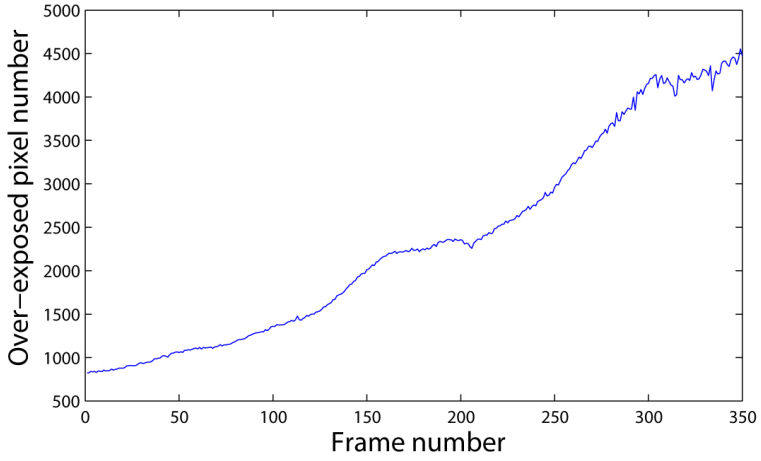
(c)

Figure 5.4: *Three tunnel scenes with the searching area (80 × 80) for over-exposed pixels centered in IBR location.*

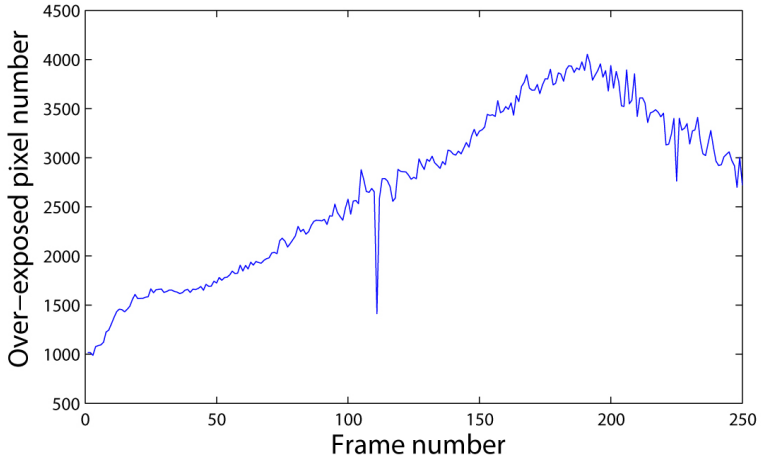
image enhancement algorithms [44][45] have been developed. Gamma correction (mapping by a power function) is selected as the tool for re-balancing the brightness of the bright region and dark region in the captured frame after exposure compensation. It



(a)



(b)



(c)

Figure 5.5: *Size change of over-exposed region in practical daytime tunnel scenes shown in Figure 5.4.*

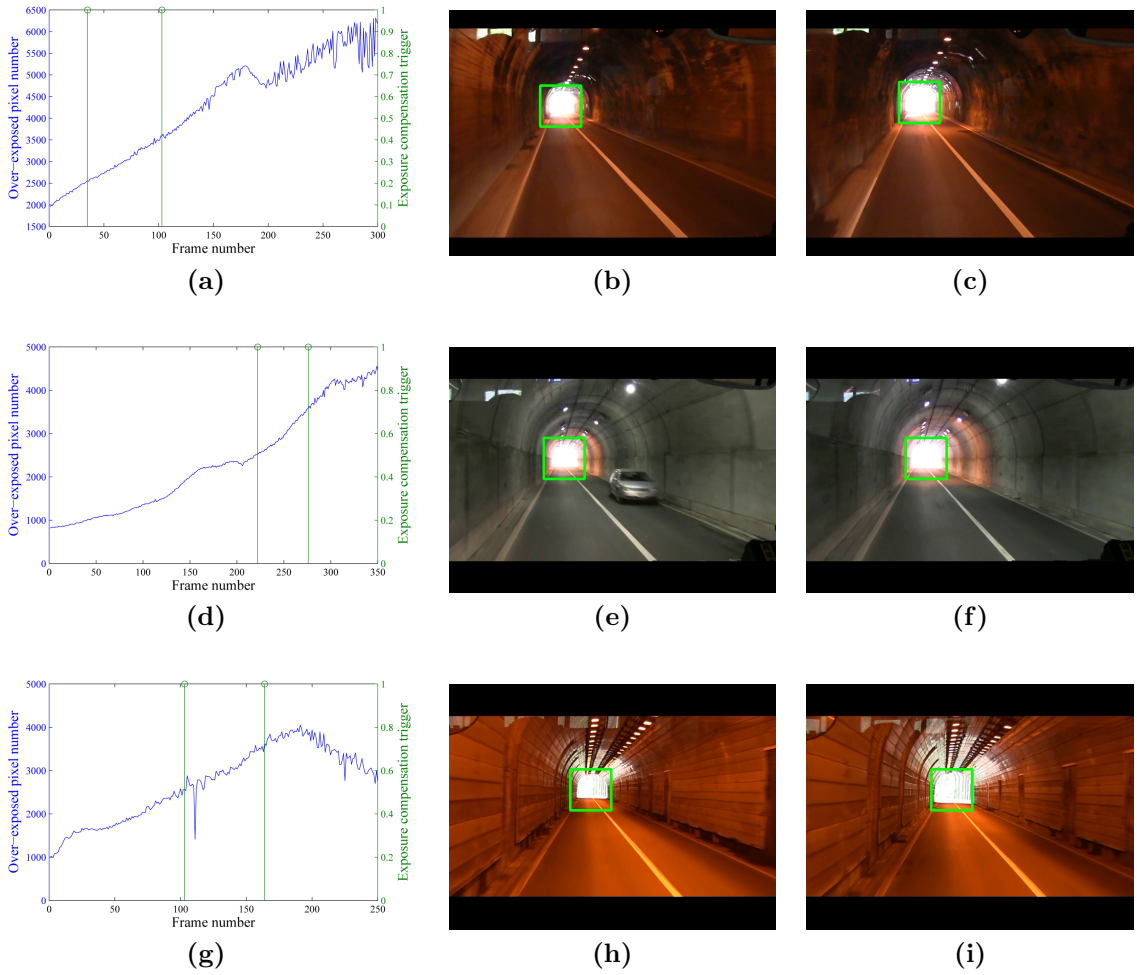
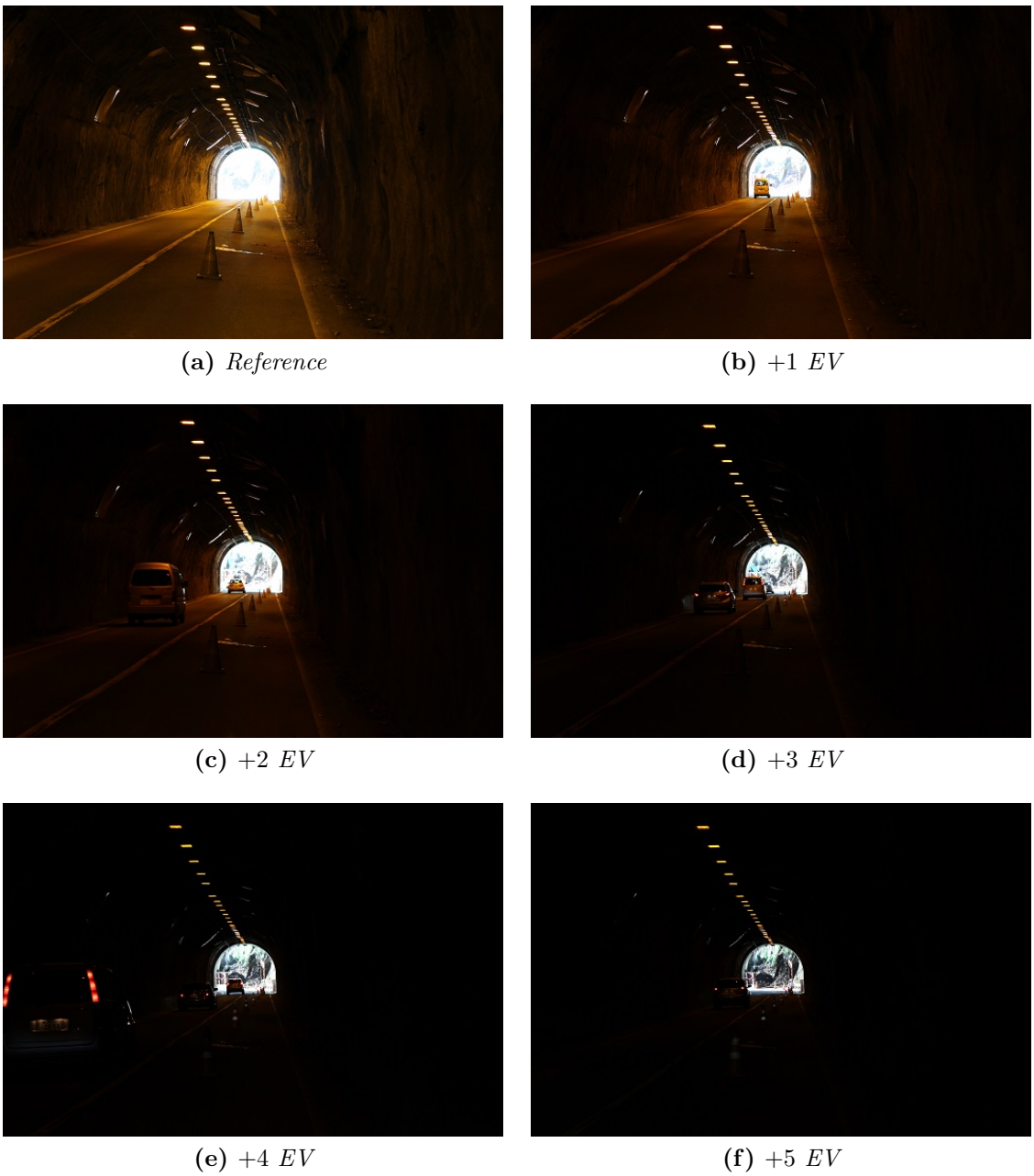


Figure 5.6: *Examples of applying the triggering strategy (searching area set to* 80×80*,* $A_{trigger}$ *set to 2500 and 3500) to three tunnel scenes in Figure 5.4.(a)(d)(g) The size change of over-exposed region with two overlaid triggers. (b)(e)(h) The frame correspond to the trigger with* $A_{trigger}$ *set to 2500. (c)(f)(i) The frame correspond to the trigger with* $A_{trigger}$ *set to 3500.*

assigns a larger portion of the total intensity level to the dark region, which is in accordance with the high sensitivity of human vision to a dark scene. The process of gamma correction [46] is globally applied to the captured frame and formulated as

$$V_{out} = {V_{in}}^{\gamma}, \tag{5.2}$$

where V_{in} is the input signal strength, and V_{out} is the output signal strength, and γ is the parameter for controlling the shape of the power function. When the value of γ is less than 1, dark region of an image is enhanced. When the value of γ equals 1, the output image stays the same with the input one. For enhancing a captured frame of daytime tunnel scene after exposure compensation, γ is set to a value less than 1. An example for showing the enhancement of dark region by gamma correction is given in Figure 5.8.



(a) *Reference* **(b)** +1 *EV*

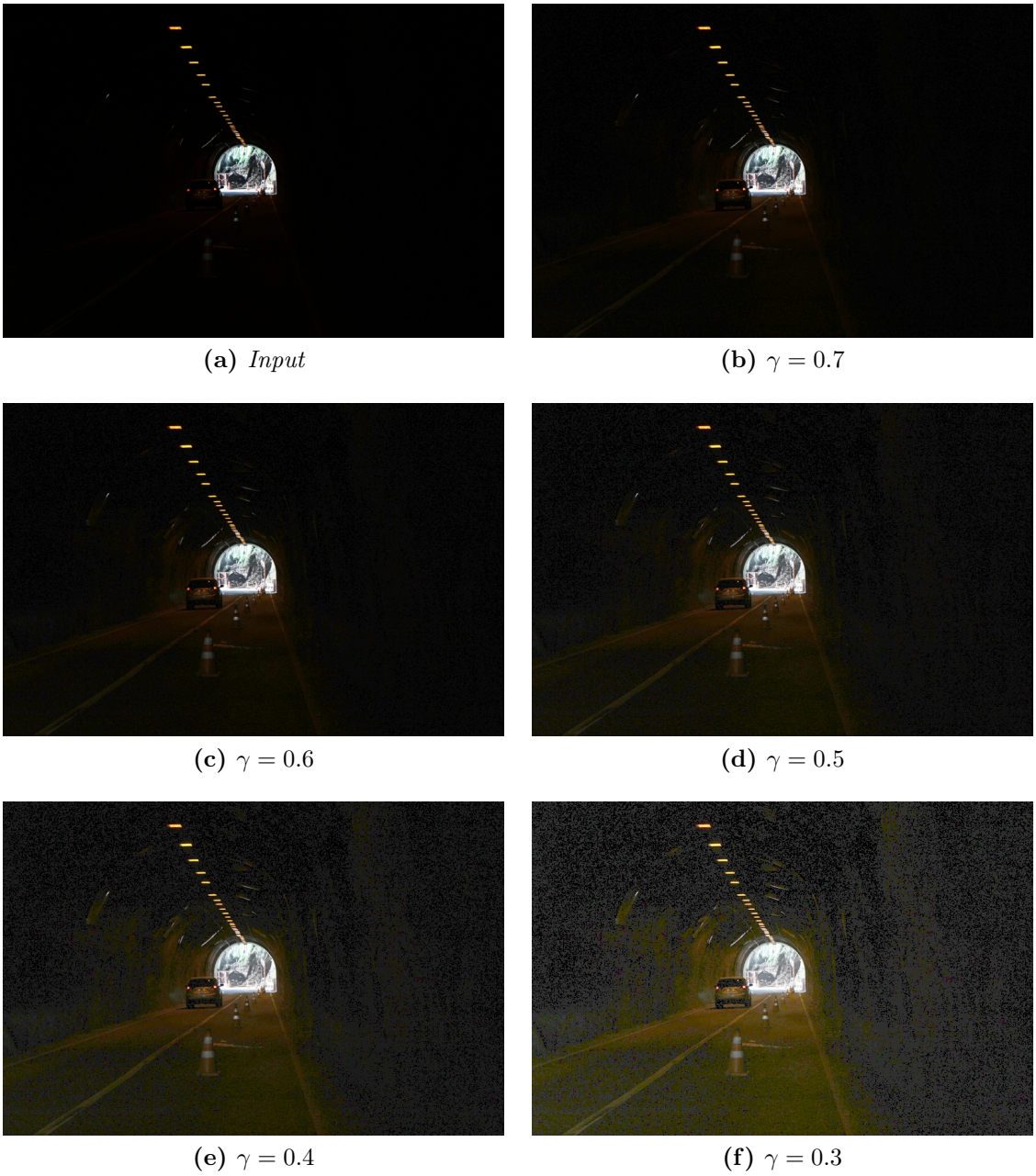
(c) +2 *EV* **(d)** +3 *EV*

(e) +4 *EV* **(f)** +5 *EV*

Figure 5.7: *An image set of daytime tunnel exit showing influence of EV adjustment by controlling shutter duration alone (*$f/5.6$*, ISO:800).*

When a smaller γ value is selected, the dark region in the frame is more intensively enhanced. Note that the input image Figure 5.8(a) is captured in still format to preserve the details in dark region. However, when applying gamma correction to video sequence under exposure compensation, the input signal V_{in} has to be the raw data from the image sensor before the encoding process.

For different daytime tunnel scenes, the value of γ has to be properly set so that the whole frame, which is captured with exposure compensation, is visually pleasant after



(a) *Input* **(b)** $\gamma = 0.7$

(c) $\gamma = 0.6$ **(d)** $\gamma = 0.5$

(e) $\gamma = 0.4$ **(f)** $\gamma = 0.3$

Figure 5.8: *Enhancement of dark region in a captured still image by gamma correction (γ = 0.7, 0.6, 0.5, 0.4, and 0.3).*

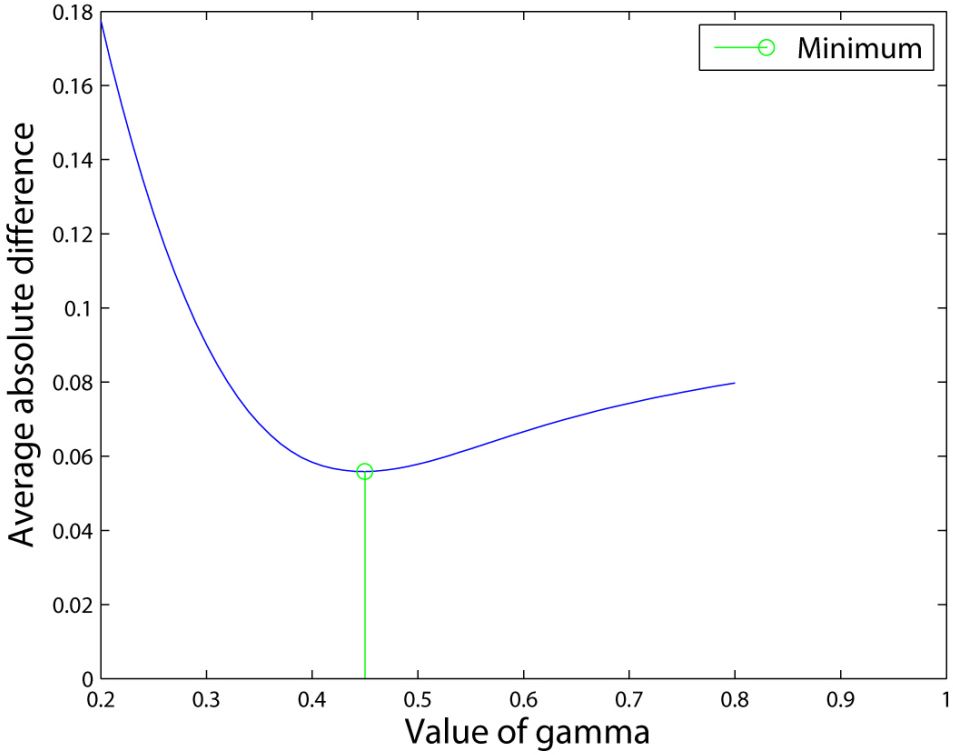
gamma correction. So, the value of γ is adjusted by minimizing the difference between the the gamma-corrected frame and another historical frame captured just before the exposure compensation. In order to focus on the global viewing effects, both the gamma-corrected frame and the historical frame are compared after blurring their details. This process for determining the value of γ has five main steps. First, the initial value of γ is set. Second, the frame captured with exposure compensation is processed by gamma correction. Third, the gamma-corrected frame and the historical frame are blurred using

circular averaging filter. Fourth, the average absolute difference between the two blurred frames is calculated, and the minimum value of the average absolute difference is updated. Fifth, if there is another γ value for comparison, the process jumps to the second step after setting the γ value. For a set of γ values, the average absolute difference between the corresponding set of gamma-corrected frames (after blurring) and a historical reference frame (after blurring) is given in Figure 5.9. Figure 5.9(a) shows the original frame before gamma correction, and Figure 5.9(b) shows the historical reference frame. Both Figure 5.9(a) and Figure 5.9(b) are captured in still format by manually setting the camera parameters. When the value of γ is set as excessively large (close to 1) or small (close to 0), the average absolute difference shows an increasing tendency. And the optimal γ value, which results in the minimal average absolute difference, is closely related with EV applied in exposure compensation. The gamma-corrected frame using the optimal γ value from Figure 5.9(c) is presented in Figure 5.10 together with the histogram. Note that the histograms in Figure 5.10(b)(d)(f) are derived from the converted gray images.

The color tone of the dark region of a frame may be distorted by the concurrently enhanced quantization noise. An example of the color distortion caused by less quantization level is shown in Figure 5.11. Figure 5.11(a) is a standard color checker quantized with 8-bit in each of the three color channels, and Figure 5.11(b) shows another version quantized with 2-bit in each of the three color channels. And the color tone in Figure 5.11(b) is greatly damaged, which may also happen when the scene is captured with less exposure.

One important phenomenon causing poor image quality is the electronic noise generated in the image capturing process. Various methods have been developed for image noise estimation [47][48][49] and image denoising [50][51][52][53][54]. On the other hand, electronic noise is possible to be used for alleviating quantization noise, which is similar to the dithering technique used in the signal processing field. The benefit by bringing in electronic noise is that the uncertainty of the noise after quantization provides cues for recovering the color or details in the dim region. When the color checker in Figure 5.11 is quantized after added a certain level of uniformly distributed noise, the color distortion is made less for human observers as shown in Figure 5.12(a). Also, when the pixel values inside each block of the color checker is averaged, the color inside each block is well recovered. Also, the methodologies like SR (stochastic resonance) [55] and DSR

(a) (b)



(c)

Figure 5.9: *The change of average absolute difference with respect to value of γ. (a) The original frame before gamma correction. (b) The historical reference frame. (c) Average absolute difference for varying γ value.*

(dynamic stochastic resonance) [56][57] also show the potential of internal noise for image enhancement purpose.

The vision enhancement process of IBR on imaging device is summarized in Figure 5.13. For the captured frame, the IBR location is detected, and the number of over-exposed pixels is monitored. When this number exceeds a pre-defined threshold in N out of M captured frames, the exposure compensation (adjustment by EV) is triggered. From the point at which exposure compensation is triggered, the captured frame is enhanced

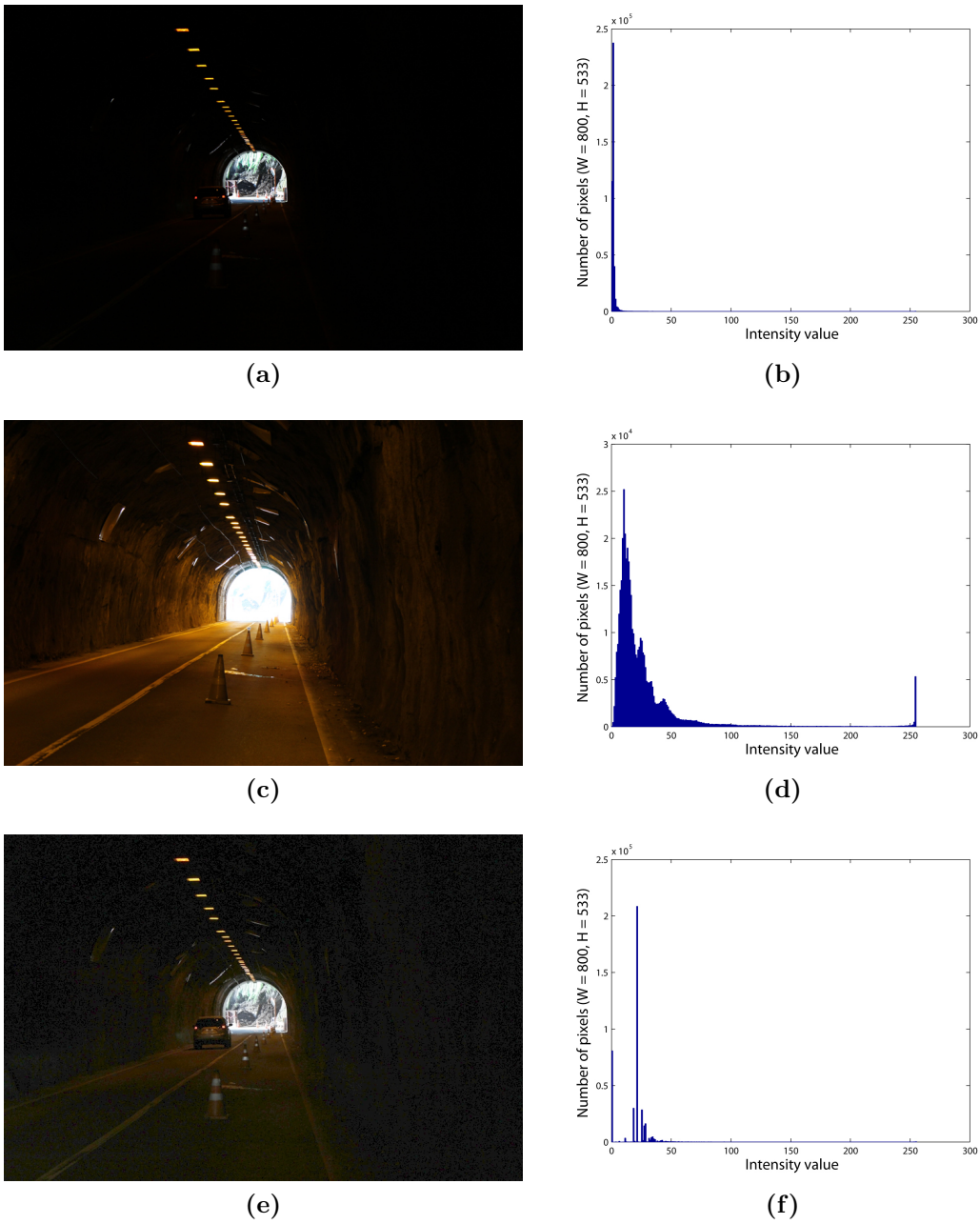
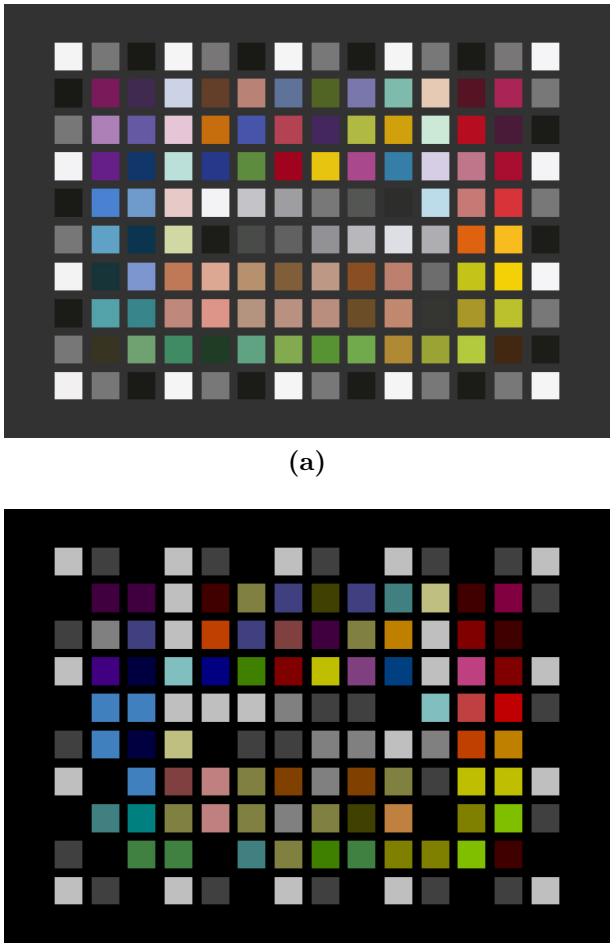


Figure 5.10: *The gamma-corrected frame using the optimal γ value from Figure 5.9(c) ($\gamma = 0.45$) together with the histogram. (a) The original frame before gamma correction. (b) The histogram of the original frame before gamma correction. (c) The historical reference frame. (d) The histogram of the historical reference frame. (e) The gamma-corrected frame using the optimal γ value. (f) The histogram of the optimal gamma-corrected frame.*

by gamma correction before the encoding process. Note that the exposure compensation is realized by sending a EV bias instruction to the image sensor, and the determination of shutter duration and signal gain is always accomplished by the image sensor itself.



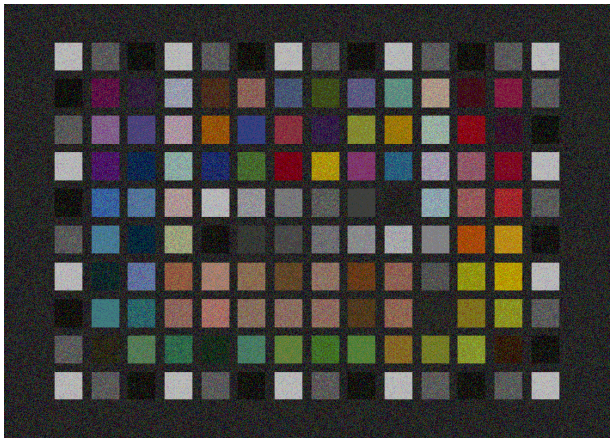
(a)

(b)

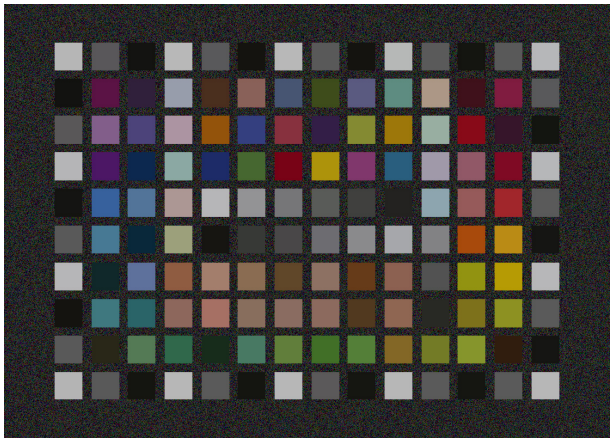
Figure 5.11: *A color checker quantized with 8-bit and 2-bit. (a) Quantized image of 8-bit. (b) Quantized image of 2-bit.*

5.2 Vision Enhancement of Informative Bright Region by Liquid Crystal Glasses

Apart from vision enhancement on image sensors, the information of IBR location is also possible to be applied to enhancement of human vision. When the size of over-exposed region in a daytime tunnel scene achieves a certain degree, the light energy casted on unit area of light-sensitive tissue in human eye is reduced by some device with functionality like light transmittance adjustment. For this purpose, liquid crystal glasses are taken as the interface for enhancing human vision in daytime tunnel scenes using the location information given by the IBR detection algorithm. The liquid crystal glasses have two basic layers. One layer is polarizing film, and the other layer is liquid crystal panel. When the voltage is applied on the liquid crystal panel, the liquid crystal molecules are arranged in one direction so that the light wave passing the panel is waving in the direction



(a)



(b)

Figure 5.12: *A color checker quantized with 2-bit after adding uniformly distributed noise. (a) Quantized image of 2-bit with added noise. (b) Recovered image by averaging the pixels in each block.*

perpendicular to the polarizing film, which causes less light to be transmitted through the glasses. By switching the applied voltage on liquid crystal glasses between "on" and "off", the light transmittance changes between "low" and "high". The two basic states of liquid crystal glasses are presented in Figure 5.14 together with the applied voltage signal.

The connection of hardware is shown in Figure 5.15. The single-board computer runs the IBR detection algorithm, and the liquid crystal glasses is switched to the low transmittance state when the over-exposed region reaches a predefined size.

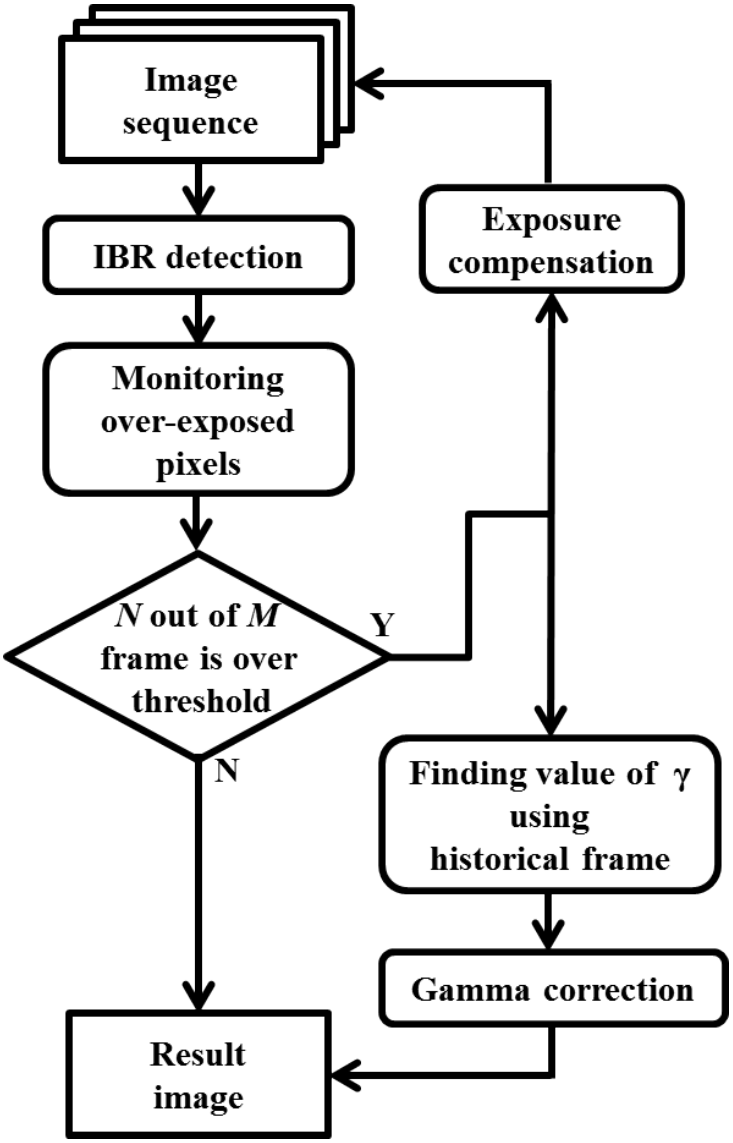
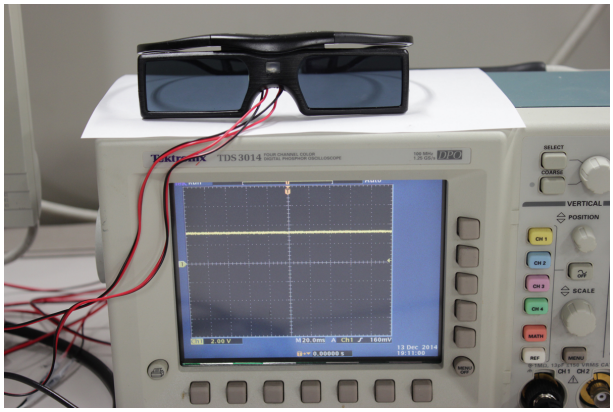


Figure 5.13: *The vision enhancement process of IBR on imaging device.*

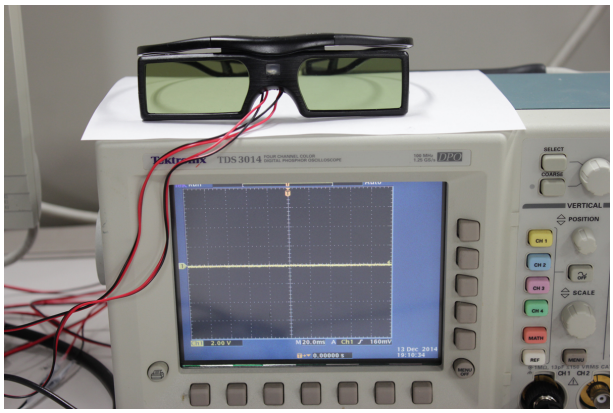
5.3 Experiments of Enhancing Tunnel Scenes on General Roads

A partial experiment is carried out to show the applicability of the proposal on low cost platforms. The IBR detection algorithm is ported into a single-board computer (Raspberry Pi Model B) equipped with an ARM11 processer (700MHz), 512MB of memory, and a Linux operation system. The single-board computer has a camera module attached with it by Camera Serial Interface (CSI-2). The single-board computer and the camera module are fixed on a plastic box as shown in Figure 5.16.

The relation between the ported IBR detection algorithm and the function modules of



(a) *low transmittance*



(b) *high transmittance*

Figure 5.14: *The light transmittance of the liquid crystal glasses controlled by the applied voltage.*

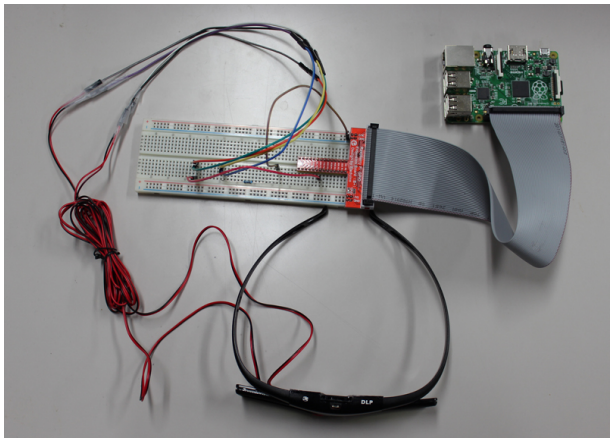


Figure 5.15: *The hardware connection of single-board computer and liquid crystal glasses.*

the camera is abstracted and shown in Figure 5.17. The image sequence captured by the camera module passes through the video port of the camera component, and is encoded with H.264 standard by the encoder component. The encoded video stream is saved on a storage media (SD card). In the meantime, the algorithm for locating IBR is running

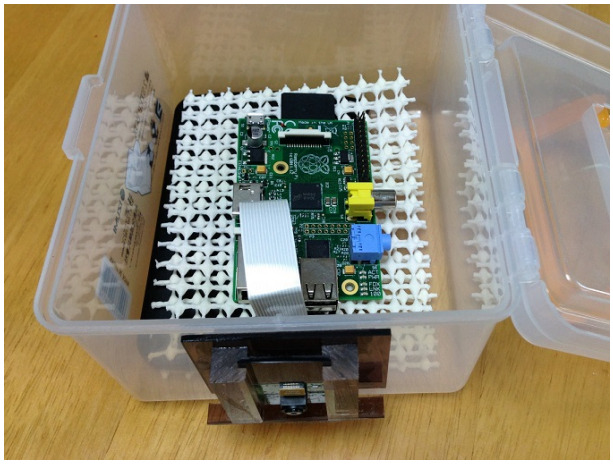


Figure 5.16: *The single-board computer (Raspberry Pi Model B) and its camera module fixed on a plastic box.*

on the image sequence from the preview port of the camera component. The size of the output image from both the video port and the preview port is 320×240. And exposure compensation is triggered by monitoring the size change of over-exposed region (IBR) using the strategy described in Section 5.1.1.

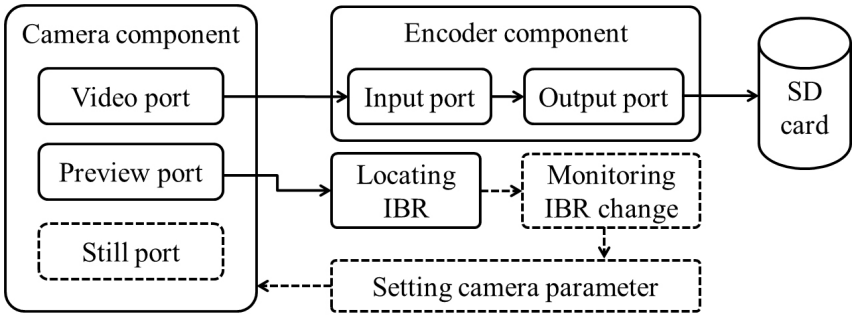
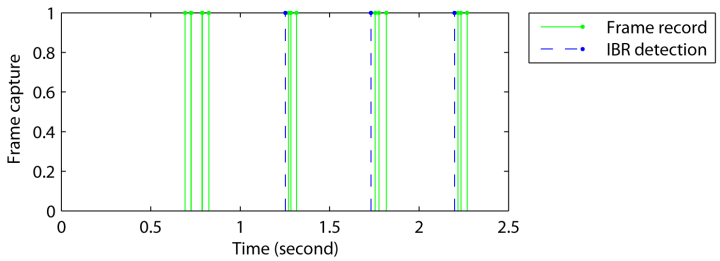


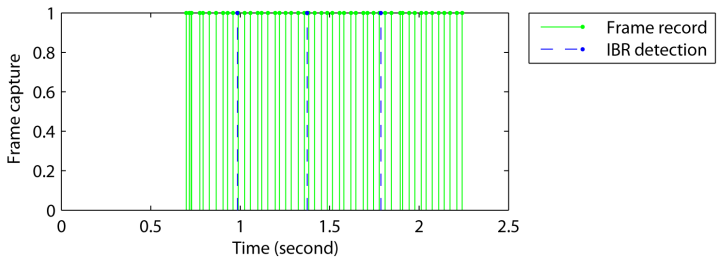
Figure 5.17: *The relation between the IBR locating algorithm and function modules of the camera.*

The time duration for encoding each frame (on the output port of the encoder component) and for IBR detection and exposure compensation triggering (on the preview port of the camera component) are given in Figure 5.18 under both serial and parallel processing strategies. Note that more computational time is consumed to locate IBR on each frame than to encode the captured frame. And, more frames are encoded for unit time interval when the two functionalities (frame encoding and IBR detection) are processed in parallel.

The device in Figure 5.16 is mounted on a vehicle and runs the algorithm ported on it



(a) *serial processing*



(b) *parallel processing*

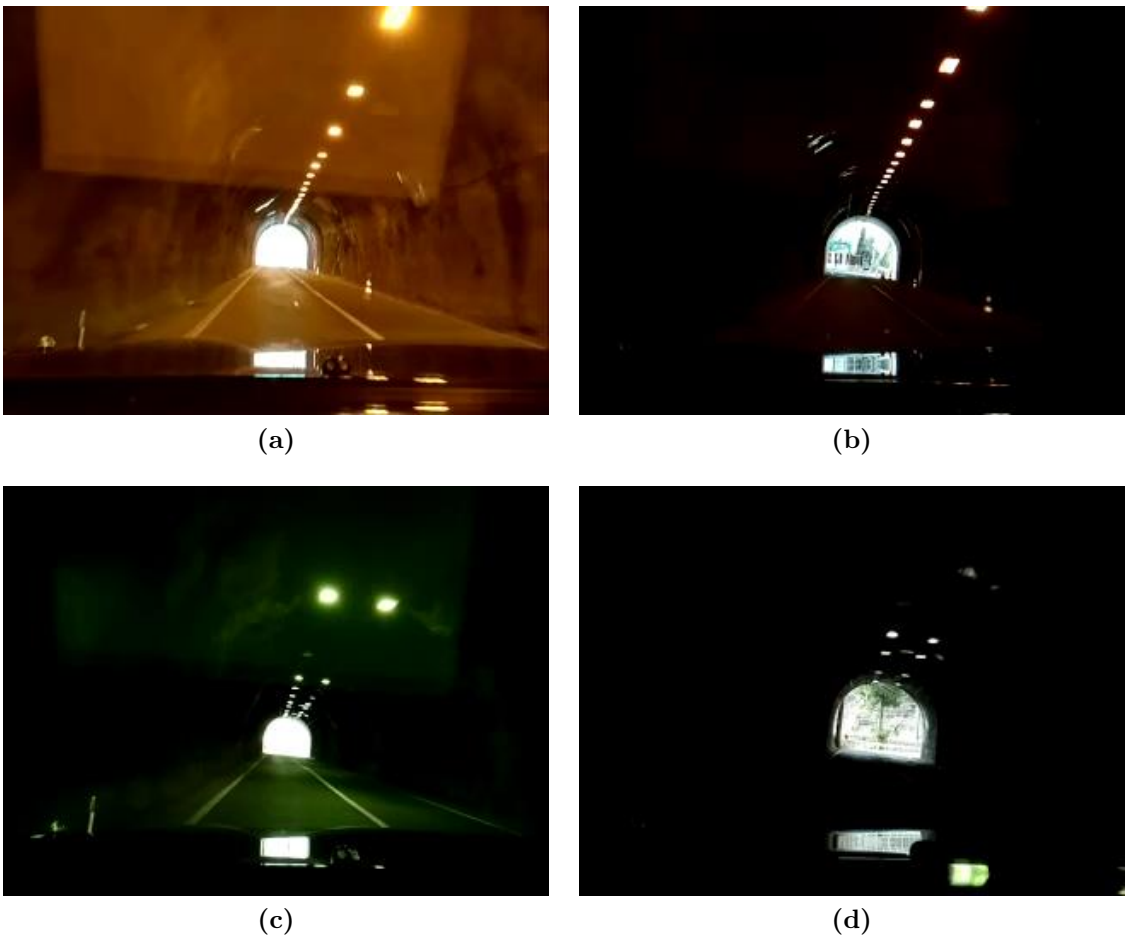
Figure 5.18: *The time duration for encoding each frame (on the output port of the encoder component) and for IBR detection and exposure compensation triggering (on the preview port of the camera component).*

when driving through tunnels on general roads. As is briefly stated, the ported algorithm includes three main functionalities: IBR detection, exposure compensation triggering, and frame encoding. For each of two typical tunnel scenes, two frames captured before and after the triggering of exposure compensation are shown in Figure 5.19. Note that the gamma correction process described in Section 5.1.2 is not applied in this experiment due to the limitation of software interface designed for the camera module. When exposure compensation is triggered, the parameter setting of the image sensor is more adapted to the tunnel exit region. And the visual information, which is previously lost due to over-exposure, is clearly visible.

Another partial experiment is carried out to demonstrate the vision enhancement method for IBR using liquid crystal glasses described in Section 5.2. The state change of the liquid crystal glasses together with the IBR detection results is shown in Figure 5.20 for two recorded video sequences of daytime tunnel scenes.

5.4 Chapter Summary

A method is proposed for visually enhancing the IBR on imaging device. Exposure compensation of the imaging device is triggered by the over-exposed pixels in searching

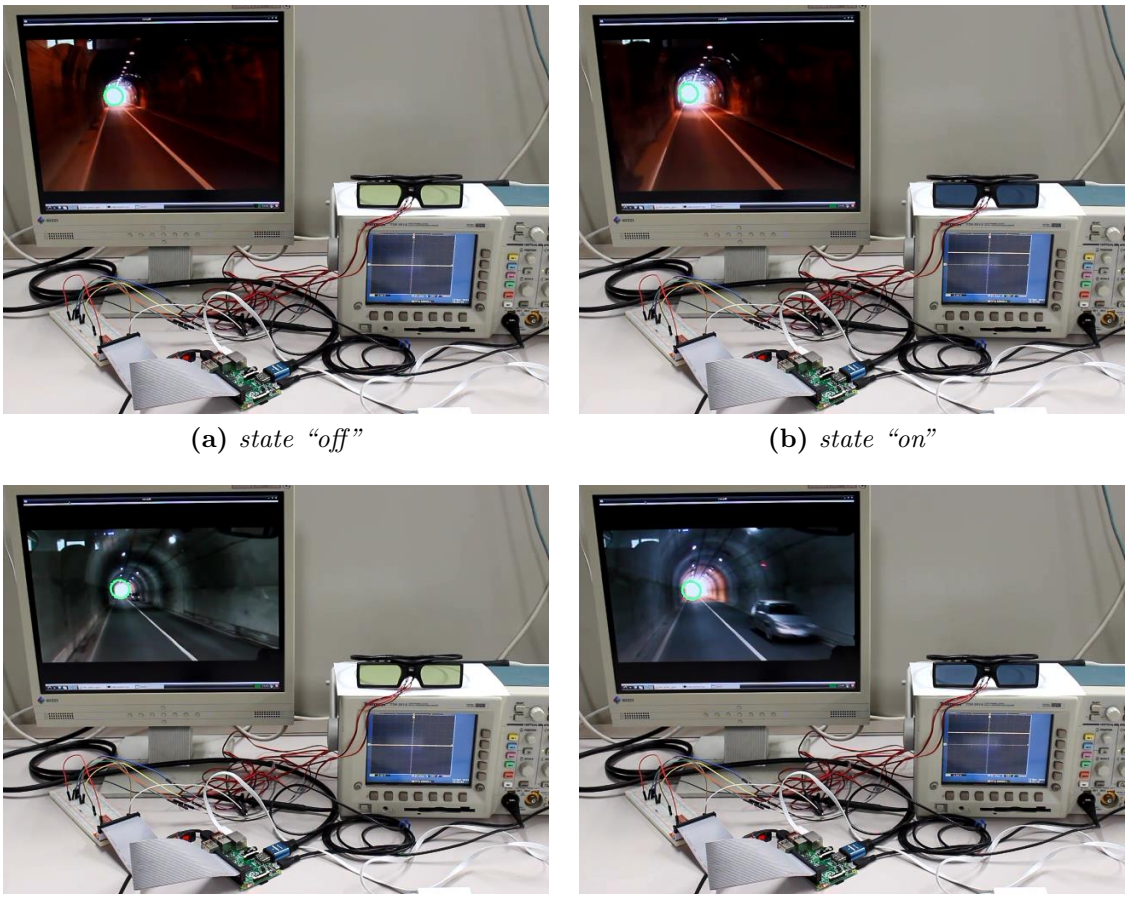


(a) **(b)**

(c) **(d)**

Figure 5.19: *Two frames captured before and after the triggering of exposure compensation (+4 EV) in two typical tunnel scenes.(a)(c) Before exposure compensation. (b)(d) After exposure compensation.*

area centered at the IBR location. And the dark region in the captured frame after exposure compensation is then enhanced by gamma correction. Another vision enhancement method is also introduced for human drivers by controlling the transmittance of liquid crystal glasses with the IBR information. Experimental results shows that the IBR detection algorithm is running properly on a single-board computer for the purpose of triggering the exposure compensation. When the IBR detection algorithm runs in parallel with the encoding process of frames, more frames are recorded in unit time interval. A partial experiment also shows that the transmittance of liquid crystal glasses is possible to be controlled by monitoring the change of the IBR.



(a) *state "off"* **(b)** *state "on"*

(c) *state "off"* **(d)** *state "on"*

Figure 5.20: *The state change of the liquid crystal glasses together with the IBR detection results for two video sequences of daytime tunnel scenes.*

Chapter 6

Conclusion

An NPR method, which focuses on the gradient value of image source, is proposed for rendering HDR images to mimic the visual perception of human artists and designers. Experimental results obtained from simulating the photo capture process show that an HDR image reduces averaged angular error of gradient values by 57.5% for a designed Gaussian pattern compared to that of the generated conventional 8-bit digital images. Preservation of details in both over-exposed and under-exposed regions of the scene is easily observed in rendered images. This detail preservation property of HDR images in NPR processes is crucial to reflecting human perception in visual information. Structures of objects in some scenes are well maintained when the proposal is applied to decorative mosaic rendering; while for conventional 8-bit digital images, there are some distortions in shapes of the rendering results. Rendering results for some HDR images appear dark compared to those of corresponding conventional digital images, which is mainly due to the tone mapping method used for compressing the dynamic range of HDR images. The final performance of the proposal is greatly influenced by the scene to be rendered. If the luminance condition changes rapidly in different regions of the scene to be rendered, then the proposal surely improves the rendering results. Furthermore, the evaluation experiment using the visual attention model shows that new salient regions are formed in rendering results for HDR images.

For designers and artists who seek approaches to more informative illustrations with newly formed salient regions that attract visual attention, potential rendering tools can be extended from the proposal. How gradient values of HDR images may influence other features such as colors and textures in an NPR process are, moreover, interesting topics to be studied further.

To locate IBR in which visual information is missing due to limited dynamic range of image sensor, an algorithm is proposed via combining geometric properties of visual cues into a confidence map. The experimental results using the tunnel scene dataset show that the locations of IBRs given by the algorithm have averaged error of 4.14 pixels out of 640 pixels on X-axis and 4.23 pixels out of 480 pixels on Y-axis in comparison with the IBRs labeled by the human observers. For varying tunnel scenes with different structures and lighting conditions on local roads, the averaged location error ranges from 4.03 to 4.28 pixels on X-axis and from 3.08 to 4.98 pixels on Y-axis. The view angle compensation step makes the location error on X-axis reduce to the same level of that on Y-axis by 74.6%. The average processing speed of the algorithm on one frame (640×480) is around 10ms for varying tunnel scenes in the dataset.

The accuracy (within 1% on each axis) of the algorithm is appropriate for image sensor control due to the fact that shutter time is supposed to be made shorten when the IBR covers a certain portion of whole frame. The processing speed (around 10ms per frame) of algorithm is also able to keep pace with the frame rate (usually 30 frames per second) of conventional cameras.

The low computation cost of the algorithm supports its application in devices with limited computing power like vehicle driving recorder. The algorithm aims at providing good timing for shutter and gain control of image sensor inside these devices so that the critical visual information inside the IBR are made visible at an earlier stage, which contributes to earlier record of over-exposed region of tunnel exits and enhancement of driving safety.

Based on the IBR location detection algorithm, a method is proposed for visually enhancing the IBR on imaging device. Exposure compensation of the imaging device is triggered by monitoring the change of the over-exposed pixels inside the searching area centered at the IBR location. And the dark region in the captured frame after exposure compensation is then enhanced by gamma correction. Another vision enhancement method is also introduced for human drivers by controlling the transmittance of liquid crystal glasses with the IBR information. Experimental results shows that the IBR detection algorithm is running properly on a single-board computer for the purpose of triggering the exposure compensation. When the IBR detection algorithm runs in parallel with the encoding process of frames, more frames are recorded in unit time interval. A partial

experiment also shows that the transmittance of liquid crystal glasses is possible to be controlled by monitoring the change of the IBR.

Apart from the daytime tunnel scene, car-mounted imaging devices also suffer from other road scenes with undesirable luminance conditions. For example, the imaging device meets the phenomenon of veiling glare when it is moving toward a direction with strong light sources. Under such luminance conditions, the visual cues (e.g., marking of traffic lanes and light sources) is possible to be fused for setting the parameters of an image sensor (i.e., shutter duration, aperture size, and signal gain). The same strategy may also be applied in road scenes with both strongly lit regions and weakly lit regions. For road scenes with differing luminance condition, several sets of image sensor parameters are provided based on the fused visual cues rather than the illuminance on the image sensor. Thus, visual information of different scene regions may be captured using image sensor parameters related with the visual cues.

For extending the flexibility of image sensor in scene understanding applications, the image sensor parameters determined from visual cues are necessary supplementary information for each captured frame. From the low-level image features, visual cues have to be selected through a balance between the computational cost and the coverage of varying scenes. From the perspective of practical application, the acceleration of the visual cue computing and fusing process have to be real-time capable. The development of customized hardware (e.g., FPGA) for accelerated visual cue computing and image sensor controlling may attach practical value to the proposal in various fields (e.g., autonomous vehicles and service robots).

Since the visual cue based approach sets its target on the information acquisition step of different vision applications, the image sensor is possible to capture the scene under content-related parameter setting. Considering the need of an extremely large image set in the training process of well-designed deep learning models, the obtainable scene details are extended if the proposal is further developed and applied in the visual information acquisition process.

Bibliography

[1] T. Strothotte and S. Schlechtweg. *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*. Morgan Kaufmann, 2002.

[2] A. Hertzmann. Painterly rendering with curved brush strokes of multiple sizes. In *Proc. of SIGGRAPH 98*, pages 453–460, 1998.

[3] G. Winkenbach and D. H. Salesin. Computer-generated pen-and-ink illustration. In *Proc. of SIGGRAPH 94*, pages 91–100, 1994.

[4] W. Zhang, X. An, and S. Pan. An improved recursive retinex for rendering high dynamic range photographs. In *2011 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR 2011)*, pages 121–125, 2011.

[5] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of SIGGRAPH 97*, pages 369–378, 1997.

[6] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, 2002.

[7] C. Schlick. Quantization techniques for visualization of high dynamic range pictures. In *Proc. of the 5th Eurographics Rendering Workshop*, pages 7–20, 1994.

[8] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, 2002.

[9] M. Colbert, E. Reinhard, and C. E. Hughes. Painting in high dynamic range. *J. Visual Comm. and Image R.*, 18:387–396, 2007.

[10] R. Vergne and X. Granier. Stylisation d'objets éclairés par des cartes d'environnement HDR. In *20èmes Journées de l'Association Française d'Informatique Graphique (AFIG)*, 2007.

[11] K. Smith and G. Krawczyk. NPR for HDR: Stylizing with high dynamic range photographs. In *Proc. of the 4th International Symposium on Non-Photorealistic Animation and Rendering (NPAR 2006)*, 2006.

[12] A. Hausner. Simulating decorative mosaic. In *Proc. of SIGGRAPH 01*, pages 573–578, 2001.

[13] Y. Wang, E. K. Teoh, and D. Shen. Lane detection using B-snake. In *IEEE Int. Conf. on Information, Intelligence, and Systems (ICIIS 99)*, pages 438–443, 1999.

[14] Z. Kim. Robust lane detection and tracking in challenging scenarios. *IEEE Trans. on Intelligent Transportation Systems*, 9(1):16–26, 2008.

[15] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Proc. of IEEE Intelligent Vehicles Symposium*, pages 255–260, 2005.

[16] A. Ruta, Y. Li, and X. Liu. Real-time traffic sign recognition from video by class-specific discriminative features. *Pattern Recognition*, 43(1):416–430, 2010.

[17] P. Stone, P. Beeson, T. Merili, and R. Madigan. DARPA urban challenge technical report - Austin Robot Technology. `http://archive.darpa.mil/grandchallenge/`, 2007.

[18] J. Lu, F. Dong, and K. Hirota. Gradient-related non-photorealistic rendering for high dynamic range images. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 17(4):628–636, 2013.

[19] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319325, 2003.

[20] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile HDR video production system. *ACM Transactions on Graphics*, 30(4):Article No.41, 2011.

[21] Y. Tamura and N. Mikami. Study on traffic accidents on tunnel zone. *Yamaguchi University Faculty of Engineering Research Report*, 55(2):73–78, 2004.

[22] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications, PG '07*, pages 382–390, 2007.

[23] J. J. McCann. Art, science, and appearance in HDR. *Journal of the Society for Information Display*, 15(9):709719, 2007.

[24] C. E. K. Mees. *The Fundamentals of Photography.* Kodak Research Laboratory, 1920.

[25] C. E. K. Mees. *From Dry Plates to Ektachrome Film:* A *Story of Photographic Research.* Ziff-Davis Publishing Company, 1961.

[26] A. Adams. *The Negative.* New York Graphical Society, 1981.

[27] J. J. McCann and A. Rizzi. Camera and visual veiling glare in HDR images. *Journal of the Society for Information Display*, 15(9):721730, 2007.

[28] L. Wang, X. Wang, P.-P. Sloan, L.-Y. Wei, X. Tong, and B. Guo. Rendering from compressed high dynamic range textures on programmable graphics hardware. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games (I3D 07)*, pages 17–24, 2007.

[29] M. N. Inanici. Evaluation of high dynamic range photography as a luminance data acquisition system. *Lighting Research and Technology*, 38(2):123–136, 2006.

[30] S. W. Hasinoff, F. Durandand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 553–560, 2010.

[31] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Transactions on Graphics*, 23(3):760–768, 2004.

[32] E. Land and J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.

[33] J. Frankle and J. J. McCann. Method and apparatus for lightness imaging (u.s. patent no.4384336), 1983.

[34] L. Meylan and S. Susstrunk. High dynamic range image rendering with a retinex-based adaptive filter. *IEEE Transactions on Image Processing*, 15(9):2820–2830, 2006.

[35] H. Graf, C. Harendt, T. Engelhardt, C. Scherjon, K. Warkentin, H. Richter, and J. N. Burghartz. High dynamic range cmos imager technologies for biomedical applications. *IEEE Journal of Solid-State Circuits*, 1(1):281–289, 2009.

[36] A. Santella and D. DeCarlo. Visual interest and NPR: An evaluation and manifesto. In *Proc. of the 4th International Symposium on Non-Photorealistic Animation and Rendering (NPAR 2004)*, pages 71–78, 2004.

[37] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207, 2013.

[38] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19:1395–1407, 2006.

[39] Saliency Toolbox. `http://www.saliencytoolbox.net/`, 2008.

[40] J. Matas, C. Galambos, and J. Kittler. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding*, 78(1):119–137, 2000.

[41] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang. Probabilistic exposure fusion. *IEEE Transactions on Image Processing*, 21(1):341–357, 2012.

[42] `http://en.wikipedia.org/wiki/Exposure_(photography)`.

[43] `http://en.wikipedia.org/wiki/Exposure_value`.

[44] J. Tang, E. Peli, and S. Acton. Image enhancement using a contrast measure in the compressed domain. *IEEE Signal Processing Letters*, 10(10):289–292, 2003.

[45] J. Mukherjee and S. K. Mitra. Enhancement of color images by scaling the dct coefficients. *IEEE Transactions on Image Processing*, 17(10):1783–1794, 2008.

[46] `http://en.wikipedia.org/wiki/Gamma_correction`.

[47] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang. Noise estimation from a single image. In *2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR06)*, pages 901–908, 2006.

[48] A. Amer and E. Dubois. Fast and reliable structure-oriented video noise estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1):1–6, 2005.

[49] S. Tai and S. Yang. A fast method for image noise estimation using Laplacian operator and adaptive edge detection. In *3rd International Symposium on Communications, Control and Signal Processing (ISCCSP08)*, pages 1077–1081, 2008.

[50] A. Buades, B. Coll, and J. M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008.

[51] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003.

[52] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

[53] J. L. Starck, E. J. Candes, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2002.

[54] S. G. Chang, B. Yu, and M. Vetterli. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 9(9):1532–1546, 2000.

[55] E. Simonotto, M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss. Visual perception of stochastic resonance. *Physical Review Letters*, 78(6):1186–1189, 1997.

[56] R. K. Jha, R. Chouhan, K. Aizawa, and P. K. Biswas. Dark and low-contrast image enhancement using dynamic stochastic resonance in discrete cosine transform domain. *APSIPA Transactions on Signal and Information Processing*, 2:e6, 2013.

[57] R. Chouhan, R. K. Jha, and P. K. Biswas. Noise-enhanced contrast stretching of dark images in SVD-DWT domain. In *2013 Annual IEEE India Conference (INDICON)*, pages 1–6, 2013.

Related Publications

- Journal Papers

  J1 Jiajun Lu, Fangyan Dong, and Kaoru Hirota, *Gradient-related Non-Photorealistic Rendering for High Dynamic Range Images*, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.17, No.4, pp.628-636, 2013.

  J2 Jiajun Lu, Fangyan Dong, and Kaoru Hirota, *Location Detection of Informative Bright Region in Tunnel Scenes Using Lighting and Traffic Lane Cues*, Journal of Advanced Computational Intelligence and Intelligent Informatics, to appear in Vol.19, No.2, 2015.

- International Conference Papers

  C1 Jiajun Lu, Fangyan Dong, and Kaoru Hirota, *Non-photorealistic Rendering for High Dynamic Range Images*, Fifth International Symposium on Computational Intelligence and Industrial Applications (ISCIIA2012), Session TUb1, 2012.

  C2 Jiajun Lu, Fangyan Dong, and Kaoru Hirota, *Locating Informative Bright Region in Tunnel Scenes using Lighting and Traffic Lane Cues* , Proceedings of The Joint International Conference of ITCA 2014 and ISCIIA 2014 (ITCA&ISCIIA2014), pp.75-82, 2014.