

ZIJIAN ZHU

*High Quality High Dynamic
Range Imaging*

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences
No 113

Academic Dissertation

To be presented by permission of the Faculty of Science and Forestry for public examination in the Louhela auditorium at the University of Eastern Finland, Joensuu, on August, 7, 2013, at 12 o'clock noon.

School of Computing

Kopijyvä Oy
Joensuu, 2013
Editor: Prof. Pertti Pasanen

Distribution:
Eastern Finland University Library / Sales of publications
P.O.Box 107, FI-80101 Joensuu, Finland
tel. +358-50-3058396
<http://www.uef.fi/kirjasto>

ISBN: 978-952-61-1171-1 (Printed)
ISSNL: 1798-5668
ISSN: 1798-5668
ISBN: 978-952-61-1172-8 (PDF)
ISSN: 1798-5676 (PDF)

Author's address: University of Eastern Finland
School of Computing
P.O.Box 111
80101 JOENSUU
FINLAND
email: zhuzj@i2r.a-star.edu.sg

Supervisors: Professor Pasi Fränti, Ph.D.
University of Eastern Finland
School of Computing
P.O.Box 111
80101 JOENSUU
FINLAND
email: pasi.franti@uef.fi

Professor Susanto Rahardja, Ph.D.
Institute for Infocomm Research
Department of Signal Processing
1 Fusionopolis Way, #21-01 Connexis
138632 SINGAPORE
SINGAPORE
email: rsusanto@i2r.a-star.edu.sg

Reviewers: Professor Rafal Mantiuk, Ph.D
Bangor University
School of Computer Science
Dean Street, BANGOR, LL57 1UT
UNITED KINGDOM
email: mantiuk@bangor.ac.uk

Dr. Radu Ciprian Bilcu, Ph.D
Nokia Research Center
Multimodal Sensing and Context
Visiokatu 1
33720 TAMPERE
FINLAND
email: radu.bilcu@nokia.com

Opponent: Dr. Alessandro Foi, Ph.D
Academy of Finland Research Fellow
Department of Signal Processing
Tampere University of Technology
33101 TAMPERE
FINLAND
email: alessandro.foi@tut.fi

ABSTRACT

A *high dynamic range* (HDR) image uses a large bit depth up to 32-bit per pixel per color channel. Due to hardware limitation, it can neither be captured by conventional camera in a single photo, nor be displayed on a conventional monitor. In this thesis, a software solution of HDR imaging is introduced from the synthesis to the display of high quality HDR images.

Contrary to conventional imaging method, HDR imaging uses multiple exposures. Therefore, the variance of the same pixel in different images is non-linear and difficult to measure. To solve this, a new *inter-pixel relationship function* (IRF) is proposed using both spatial and temporal correlations. It is widely used in many HDR imaging components.

Multiple input images produce a new ghosting artifact due to moving object. A real time de-ghosting method is first proposed using bi-directional comparison and IRF based moving object detection and patching. It is lightweight in terms of both time complexity and physical memory consumption, which makes it suitable for mobile devices. We further extend it by merging the IRF with a histogram intensity mapping and adopting a new threshold model based on statistical study. The extended de-ghosting scheme is more robust in the scenarios where the moving object occupies larger areas.

Noise in the input images will be kept in the synthesized HDR image. A 2D de-noising factor is proposed for the synthesis of 32-bit HDR image, and a noise reduced tone-mapping is proposed to map the 32-bit HDR image back to 8-bit for the display based on *information content weighting* (ICW). Both methods have low time complexity.

It is also important to measure if the input images are from the same scene, as huge artifacts may appear if a wrong image is involved in the synthesis process. We proposed a structure similarly based metric which is robust to the images with different exposures. We also proposed a metric for the measurement of two 32-bit HDR images based on their histograms of radiance maps.

AMS Classification: 68U10; 94A08

Universal Decimal Classification: 004.932

*Library of Congress Subject Headings: High dynamic range imaging;
Computational photography; Image enhancement - Methods; Image systems -
Image quality*

Yleinen suomalainen asiasanasto: digitaalikuvaus; kuvankäsittely

Acknowledgements

The efforts of my PhD studies span from Singapore to Finland crossing more than 9000 kilometers. It would not have been possible without the kind support and help of many people. And I would like to extend my sincere thanks to all of them.

I would like to express my great appreciation to Prof. Pasi Fränti, my supervisor, for his support and guidance over the years. His willingness to give his time and constructive suggestions so generously has been very much appreciated.

I would like to express my deep gratitude to Prof. Susanto Rahardja, my co-supervisor, for his strategic suggestions during the planning and development of my research work. He is also the one who provided me a connection from Singapore to Finland.

I would like to thank Prof. Mantiuk and Dr. Bilcu, the reviewers of the thesis, for their helpful comments, and Dr. Foi, my opponent, for the valuable time.

Besides, I wish to thank my colleagues in Image Processing Group in Institute for Infocomm Research for co-authoring the papers. Particularly, I would like to thank to Dr. Zhengguo Li for his great suggestions, inspirations and critiques. Nevertheless, my grateful thanks are extended to the Speech and Image Processing Unit in the University of Eastern Finland for their help in offering me the resources during my trips in Finland.

Thanks and appreciation to the support from the University of Eastern Finland, Institute for Infocomm Research and Nokia Foundation.

Last but not least, I would like to express my special thanks to my family for their understanding and support.

Singapore Jun 25th 2013

Zijian Zhu

LIST OF ABBREVIATIONS

CCD	Charge-Coupled Device
CRF	Camera Response Function
CMOS	Complementary Metal–Oxide–Semiconductor
DR	Dynamic Range
EV	Exposure Value
HDR	High Dynamic Range
HVS	Human Visual System
ICW	Information Content Weighting
IFC	Information Fidelity Criterion
IMF	Intensity Mapping Function
IRF	Inter-pixel Relationship Function
ISO	Light Sensitivity Setting
LDR	Low Dynamic Range
PSNR	Peak Signal-to-Noise Ratio
SNR	Signal-to-Noise Ratio
WLS	Weighted Least Squares

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following selection of the author's publications in the field of high dynamic range imaging:

- [P1] **Z. J. Zhu**, Z. G. Li, S. Rahardja, and P. Fränti, "Recover real world scene: high quality image inpainting using multi-exposed references", *Electronics Letters*, pp. 1310-1312, Vol. 45, Issue 25, Dec 2009.
- [P2] Z. G. Li, **Z. J. Zhu**, and S. Rahardja, "Structural similarity indices for high dynamic range imaging", *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1145-1148, Prague, Czech Republic, May 2011.
- [P3] **Z. J. Zhu**, S. Rahardja, Z. G. Li, P. Fränti, "Detecting and composing near-identical HDR images without exposure information", *IEEE International Conference on Image Processing*, pp. 3337-3340, Hong Kong, Sep 2010.
- [P4] **Z. J. Zhu**, S. Rahardja, S.Q. Wu, and P. Fränti, "Real-time ghost removal for composing high dynamic range images", *IEEE Conference on Industrial Electronics and Applications*, pp. 1627-1631, Taichung, Taiwan, Jun 2010.
- [P5] **Z. J. Zhu**, Z. G. Li, S. Rahardja, and P. Fränti, "De-ghosting of HDR images with double-credit intensity mapping", *IEEE International Conference on Image Processing*, pp. 1361-1364, Brussels, Belgium, Sep 2011.
- [P6] Z. G. Li, **Z. J. Zhu**, and S. Rahardja, "Fast movement detection for high dynamic range imaging", *IEEE International Conference on Image Processing*, pp. 373-376, Brussels, Belgium, Sep 2011.
- [P7] **Z. J. Zhu**, Z. G. Li, S. Q. Wu, J. H. Zheng, W. Yao, P. Fränti, and S. Rahardja, "Noise reduced high dynamic range tone-mapping based on information content weighting", Manuscript
- [P8] **Z. J. Zhu**, Z. G. Li, S. Rahardja, and P. Fränti, "2D denoising factor for high dynamic range imaging", *ACM SIGGRAPH Posters*, pp. 46:1-46:1, Los Angeles, US, 2012.

Throughout the overview, these papers will be referred to as [P1]-[P8]. The papers have been included in this thesis with permission of their copyright holders.

AUTHOR'S CONTRIBUTION

In [P1, P3-P5], the ideas, implementation and experiments are original from the author. The papers were written by the author too. In [P2], author contributed the idea of how to use the intensity mapping for balancing the different dynamic range. In [P6], author contributed to the idea of how to improve the speed for moving object detection, as well as the C implementation on both PC and mobile phone. The HDR de-noising [P7-P8] was inspired by Dr. Li Zhengguo's early work on weighted guided filter and weighted frame average, while the author proposed the methods to solve the problems. At the same time, author was responsible for the implementation, experiments and composing the papers.

Contents

1 Introduction	1
1.1 Dynamic range	2
1.2 High dynamic range acquisition	2
1.3 High dynamic range display	4
1.4 Workflow of software solution of HDR imaging	5
1.5 Structure of the thesis	6
2 Intensity Mapping	7
2.1 Background	8
2.2 Inter-pixel relationship function	9
2.3 Inpainting with exposure bracketing images	11
2.4 Blind exposure value estimation	14
3 De-ghosting	17
3.1 Image registration	17
3.2 Review of de-ghosting methods	18
3.3 Realtime de-ghosting framework	19
3.4 Robust de-ghosting	24
3.5 Motion detection for sub-sampled images	28
4 HDR De-noising	31
4.1 De-noising of LDR images	31
4.2 Noise-reduced HDR synthesis using 2D-denoising factor ...	31
4.3 Noise reduced tone-mapping	37

5	Image Difference Metrics	47
5.1	Image difference metrics for exposure bracketing images ...	47
5.2	Image difference metrics for HDR radiance map	51
6	Summary of Contributions	55
6.1	Contribution of the thesis	55
6.2	Summary of results.....	57
7	Conclusion	67
8	Reference	69

1 Introduction

When a scene with high light contrast is captured by a camera, either the dark area or the bright area will be saturated in the output image, as shown in Fig. 1.1. This is due to the limitation of the camera sensor, and has been existed since the first camera was invented. To solve this, *high dynamic range* (HDR) imaging has been invented by recovering the real world scene using multiple conventional *low dynamic range* (LDR) images [1, 2, 3, 4, 5, 6].

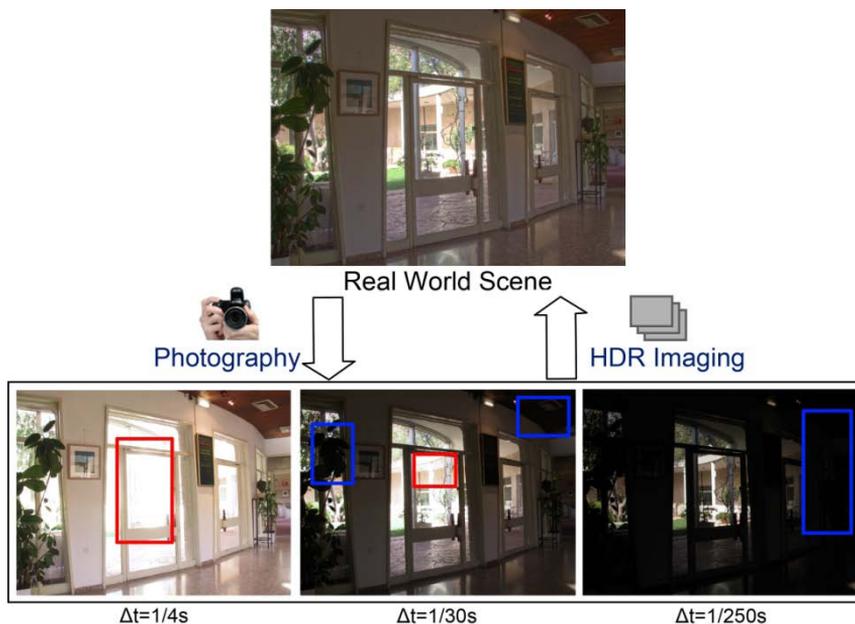


Fig. 1.1 Problem in capturing high dynamic range scene, where (Top) is the real world scene perceived by human eyes, and (Bottom) are the images captured by camera using different exposure times (Δt). Both over-exposed saturation (red square) and under-exposed saturation (blue square) are present, even in the 'best exposed' image (Bottom Middle). The 'real world scene' image was synthesized through HDR imaging by using the bottom 3 images.

1.1 DYNAMIC RANGE

The *dynamic range* (DR) of a scene is defined as the range from the lowest light intensity to the highest light intensity in the scene. It is named as *scene contrast* too. In real world, the full range of light intensities that can be perceived by human vision system spans from as low as star light to as high as the sunshine. The human vision system is capable of perceiving light over a range of 5 orders of magnitude simultaneously. On top of that, the pupil adapts very fast to the changes in lighting conditions, which result to a concurrently perception of nearly 10 orders of magnitude [1, 7]. In comparison to this biological sensor, modern cameras can only capture a small dynamic range of about two orders of magnitude in a single exposure, as shown in Fig. 1.2. Thus, information is lost in the very bright and very dark areas, named *saturated* areas.

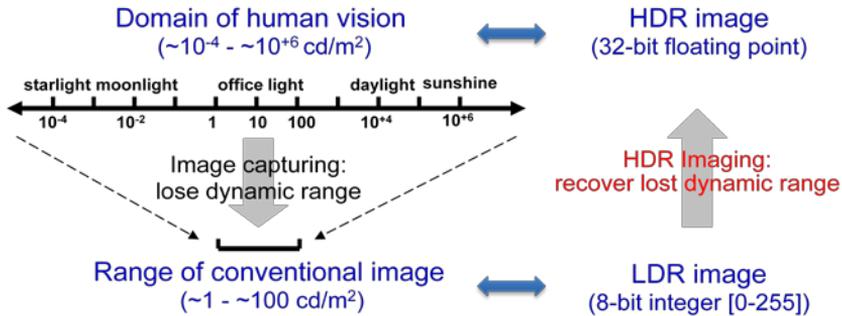


Fig. 1.2 The domain of human vision compare to the range of conventional LDR images. The image capturing process lose information due to sensor limitation, while HDR imaging recover the lost dynamic range through reverse engineering by merging multiple shots with different exposures.

1.2 HIGH DYNAMIC RANGE ACQUISITION

In order to recover the real world scene, various HDR synthesis methods have been proposed with different acquisition methods, as listed in Table 1.1.

It is possible to capture an HDR image using a single shot with the modification of camera hardware. Capturing image

gradients rather than actual pixel intensities was shown to increase the dynamic range at the cost of computationally expensive *Poisson solver* [8]. As far as I know, the gradient camera is still a theoretical solution. In *assorted pixels* method, multiple neighboring pixels with fixed pattern filter are used to capture at different exposures [9, 10]. Each pixel in the final HDR image was reconstructed by 2-4 neighboring pixels, which results a loss of camera resolution. To overcome this problem, an alternative design uses an aligned spatial light modulator with programming imaging [11, 12]. However, such a design is difficult to implement with extra hardware expenses.

Table 1.1 Typical acquisition approaches for capturing high dynamic range images.

Acquisition Methods	HDR Synthesis Methods
Single image capture	Gradient camera Assorted pixels Dynamic range adaptive imaging
Single device multiple image capture	Exposure bracketing (HDR video) Generalized mosaics
Multi-device capture	Split aperture imaging Optical splitting trees

The second acquisition approach uses single camera with multiple shots. Among all the methods, exposure bracketing [1, 2, 13, 14, 15, 16] is most widely used as no hardware modification is required. Within a short period of time, multiple shots with different exposures are captured to minimize the change of the lighting condition as well as object movement. Then, intensity mapping is estimated using image correlation and an HDR radiance map is recovered by using the best exposure information in different input images. Similar approach was adopted to generate HDR video too [17, 18, 19, 20]. In addition to exposure bracketing, a static filter with varying transitivity, named *generalized mosaicing* [21], was proposed to be mount in front of the camera sensor to mimic the exposure bracketing effect.

Third acquisition approach uses multiple devices to capture the same scene with the help of prisms [22] or beam-splitters [23,

24, 25]. While these approaches allow dynamic content to be recorded, the additional optical elements split the emitting light and make it difficult to capture the low light scenes in their best exposures.

1.3 HIGH DYNAMIC RANGE DISPLAY

Similar to image capturing system, the same hardware limitation exists in display system, where the HDR contents cannot be displayed on the conventional LDR display devices, such as monitors, printers and projectors.

Nowadays, most HDR images are converted back into LDR images using HDR tone-mapping before being presented on conventional display devices. There are two major categories of HDR tone-mapping algorithms, global operators [26, 27, 28] and local operators [29, 30, 31, 32]. Global operators are optimized on the whole image without local adjustment based on neighboring information. It is fast with no intensity inversing, meaning the brighter pixels in the HDR image remains brighter or equal in the tone-mapping result. A good application is HDR video [33]. The tradeoffs of the global operators are the degradation of visual quality, such as lack of local contrast and lose of small details. In order to improve the visual quality, local operators use the neighboring information for every pixel to preserve the fine details. These operators are usually computationally expensive, and come with halo artifacts [29, 34] in some scenarios.

Besides software solution, research based on hardware solutions has been carried out to display the original HDR content. A dual modulation display system [35] was invented by adding dynamic backlight using light-emitting diode (LED) to the conventional liquid crystal display (LCD), where the LCD provides color, resolution and normal contrast, while the programmable LED provides additional rear-illumination. Similar ideas have also been applied to increase the dynamic range of printer [36] and microscopy [37]. Another approach

focused on HDR projection by reallocating the light inside a conventional projector using an analog micro-mirror array in the optical path [38]. The light reallocation does not only increase the dynamic range of the projector display, but also reduces the device heat due to dumping excessive background light out instead of physically enclosed in the device.

1.4 WORKFLOW OF SOFTWARE SOLUTION OF HDR IMAGING

In this thesis, we focus on the software solution of HDR imaging. Therefore, exposure bracketing is used for HDR acquisition, and HDR tone-mapping is adopted for displaying. This solution is most widely used nowadays in consumer industry due to their capabilities of directly applying on existing hardware, which reduces the total cost.

As shown in Fig. 1.3, a typical software solution of HDR imaging is not only one or two, but series of technologies that transforms the captured LDR images (usually 8-bit per color channel) into HDR images (usually 32-bit per color channel), and then display them on existing conventional display devices. The modules in amber color in Fig. 1.3 are the research topics covered in this thesis. The modules in green color are technologies used in our software solution of HDR imaging system, which are not covered in this thesis.

Nevertheless, most consumer-level smart phones, such as iPhone 5, Samsung Galaxy S3, and HTC 1X¹, embedded HDR acquisition using exposure bracketing and display using *exposure fusion* [39, 40, 41, 42, 43, 44, 45], a combined fast process of HDR synthesis and HDR tone-mapping. We use exposure fusion for our mobile HDR solution, which is not covered in this thesis.

¹ Information and datasheet can be found at
iPhone 5: <http://www.apple.com/iphone/>
Galaxy S3: <http://www.samsung.com/global/galaxys3/>
1X: <http://www.htc.com/www/smartphones/htc-one-x/>

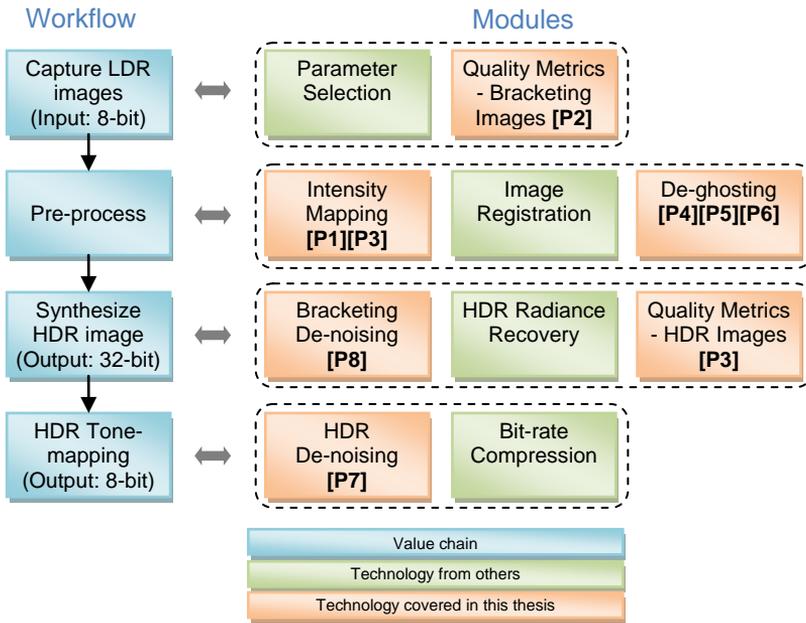


Fig. 1.3 A typical workflow of software solution for high dynamic range imaging.

1.5 STRUCTURE OF THE THESIS

The rest of the thesis is organised as follows. Intensity mapping is discussed in Chapter 2. HDR de-ghosting is studied in Chapter 3. Bracketing image de-noising and tone-mapping de-noising are summarized in Chapter 4. Image difference metrics for the exposure bracketing images and the HDR images are presented in Chapter 5. A summary of the contributions is given in Chapter 6 and conclusions drawn in Chapter 7.

2 Intensity Mapping

Exposure bracketing is a photography technique to capture the same scene with multiple shots using various exposure settings. It allows high dynamic range contents to be stored at different exposure levels [46, 47, 48, 49].



Fig. 2.1 A sequence of exposure bracketing images from the highest exposure (top left) to the lowest exposure (bottom right). The image of high exposure captured the detail of dark area with the saturation in the bright area, while the image of low exposure captured the bright area with the saturation in the dark area.

HDR images are usually synthesized by recovering the radiance map using the correlations among these bracketing images. Among all the correlations, intensity mapping is the most important feature defined as

$$\psi_{i,j}(Z_i(p)) = Z_j(p) \quad (2.1)$$

where Z_i and Z_j are the intensity value of the same co-located pixel p at different exposures of i and j , and $\psi_{i,j}$ is the intensity mapping from image i to image j . It is noted that the intensity mapping is a directional function ($\psi_{i,j} \neq \psi_{j,i}$). The *intensity mapping function* (IMF) is not only hardware dependant, but also

will change according to different capturing settings. And in most cases, it is not a linear function.

The intensity mapping is the key to many HDR applications, such as image in-painting with exposure bracketing images, HDR image synthesis, and HDR de-ghosting.

2.1 BACKGROUND

When taking a photo, each pixel in the digitized output image will contain a pixel value known as *intensity*. However, this value is rarely true measurement of the real radiance of that pixel [14]. Instead, it is the output of a mapping function, known as *camera response function* (CRF) [1, 2, 50], which maps the real world radiance to a pixel value within the display range.

Intuitively, CRF is a good candidate for intensity mapping. It is a characteristic of the camera and the exposures and, thus, is not scene dependant and is commonly estimated from the exposure bracketing images. A popular CRF estimation is proposed in [2] by minimizing the quadratic objective function

$$\mathcal{O} = \sum_{p=1}^P \sum_{i=1}^N [g(Z_{i,p}) - \ln E_p - \ln \Delta t_i]^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} g''(z)^2 \quad (2.2)$$

where g is the log reverse CRF function, Z_{ij} is the known pixel value at pixel location p (total P pixels) in the input image i (total N images), E denotes the radiance value, Δt denotes the exposure time, and λ represents a weight for the smoothness term. The CRF estimation is computationally expensive, and not very robust to noise, moving object or camera movement. Using CRF, the intensity mapping can be represented as

$$\psi_{i,j}(Z_i(p)) = g^{-1}(g(Z_i(p)) - \ln \Delta t_i + \ln \Delta t_j) \quad (2.3)$$

where the exposure time must be known in advance.

Another important approach uses a histogram based IMF[51], which determines the intensity mapping by using cumulative histogram. An IMF (ψ) is defined as

$$\psi_{i,j}(Z(p)) = H_j^{-1}(H_i(Z(p))) \quad (2.4)$$

where $H(z) = \int_0^z h(x)dx$ is the cumulative histogram of the original histogram h . Compared to the CRF-based approach, the histogram based intensity mapping does not require exposure time information, and does not require hard image alignment. A drawback is that it only makes use of the pixel statistic information, but not any spatial correlation.

2.2 INTER-PIXEL RELATIONSHIP FUNCTION

The *inter-pixel relationship function* (IRF) is proposed in [P1], and the idea is illustrated in Fig. 2.2. If the intensity at A' and B' are the same at one exposure time in the reference, they shall be the same at the other exposure times (not for saturated pixels). This is according to photography reciprocity [52]: when exposure time changes, the pixel values change correspondingly. However, during the image capturing process, sensor noise, sampling noise and compression noise are commonly generated. Thus, it is more accurate to find all the pixels with the same intensity in the reference image and calculate their co-location values in the test image using mean average. The IRF is defined as

$$\psi_{i,j}(Z_i(p)) = \frac{\sum_{x \in \Omega(Z_i(p))} Z_j(x)}{|\Omega(Z_i(p))|} \quad (2.5)$$

where $|\Omega(Z_i(p))|$ is the cardinality of the spatially co-located pixels set $\Omega(Z_i(p)) : \{x | Z_i(x) = Z_i(p)\}$.

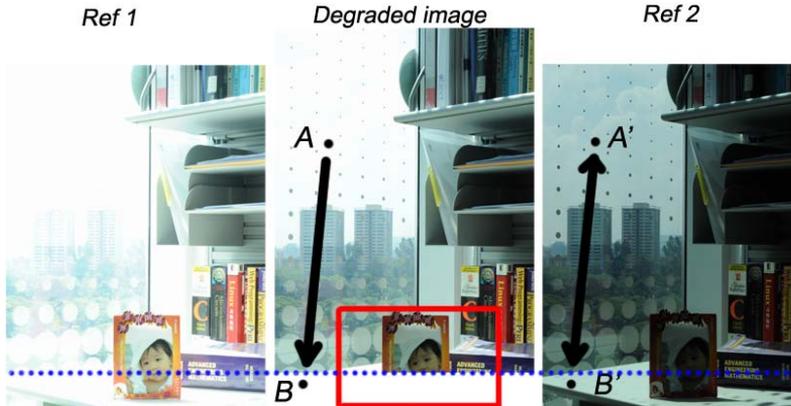


Fig. 2.2 Three differently exposed images with degraded image due to camera shake. The pixel value of B can be copied from A as their co-location A' and B' have the same intensity in the reference image.

The IRF has three useful characteristics inherited from the physical camera response. (1) The IRF is a monotonically increasing function. (2) The pixels located at left end (dark pixels), as shown in Fig. 2.3, and right end (bright pixels) are highly compressed due to dynamic range limit. (3) When choosing different reference images, shorter exposure time leads to smaller slope at the left end and bigger slope at the right end.

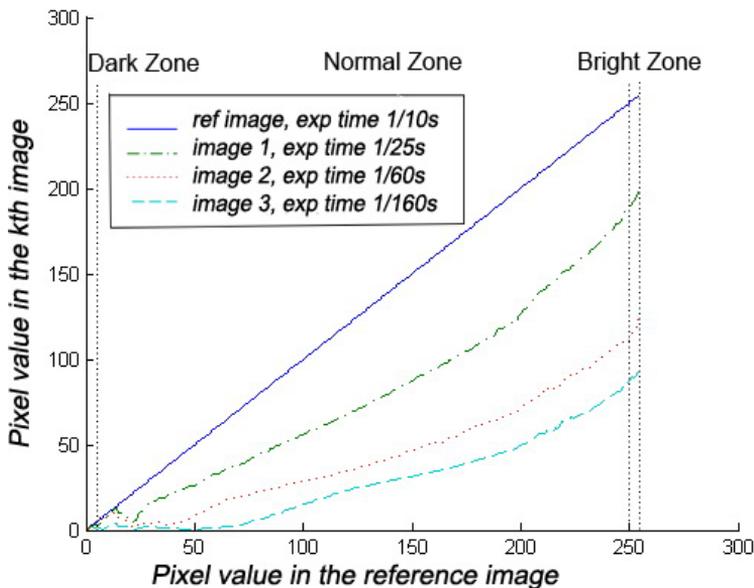


Fig. 2.3 An example of IRF function. Multiple IRF curves are presented with different images (captured at different exposure time) to the same reference image.

2.3 INPAINTING WITH EXPOSURE BRACKETING IMAGES

Images of same scene can be captured with different exposures and combined with computing power to synthesize image that overcomes limitation of conventional cameras. However, useful data can be lost due to camera shake, especially when capturing by a hand held device, which generates noticeable artifact in the synthesized image. In other words, different from the traditional image in-painting [53, 54, 55, 56] and scene completion [57], which generate only photorealistic patches, the degraded image in-painting in digital photography requires true luminance value of real world scene. Therefore, the challenge of patching is to find useful relations between missing pixels and the remaining pixels.

In [P1], an IRF is proposed to in-paint the missing area with a dual patching, where dual reference images are used. In order to increase the accuracy, the reference image is selected to have the smallest exposure difference with the degraded image. However, the dynamic range lost caused by the characteristic 2, described in chapter 2.2, is still inevitable.

If the reference image has shorter exposure time than the degraded image, then as can be seen from the characteristic 3, the dark pixels in the reference image are mapped from a large contrast to a small one. In other words, it is a compressing process mapping multiple values to one, which in turn makes the IRF in this area more reliable. On the contrary, a highly compressed bright pixel in the reference image is mapped into multiple values in the degraded image, which causes inaccuracy due to the dynamic range lost. Thus, multiple reference images, with longer and shorter exposure time respectively, can be adopted to recover the lost dynamic range and enhance the patching accuracy. The missing pixel intensity is calculated by

$$Z_t(p) = \frac{\psi_{i,t}(Z_i(p)) \cdot \omega_{i,t}(Z_i(p)) + \psi_{j,t}(Z_j(p)) \cdot \omega_{j,t}(Z_j(p))}{\omega_{i,t}(Z_i(p)) + \omega_{j,t}(Z_j(p))} \quad (2.6)$$

where image i and j are the two reference images, image t is the degraded image to be patched, and ω is the weighting function defined as

$$\omega_{i,t}(z) = \begin{cases} \log(z + 1), & \Delta t_i \geq \Delta t_t \\ \log(256 - z), & \Delta t_i < \Delta t_t \end{cases}, z \in [0,255] \quad (2.7)$$

which is plotted in Fig. 2.4.

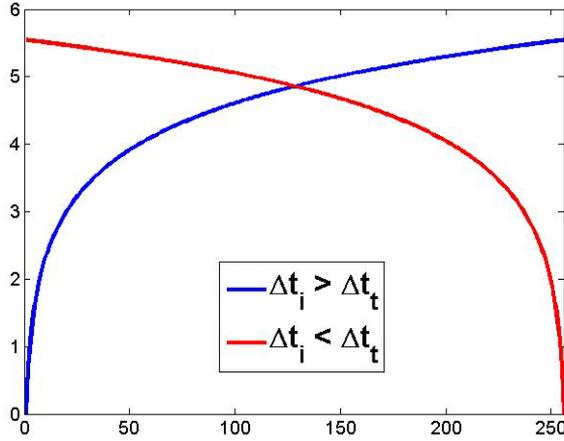


Fig. 2.4 Plot of the weighing function ω .

The missing area in an exposure bracketing image set can be recovered using the proposed intensity mapping. Compare with other methods, the exemplar-based in-painting [54] works well on simple texture, such as the table top. However, obvious errors can be seen at complex contents, like the baby's face and the title of the book as shown in Fig. 2.5 (b), due to short of reference. The CRF-based method [58] recovers the contents by the luminance shift from the reference pixels, where obvious artifact can be seen at the border. The method proposed in [P1] recovers the original scene effectively in terms of speed and quality. We tabulate the *peak signal-to-noise ratio* (PSNR) of each method in Table 2.1. The test was set up similar to Fig. 2.5 (a), where part of the image is missing due to camera movement, but can be patched using two references (a brighter reference

and a darker reference). This scenario is common in HDR imaging. The result was averaged from ten test sets.

Table 2.1 Typical acquisition approaches for capturing high dynamic range images.

Method	Inpainting [54]	CRF based [58]	IRF dual patching [P1]
PSNR (dB)	11.75	20.54	33.66

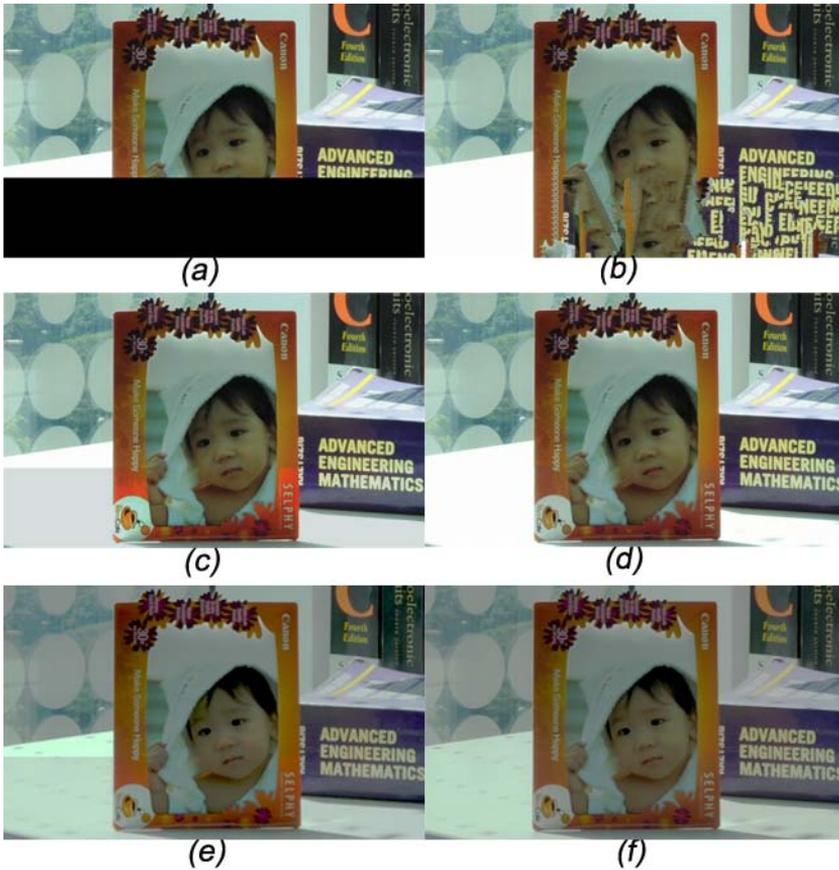


Fig. 2.5 Patching results of a degraded image from an exposure bracketing set shown in Fig. 2.2. (a) Original degraded image, (b) fixed image using exemplar-based inpainting [54], (c) fixed image using CRF [58], (d) fixed image using the proposed method, (e) HDR image synthesized with degraded image, and (f) HDR image synthesized with patched image [P1].

The structure of the proposed method, intensity mapping plus dual patching, is extended to patch-based in-painting in

our later research works [59, 60]. It is proven to be an important tool for HDR de-ghosting [61, 62, 63].

2.4 BLIND EXPOSURE VALUE ESTIMATION

The exposure time is an important feature in synthesizing HDR images [1, 2, 50]. However, this information can be easily lost during copy or editing. Some methods have been introduced to recover the missing exposures. A radiometric model [14] was introduced to estimate the exposure ratio with an initial value set by user and an iterative process to fine-tune it. The results vary due to different initial settings and the order of the polynomial model used in the estimation. Another rough approximation was obtained by using IMF [51] between two images. However, as realized by the authors themselves, there are a lot of restrictions and assumptions involved. Thus, they used the approximated exposure ratio as the initial value to the radiometric model [14] in their experiments.

In [P3], we present an accurate algorithm to recover the rate in between the different exposure ratios by using IRF [P1] and CRF [2]. For any pair of differently exposed images, each co-located pixels represent the same intensity. Thus, if they are not saturated, their reverse CRF mapping difference shall be a constant that represents the radiance difference of their exposure ratio. The exposure ratio is defined using *exposure value* (EV) interval as $EV_i = \log_2(\Delta t_i / \Delta t_{i+1})$. For any three images in an exposure set $(i, i + 1, i + 2)$, we select one image (i) as reference. The function CRF (g) is calculated using the reference image and the first test image, while the pixel correlations are calculated using IRF between the reference image and the second image. As such, the EV unit in between the second test image and the reference image calculated by

$$\delta_{i+2,i+1,i} = \frac{\sum_{z=\hat{Z}_{i,i+2}}^{\hat{Z}_{i,i+2}} [g_{i,i+1}(z) - g_{i,i+1}(\varphi_{i,i+2}(z))]}{\hat{Z}_{i,i+2} - \hat{Z}_{i,i+2} + 1} \quad (2.8)$$

where \hat{Z} and \check{Z} are two constants that defines the valid region boundaries as

$$\begin{cases} \hat{Z}_{i,j} = \max(\alpha, \varphi_{i,j}^{-1}(\alpha)) \\ \check{Z}_{i,j} = \min(\beta, \varphi_{i,j}^{-1}(\beta)) \end{cases} \quad (2.9)$$

where α and β are two constants defines the saturation value. When the first and second test images are the same, a base EV unit is estimated. The unknown exposure ratio is then calculated by

$$k_{i,i+2} = \left(\frac{\delta_{i+2,i+1,i}}{\delta_{i+1,i+1,i}} \right) \cdot k_{i,i+1} \quad (2.10)$$

If there is no prior exposure knowledge, assumptions of the exposure ratio between the first two LDR images are given, as shown in the experiments in Table 2.2. The estimated EV internals have small errors of 5-10%, which will not generate obvious distortion in the final HDR image. If the initial EV was not given correctly, as shown in EXP3, all the estimated EV will have the same proportion to the given initial EV, which results in synthesizing a near-identical HDR image (Chapter 5.2).

Table 2.2 EV correction with given initial exposure ratio

		EV1	EV2	EV3	EV4
EXP1	Original	1	1	1	1
	[P3]	1 (given)	0.92	0.94	0.88
EXP2	Original	1.66	1.66	1.66	1.66
	[P3]	1.66 (given)	1.63	1.62	1.58
EXP3	Original	1	1	1	1
	[P3]	1.58 (given)	1.53	1.54	1.51

3 *De-ghosting*

When synthesizing HDR image using exposure bracketing images, ghosting artifacts usually appear due to moving object [64, 65, 66, 67], such as moving people and trees waving in the wind. To solve this, de-ghosting is applied before the HDR synthesis.

There are three major challenges in the de-ghosting process. Firstly, LDR images are taken with different exposures and cannot be compared directly. It is difficult to find a good criterion to bring different exposed images to the same comparable scale. Secondly, moving object pixels are classified by a threshold, but other factors, such as capturing parameters (ISO, shutter speed), capturing hardware and light conditions, affect the pixel co-relationship and militate against the selection of a static threshold. The third challenge is how to recover the missing dynamic range information.

3.1 IMAGE REGISTRATION

Before jump into de-ghosting, there is another important pre-processing step: *image registration* [68, 69, 70]. It is required if the images are captured using hand held devices. Otherwise, blur will appear in the synthesized HDR image. Luckily, when capturing multiple images in burst mode, misalignment are commonly seen as rotation and translation only. Thus, in most cases, only image alignment is required in HDR imaging.

Fig. 3.1 gives a simple comparison between image alignment and de-ghosting, which both cause blur in the synthesized HDR image. Blur due to mis-alignment is a global effect, while ghosting artifact due to moving object is content based. Thus, it is more difficult to do de-ghosting than alignment.

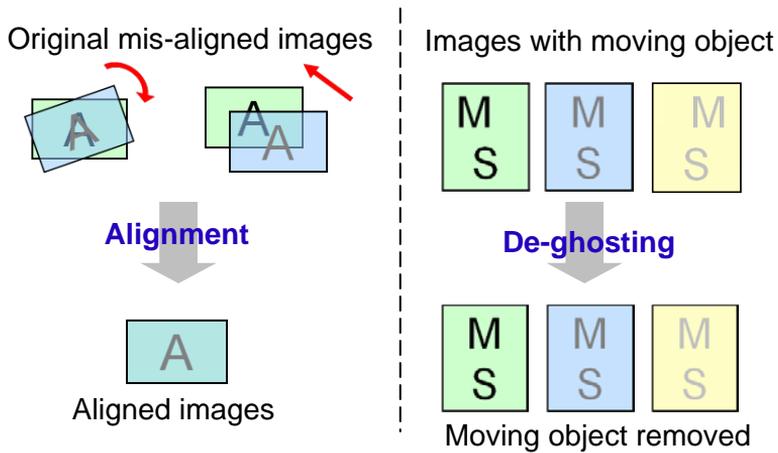


Fig. 3.1 Comparison with alignment (left) and de-ghosting (right).

We have made some study on image alignment methods for HDR imaging, as shown in Table 3.1. Some of the ideas are used in the final HDR imaging application.

Table 3.1 Comparison of image alignment methods

Method	[71]	[17]	[72]	[58]	[73]	[74]	[75]	[76]
Using feature	Median value	Pixel	Pixel	Median value	Pixel	Key points	Corner	Histogram
Support extreme DR		✓	✓					
No CRF			✓	✓		✓	✓	✓
Fast feature matching	✓					✓	✓	✓

3.2 REVIEW OF DE-GHOSTING METHODS

Different from traditional motion detection[77, 78, 79, 80] in video processing, the input images (frames) are not at the same exposure level. Thus, special processes are needed.

An iterative approach was proposed in [81] by detecting the probability of a pixel belonging to a moving object or background, which minimized the assumption on the input image sequence. However, it is difficult to remove the moving object completely, and the algorithm is computationally expensive. Local entropy was proposed in [82] by matching the structure of the distribution of the local pixel value. It is a relative fast process, as there is no need for response curve estimation. However, as realized by the authors themselves, it cannot detect even large intensity differences when two co-located areas have similar local structure. Grosch [58] proposed a method using *camera response function* (CRF). It is suitable for the scenario when CRF is known, and can therefore perform a very fast detection process. Otherwise, with moving object, it is very difficult to estimate CRF accurately. A pre-classification process is adopted to solve this problem in [83]. Gallo et al. [84] used a similar idea with an assumption that log exposure can be mapped to a straight line, which in practice may not be the case due to different capturing hardware.

3.3 REALTIME DE-GHOSTING FRAMEWORK

A real-time de-ghosting method is proposed in [P4], which includes detecting of moving object pixels and patching them using only the background pixels. Its major benefit is that it is fast and requires little computational resources, including processing power and physical memory. In all cases, the de-ghosting process involves only two images: a test image and the reference image. Because of this, the de-ghosting process is capable of working concurrently with the image capturing process, as shown in Fig. 3.2. The main function modules are the IRF estimation [P1], bi-directional comparison, moving object classification and moving object correction.

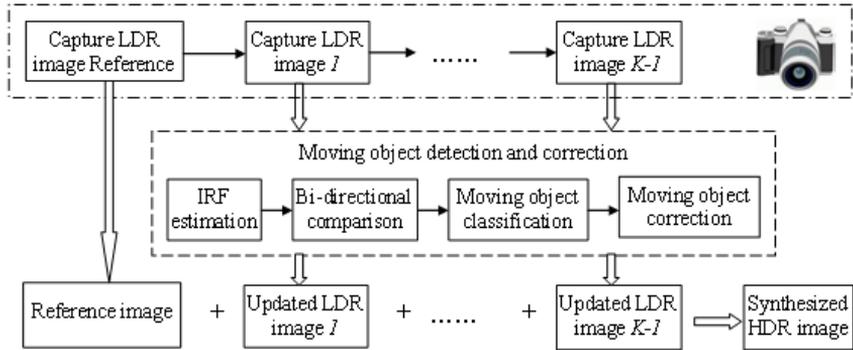


Fig. 3.2 Framework of the proposed real-time de-ghosting method.

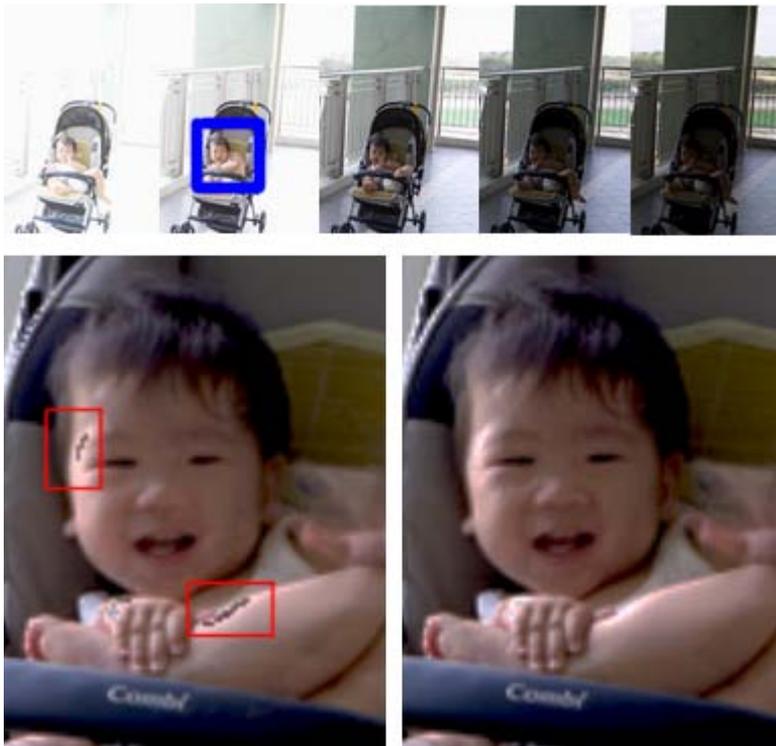


Fig. 3.3 Original input LDR images (top), and the zoom-in HDR images composed using uni-direction method (bottom left) and bi-directional method [P4] (bottom right).

The reciprocity law fails at underexposed and saturated pixels [52], as they are cut-off values due to the limitation of the dynamic range of the camera. Thus, the IRF values of those regions are unreliable. To solve this, bi-directional comparison is

proposed to mapping only from the more reliable pixel to the less reliable pixel as

$$\Gamma_{i,j}(p) = \begin{cases} |Z_j(p) - \psi_{i,j}(Z_i(p))|, & \omega(Z_i(p)) \geq \omega(Z_j(p)) \\ |\psi_{j,i}(Z_j(p)) - Z_i(p)|, & \omega(Z_i(p)) < \omega(Z_j(p)) \end{cases} \quad (3.1)$$

where ω is a triangle weighting function [2], and $\Gamma(p)$ is the error estimator at pixel p . The error estimator simulates the distance between the two co-located pixels despite the saturation area in the reference image. An example is given in Fig. 3.3.



Fig. 3.4 Original LDR images (top), and the zoom-in HDR images synthesized without outline re-evaluation (bottom left) and with outline re-evaluation [P4] (bottom right).

It is worth noticing that a simple but powerful outline re-evaluation filter can be adapted to the system before patching using IRF. The outline artifacts usually appear at the boundaries due to a similar value between moving object and the background, which makes it difficult to detect. The idea of the re-evaluation filter is to use a tight threshold and re-detect the

boundary pixels. An example of the outline re-evaluation filter is shown in Fig. 3.4.

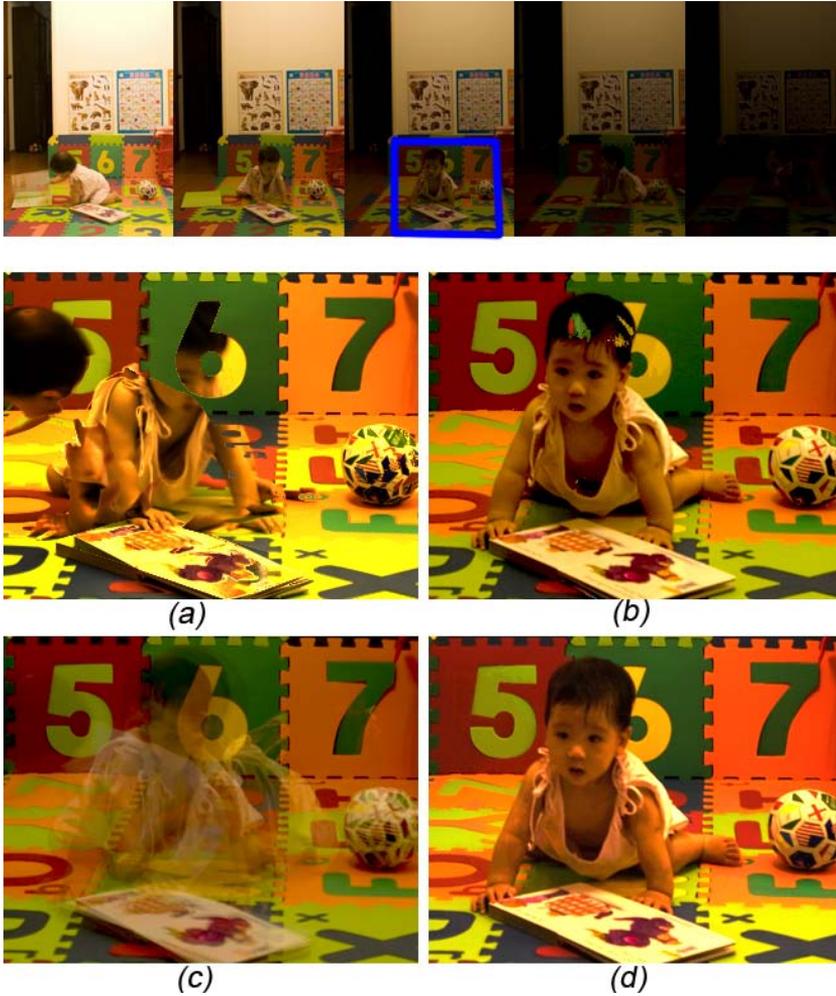


Fig. 3.5 HDR image synthesized from (top) five exposure bracketing images, using (a) *FDRTools*¹, (b) *Photomatix*², (c) *Qtpfsgui*³, and (d) **[P4]**.

Similar to [58], once the error estimator is calculated, the moving object pixels are classified using empirical fixed threshold and patched using IRF. Experiments show a

¹ *FDRTools* is available at <http://www.fdrtools.com/>

² *Photomatix* is available at <http://www.hdrsoft.com/>

³ *Qtpfsgui* is available at <http://qtpfsgui.sourceforge.net/>

significant improvement in the visual quality comparing to the off-the-shelf applications, such as Photoshop.

Our real-time de-ghosting framework has been tested on smart phones, where computation resources are limited. Combine with sub-sampling in Chapter 3.5, the processing time of the whole de-ghosting process for three 8MB (resolution 3264x2488) input images is only 700ms on Samsung Galaxy S3.

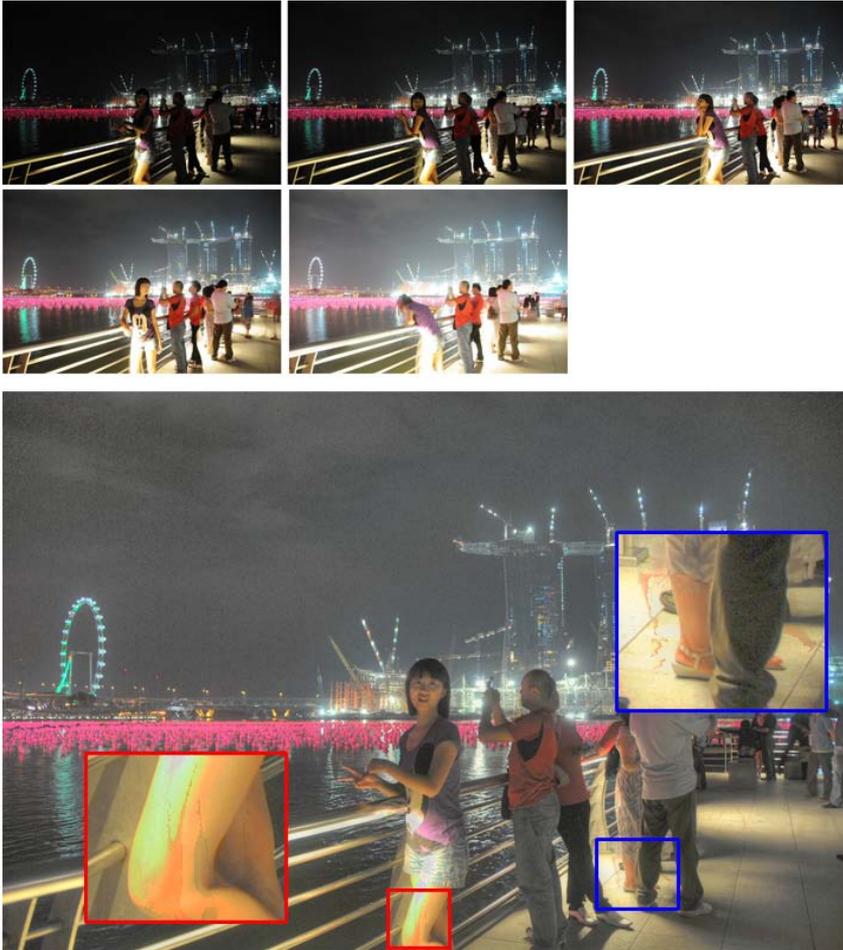


Fig. 3.6 An example of the failed case due to extreme large intensity changes, where (top) are the input image sequence, and (bottom) is the synthesized HDR image.

The tradeoff of this fast processing is the fixed threshold, which fails in some extreme scenarios, such as an example of a night scene shown in Fig. 3.6.

3.4 ROBUST DE-GHOSTING

The real-time de-ghosting method introduced in Chapter 3.3 is a useful solution for mobile applications, where processing power and memory are the most important concerns. On the other hand, when quality is the key requirement, such as an offline HDR synthesis using PC, a more robust de-ghosting method should be used. In [P5], an improved de-ghosting method is proposed using double-credit intensity mapping which suits for more scenarios and gives better visual quality.

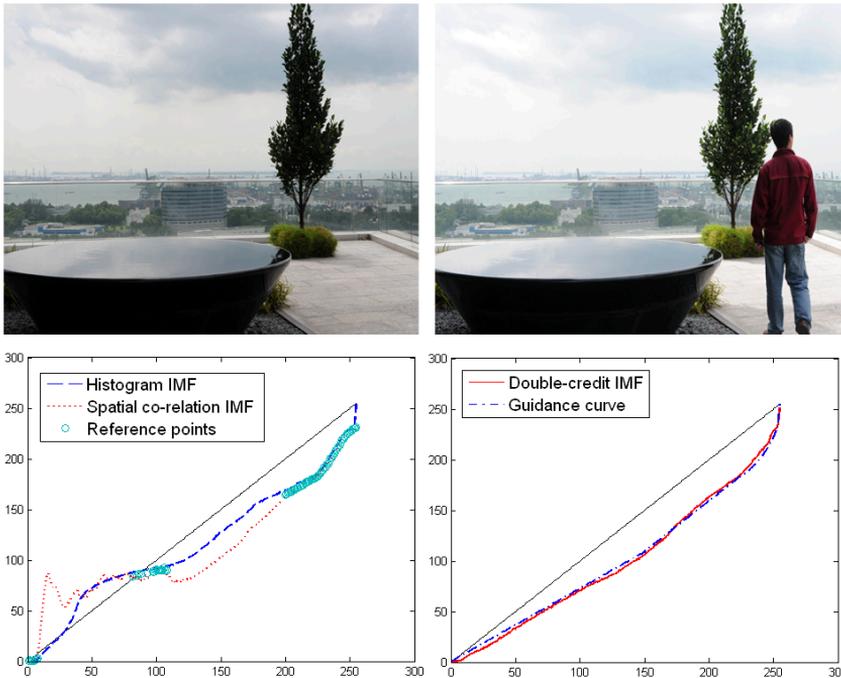


Fig. 3.7 An example of double-credit IMF estimation. Two input images with different exposures (Top); extraction of reference points (Bottom-left); guidance curve and the fine-tuned double-credit IMF (Bottom-right).

Histogram-based intensity mapping [51] uses the statistic information of the intensity distribution. However, such a process loses position information. On the other hand, the proposed IRF [P1] uses spatial correlation which preserves the position information, but can be distorted by large moving objects. Thus, a double-credit de-ghosting method is used,

which combines the advantages of both statistic information and spatial correlation, and generates HDR image with high visual quality.

The final double-credit IMF is estimated using the following steps. First, reference points are detected as the points with the same value in the histogram-based IMF as well as the spatial correlated IRF. A guidance curve is an approximation by extending the reference points with Bezier curve approximation [84, 85]. The guidance curve is a monotonic increasing function that cut across the identified reference points. It provides the initial value for the final IMF, and is useful to filter out large intensity changes due to moving object. The final double-credit IMF is then fine-tuned by calculating the spatial correlation using only the reliable pixels as

$$\psi(z) = \frac{\sum_{p \in \Omega(z)} Z(p) + \psi_G(z) \cdot \beta}{|\Omega(z)| + \beta}, z = 0, \dots, 255 \quad (3.2)$$

where $\Omega(z)$ is the set of pixels with the same intensity of z in the reference image, $|\Omega(z)|$ represents the cardinality of this set, Z denotes the pixel of the test image, $\psi_G(z)$ is the guidance curve and β is a predefined weight to the guidance curve.

In order to robustly detect moving objects in different lighting conditions, an empirical fixed threshold is no longer feasible. In [P5], a data driven training is conducted based on images captured from tripod cameras with no moving object in the scene. The training is aimed to find a threshold model suitable for different scenarios. A threshold vector is calculated based on pixel intensities (from 0 to 255) to classify 95% of the pixels as background. An example is shown at the bottom graph of Fig. 3.8. The image database includes more than 300 images from different cameras including Nikon D3, D300, Canon EOS-1, IXUS850, IXUS900, and images downloaded from Internet. The images have been captured using different capturing parameters at different lighting conditions to ensure extensive coverage. The adaptive threshold model is concluded as

$$T(z, M) = \begin{cases} a, & 0 \leq z \leq d \\ a + \frac{b-a}{M-d} \cdot (z-d), & d < z \leq M \\ b + \frac{c-b}{255-M} \cdot (z-M), & M < z \leq 255 \end{cases} \quad (3.3)$$

where z represents the original RGB value, M represents the optimum bi-directional comparison central point, and a, b, c, d are the key parameters derived from the variance vector (V) generated during the fine-tuning of the double-credit IMF estimations. We use the following:

$$\begin{cases} a = 0.8 \cdot \sqrt{\frac{1}{d} \sum_{z=1}^d V(z)} \\ b = 0.8 \cdot \sqrt{\frac{1}{d} \sum_{z=M-d/2}^{M+d/2} V(z)} \\ c = 5 \\ d = 30 \end{cases} \quad (3.4)$$

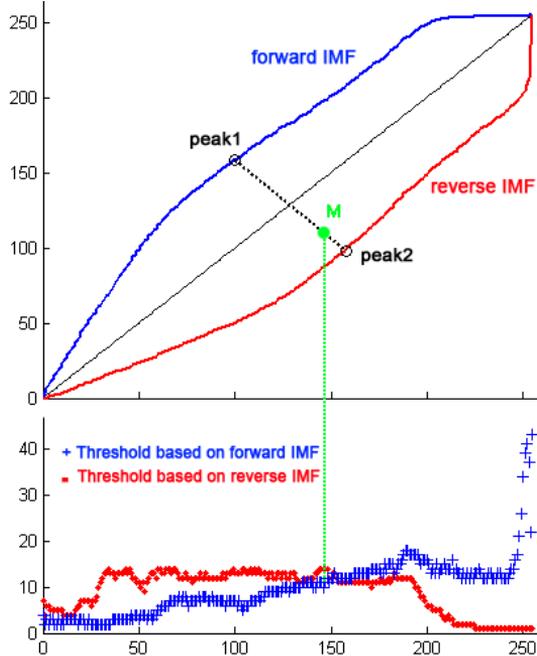


Fig. 3.8 The top graph plots the forward IMF and the reverse IMF of two static images. The bottom graph plots the thresholds, with which 95% of the pixels are classified as background. The forward IMF gives better performance on the left side of the central point M , and the reverse IMF minimizes the threshold value on the right.

Table 3.2 SNR comparison of different intensity mapping functions

SNR	Histogram	Spatial	Double-credit [P5]
SET1	20.32 dB	21.79 dB	27.50 dB
SET2	28.49 dB	18.39 dB	33.05 dB

The proposed double-credit IMF is compared with the histogram based mapping [51] and spatial correlation based mapping [P1] using two sets of images, of which each contains 30 bracketing image pairs. In SET1, the moving object (small size) appears only in one image. In SET2, the moving object appears in both images at different positions. An average improvement of 7 dB is achieved, see Table 3.2, and no ghosting artifacts appears in the visual quality test when comparing with the leading commercial software in Fig. 3.9 and Fig. 3.10.

There is an extreme case where the proposed method will fail. When the moving object is in a saturated area in the reference image, there is no information to patch during the synthesis. In this case, the result of the proposed de-ghosting algorithm may contain grey color patches. A hybrid patching algorithm was proposed in [59, 60] to handle these saturated pixels using block patching.



Fig. 3.9 De-ghosting visual comparison in a day scene with (top) input images, using (bottom-left) Photomatix, (bottom-middle) Photoshop¹, and (bottom-right) [P5].

¹ Photoshop is one of the leading image processing software, and can be found at <http://www.photoshop.com/>.



Fig. 3.10 De-ghosting visual comparison in a night scene with (top) input images, using (bottom-left) Photomatix, (bottom-middle) Photoshop and (bottom-right) [P5].

3.5 MOTION DETECTION FOR SUB-SAMPLED IMAGES

Modern smart phones, such as iPhone 5 and Galaxy S3, are capable of capturing images with resolution up to 8MB. Thus, sub-sampling based fast movement detection is proposed in [P6] to save computational resources for the de-ghosting algorithm to be running on the mobile devices.

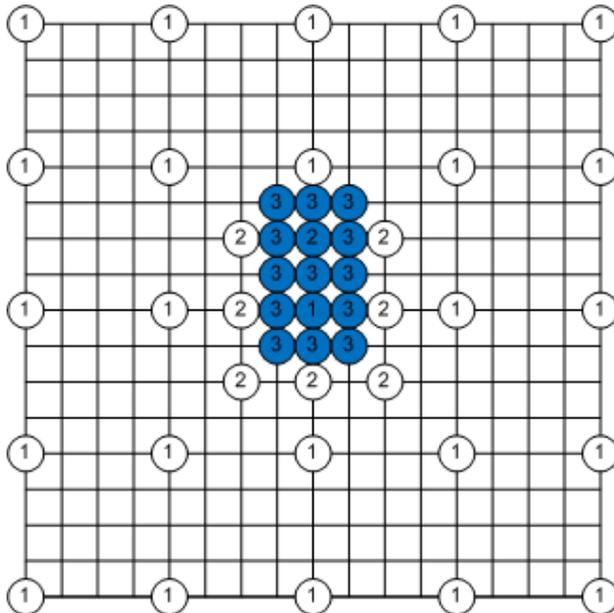


Fig. 3.11 A sub-sampling based movement detection scheme

An example with a sub-sampling factor of 4 is selected to illustrate the proposed fast movement detection scheme in Fig.

3.11. The pixels that are labeled with 1 are checked at the first round, and they are marked by the blue color if they are detected as moving object pixels. The neighboring pixels of a blue pixel, labeled with 1, are then checked by using a sub-sampling method. In Fig. 3.11, all neighboring pixels that are labeled by 2 are checked in the second round. They are also marked by the blur color if they are moving object pixels. Finally, the eight neighboring pixels of a blue pixel labeled by 2, i.e., those pixels are labeled by 3, are checked in the final round.

Only 1/16 of all pixels are detected in the first round. Since only a small portion of pixels in an LDR image belongs to moving objects, the second and third rounds of detections are only conducted for a small amount of pixels in the LDR image. As such, the complexity of the improved movement detection scheme is significantly reduced. Meanwhile, since many pixels are in the neighborhoods of two pixels, a flag is attached to each pixel to indicate whether it has been detected. With the flag, each pixel will only be processed once.

Experiments show that the overall number of processed pixels is reduced by up to 58.61%.

4 HDR De-noising

4.1 DE-NOISING OF LDR IMAGES

Image noise is usually caused by the capturing device. Common sources for noise are photon shot noise, dark current noise, and readout noise[87]. Noise may also come from special hardware, such as *fixed pattern noise* in *charge-coupled device* (CCD) camera [88, 89] and *thermal noise* and *flicker noise* in *complementary metal-oxide-semiconductor* (CMOS) device [90].

Lots of research studies have been conducted to tackle the single image de-noising problem. Simple local filters, such as *Gaussian mean filter* and *median filter*[91] bring in blur in the de-noised image. Edge preserved filters, such as bilateral filter[92], are developed to keep the sharp edges. Studies have been carried out on the *anisotropic diffusion* based techniques[93] and wavelet-based techniques[94]. Patch-based image de-noising find similar patch information inside the original image and stack them together to remove the noise while keep the edges, such as non-local mean [95, 96], BM3D[97], PLOW [98, 99].

When multiple images of the same scene are available, it is possible to merge the information from the multiple inputs. The simplest approach is the *frame average*[100], which theoretically reduces the noise variance by N (number of input images). However, when the multiple images are not captured using the same exposure, frame average fails due to un-aligned exposures. To solve this, noise-reduced HDR synthesis is studied.

4.2 NOISE-REDUCED HDR SYNTHESIS USING 2D-DENOISING FACTOR

The presence of noise in an HDR synthesis poses a serious degradation to the HDR image especially when the input

images are captured at low light condition or with high sensitivity settings. Various methods have been proposed to tackle this problem. We further classify them into three categories: 1) weighted frame average based preprocessing method; 2) intensity weighting based direct HDR synthesis method; 3) exposure time based HDR composition method.

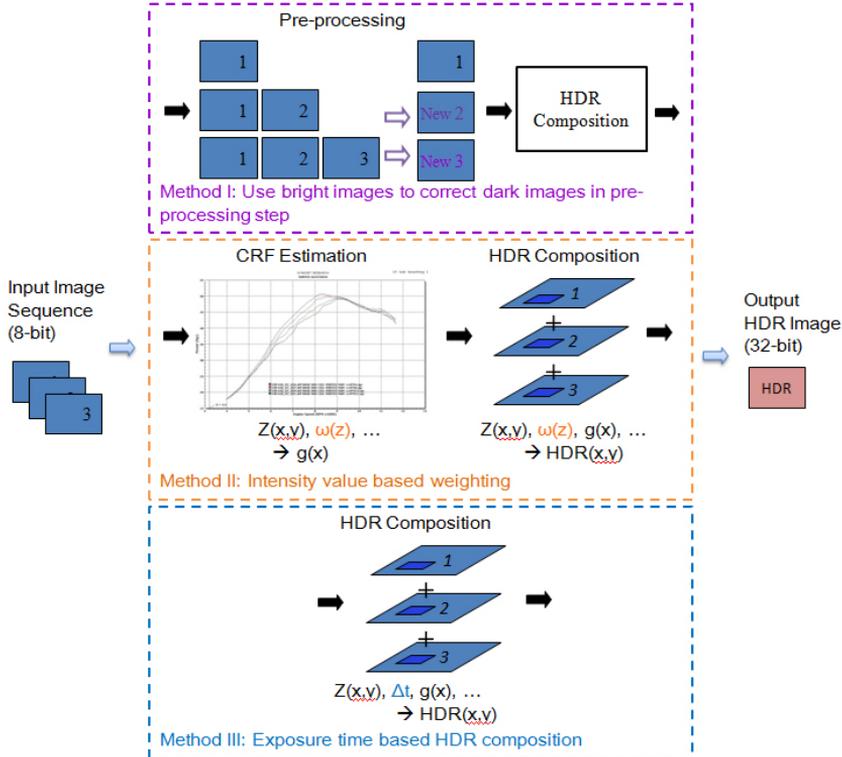


Fig. 4.1 Different approaches in noise-reduced HDR synthesis: (I) spatial-based edge preservation, (II) temporal based weighted frame average, (III) temporal-based intensity weighting, (IV) temporal-based exposure time weighting, and (proposed) temporal-based adaptive exposedness de-noising factor [P8]. The input image sequence can have more than 3 images.

The weighted frame average based methods, as shown in Fig. 4.1 (II), are based on an observation that brighter images contain less noise than the darker ones, as the brighter images have stronger signal and result in higher signal-to-noise ratios (SNRs). In order to average different exposed images, a calibration of the co-located pixels is required. CRF is used in [87] to recover the full radiance map for weighted averaging. IMF is used in [101]

to correct the dark image on LDR domain. These methods are capable of reducing noise in dark images at the cost of extra memory and computation time.

To reduce those costs, the intensity weighting based methods are designed to work directly in the HDR synthesis, as shown in Fig. 4.1 (III). A well-exposed pixel is assigned higher weight than an over/under-exposed pixel in both CRF estimation and HDR composition [87, 102, 103]. These methods are computational efficient and work well at extreme bright and dark area. But they are not as effective in the areas in between the result of weighted frame average methods.

Another approach attempts to use exposure times as additional weight [104] during the HDR composition, as shown in Fig. 4.1 (IV). However, the link between the estimation of the CRF and the proposed additional weight is missing. Color shift is another problem due to high weight at saturated pixels.

In [P8], we propose a two-dimensional (2D) denoising factor to assign higher weight to pixel with less noise based on both pixel luminance and image exposure. It is controlled by two coefficients. It preserves edges and fine details without blurring artifact. In addition, both memory and computation time are significantly reduced compare to other denoising methods.

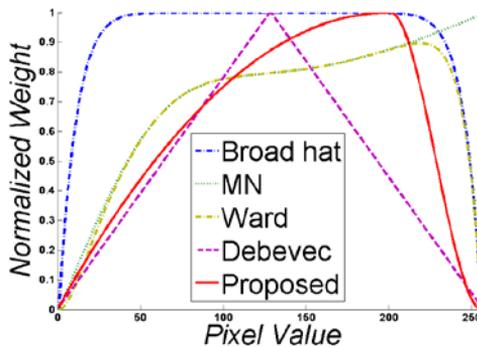


Fig. 4.2 The proposed luminance based weighting compares to other weighting functions: broad hat [1], MN [14], Ward [1] and Debevec [2].

The first dimension of the proposed denoising factor assigns a high weight to a pixel with a large luminance. Several methods have been proposed using different normalized

weighting functions, as shown in Fig. 4.2. Given an assumption that the noise distribution is independent of the measurement pixel value z , it is argued in [14] that a luminance based weighting function (*MN weight*) $\omega = f(z)/f'(z)$ achieves the best signal to noise ratio, where $f(z)$ is the radiometric response function. In [87], the MN weight is modified by replacing pixel value with luminance value. A broad hat function $h(z) = 1 - (\frac{2z}{255} - 1)^{12}$ was used to restrict the saturated pixels that may cause color cast.

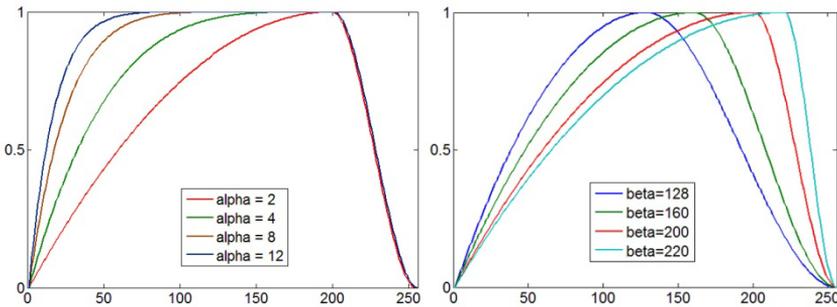


Fig. 4.3 A family of weightings in the first dimension calculated by using (left) different α with $\beta=200$, and (right) different β with $\alpha=2$.

Since the radiometric response function is usually monotonic increasing, we approximate the luminance based weight by a controllable hat function and a Hermite interpolation. Thus, we can significantly reduce processing time on response function recovery by defining a new weighting factor as

$$\omega(z) = \begin{cases} 1 - \left| \frac{z}{\beta} - 1 \right|^\alpha, & 0 \leq z < \beta \\ 1 - 3 \left(1 - \frac{255 - z}{255 - \beta} \right)^2 + 2 \left(1 - \frac{255 - z}{255 - \beta} \right)^3, & \beta \leq z < 255 \end{cases} \quad (4.1)$$

where the two key coefficient are the denoising strength coefficient α and the saturation control coefficient β . The smaller the value of α , the hat function will become steeper and this will result in better denoising effect. A large α gives high weights to pixels with small value (luminance), which remains noise in the

synthesized HDR image. A family of the proposed weighting function is plotted in Fig. 4.3. In our experiments, $\alpha = 2$ is chosen. Our tests indicate that $\alpha = 12$ generates similar results as [87]. The saturation control coefficient limits the near saturated pixels to avoid color cast due to gamut limitations (an empirical value $\beta = 200$ is used here).

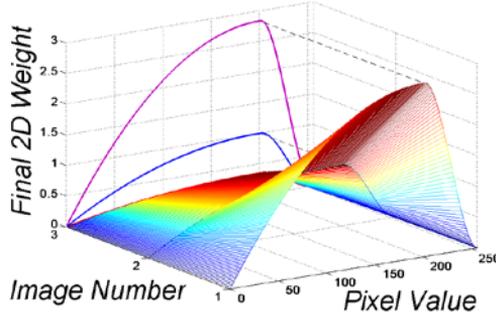


Fig. 4.4 An example of the proposed 2D denoising factor with 3 input images.

The second component of the proposed denoising factor is based on exposure time. More photons reach the camera sensor with a longer exposure time (Δt), which results in a more accurate reading. Thus, the proposed 2D weighting factor is designed to multiply the geometrically normalized exposure times with the luminance based denoising factor as

$$W(z, \Delta t_j) = \sqrt{\frac{\Delta t_j}{\sqrt[2]{\prod_{p=1}^P \Delta t_p}}} \cdot \omega(z) \quad (4.2)$$

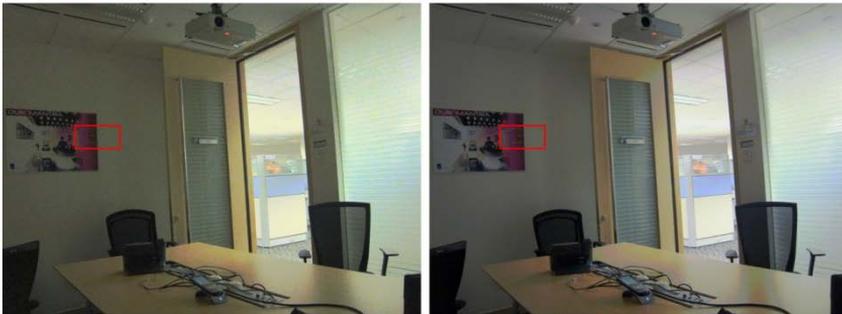
where j denotes the j th image among the P input images. The geometrical normalization avoids overwhelming big weights cause by images with very large exposure time. Thus, a modified objective function for estimating the CRF function is derived from [2] as

$$\sum_{i=1}^N \sum_{j=1}^P \{W(z, \Delta t_j) \cdot [g(Z_{i,p}) - \ln E_p - \ln \Delta t_i]\}^2 + \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} [W(z, \max(\Delta t_1, \dots, \Delta t_p)) \cdot g''(z)]^2 \quad (4.3)$$

Input Image Sequence:



Synthesized HDR Images:



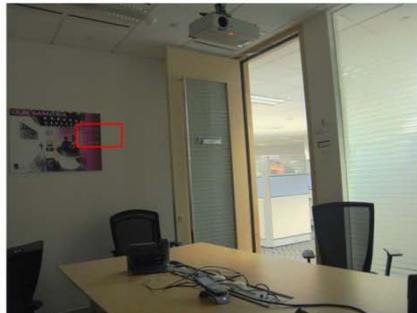
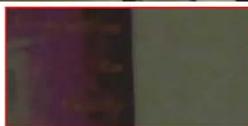
(a) Debevec and Malik 1997
[heavy noise from input images]



(c) Akyuz and Reinhard 2007
[long processing time due to luminance estimation]



(b) Yao et al. 2010
[blur due to spatial averaging]



(d) The proposed denoising factor
[best quality with only 1/5 of the processing time compare to method (c)]



Fig. 4.5 Visual quality comparison for 2D de-noising factor

The proposed denoising method is tested by comparing it with three HDR synthesis methods, as shown in Fig. 4.5. The noise is significantly reduced compared to [2] with the same processing time. No blur artifact is generated as in [105] due to pure spatial averaging. Comparing with [87], the quality of the proposed is about the same. It is achieved with only 1/5 of the processing time, since no intermediate steps for approximating response function is required.

4.3 NOISE REDUCED TONE-MAPPING

Tone-mapping is usually the last step in the software solution of HDR imaging to display the HDR content on LDR display devices. In general, the tone-mapping is designed by using either global operator [26, 27, 28] or local operator [29, 30, 31, 32, 106]. The benefits of global operator are fast, easy to implement and no lighting inverse. Although some studies [107, 108] show that global operators can produce good results in some scenarios, many research studies are conducted based on sophisticated local operators for better local contrast and fine details.

A popular local tone mapping approach is based on the *Retinex theory* [109] that an image (I) is regarded as a product of two components ($I = L \cdot R$): an *illuminance component* (L) which contains large luminance variance, and a *reflectance component* (R) which contains *intrinsic* information. *Gaussian filtering* method for decomposition was proposed in [110], and was soon be replaced by *bilateral filtering* [29, 111] due to its better edge preservation. However, *halo artifact* is usually seen in the compressed image due to edges leaking to the reflectance component. Thus, better edge-preserving methods are proposed. A *weighted least square* (WLS) framework is proposed in [34] with progressive detail layer decomposition. Other methods focused on the smoothness constraint [32, 112] and the localized data term [113] for better visual quality and fast processing.

However, despite these improvements the main problem still remains: when the original HDR image contains noise, it remains in the tone-mapping result. There are two ways of de-noising for HDR tone-mapping. One is to direct apply a single image de-noising method [95, 96] after the tone-mapping algorithm. This approach is more time costly. The other approach is to do noise reduction during the tone-mapping process, as proposed in [P7].

The process of the tone-mapping in [P7] is illustrated in Fig. 4.6, where the input HDR image is decomposed into a base layer and a detail layer [34]. The base layer contains all the high dynamic range content, and the detail layer contains the local contrast information. Noise is easily classified into detail layer as they share the same attributes as the small details, such as introducing small gradient and slightly increasing of local variance near the noise pixels.

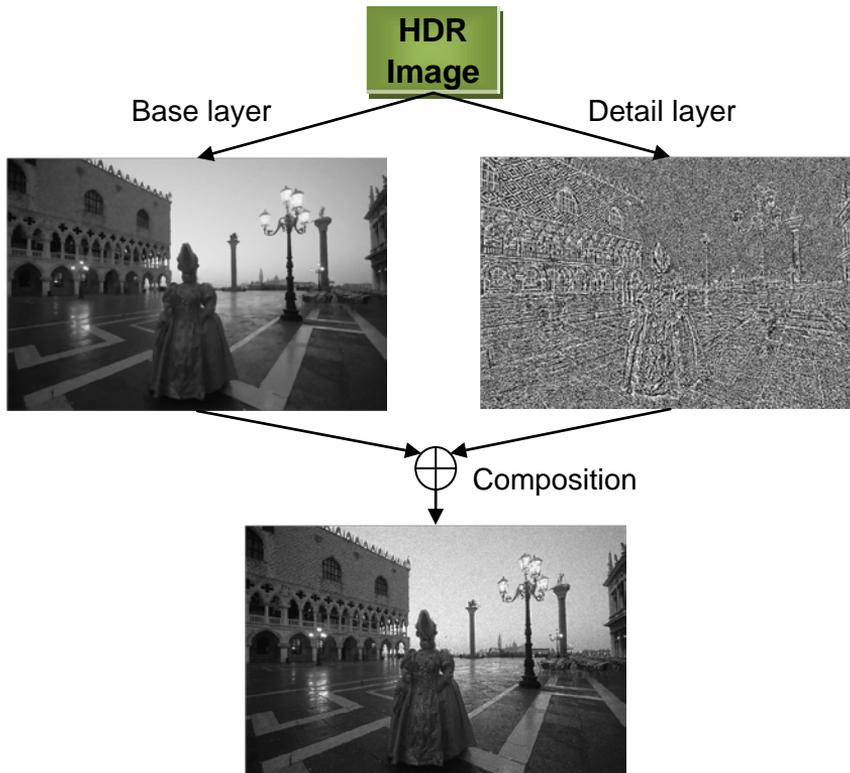


Fig. 4.6 Proposed HDR tone mapping process in [P7]

In [P7], a base layer *information content weighting* (ICW) and a detail layer ICW are proposed to reduce the noise in two steps. In edge-preserving decomposition, the data term defines the fidelity between the based layer and the original luminance. It is proposed in an L2 norm for easy calculation. First, the base layer ICW is introduced to be used on the data term as

$$\min_{B(p)} \left\{ \omega_b(I(p)) \cdot \sum_p (B(p) - I(p))^2 + \lambda \sum_p \left(\Phi \left(\frac{\partial B(p)}{\partial x}, a_x(p) \right) + \Phi \left(\frac{\partial B(p)}{\partial y}, a_y(p) \right) \right) \right\} \quad (4.4)$$

where ω_b is the spatial varying ICW based on perceptual of HVS, a_x and a_y are the smoothness coefficient, I denotes the log HDR luminance, B denotes the target base layer. Without the base layer ICW, the above equation is exactly the tone mapping using WLS edge-stopping filter [34]. Inspired by the information theory [114, 115, 116], the base layer ICW is defined as

$$\omega_b(p) = \frac{1}{2} \cdot \log \left(1 + \frac{\sigma^2(p)}{\sigma_c^2} \right) \quad (4.5)$$

where $\sigma(p)$ is the local variance at pixel p , and σ_c^2 represents the channel noising power, which is determined by the saliency coefficient (c) as

$$\sigma_c^2 = \vec{\Gamma}(c \cdot P) \quad (4.6)$$

where $\Gamma(p) = \{\sigma^2(p), p \in P\}$, and $\vec{\Gamma}$ denotes the ascending sort of Γ . For example, $c = 0.5$ indicates the median value of Γ .

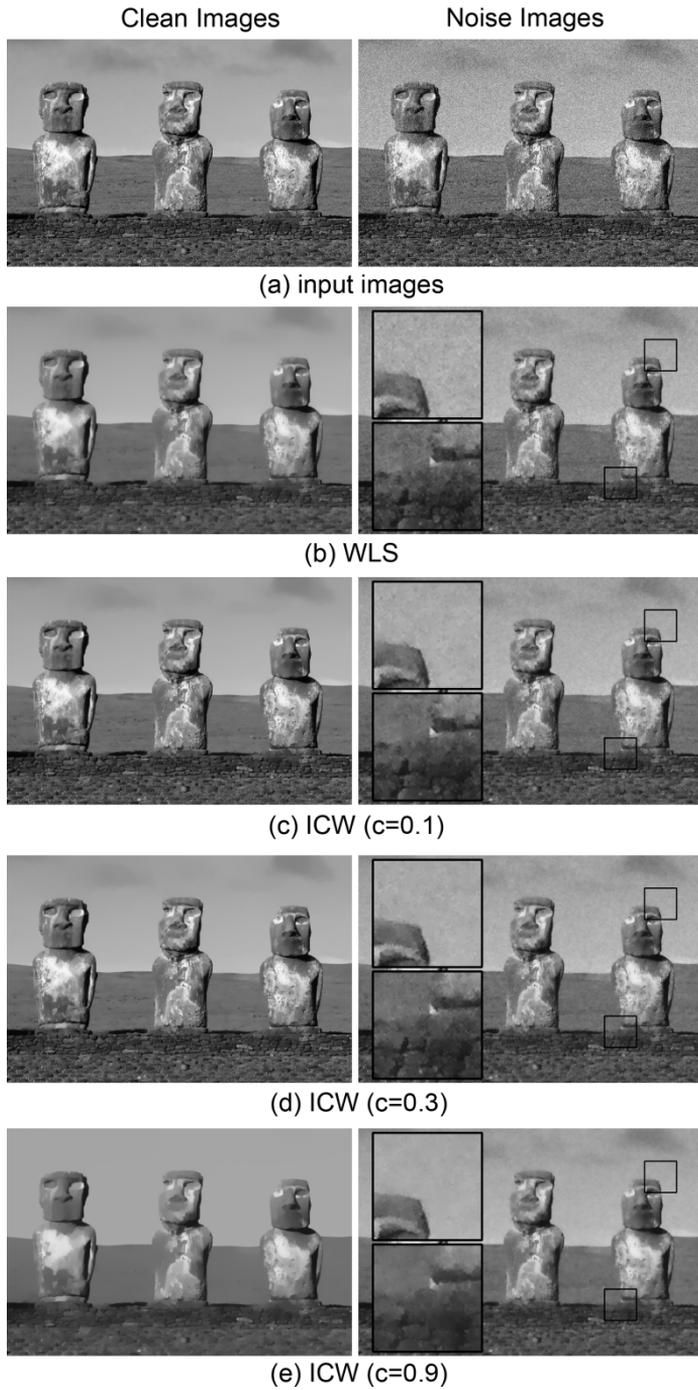


Fig. 4.7 Visual comparison of different saliency coefficients on (left column) clean image and (right column) noise image.

The saliency coefficient is an important factor that controls the saliency of the base layer ICW. A small saliency coefficient indicates a small channel noise, and results a high fidelity between base layer and the original image. A large saliency coefficient indicates a large channel noise, and results a smooth base layer. If the saliency coefficient is too big, the base layer can be over smoothed. A visual comparison of different saliency coefficient is illustrated in Fig. 4.7. In our implementation, $c = 0.3$ is chosen as a balanced value. However, if the noise level of the input HDR image is very high, a bigger saliency coefficient generates better denoising result.

During the decomposition process, the ICW helps to push the out of the base layer. Therefore, when the detail layer is subtracted from the clean base layer, more noise appears. Thus, a detail layer ICW is adopted as

$$D(p) = \omega_d(p) \cdot (I(p) - B(p)) \quad (4.7)$$

where ω_d is the detail layer ICW derived from information fidelity criteria [114, 115, 116] as

$$\omega_d(p) = \frac{1}{\alpha} \cdot \log_2(1 + \beta \cdot (\frac{\sigma_{B,I}(p)}{\sigma_B^2(p)})^\gamma) \quad (4.8)$$

where σ_B^2 denotes the local variance of the base layer in a small window, $\sigma_{B,I}$ denotes the covariance between the base layer and the original luminance, β and γ are two constants control the effectiveness of the weighting function, and α is a normalization factor which full fill the constraints of $\omega_d(p) \in [0,1]$. If $\omega_d(p)$ is bigger than 1, the detail layer is amplified. It is commonly used in detail enhancement, with the risk of amplifying the noise at the same time. An example of the effectiveness of the detail layer ICW is given in Fig. 4.8.

The de-noising level is controlled by the different selection of two smooth coefficients β and γ , as shown in Fig. 4.8. It is worth noticing that the required details may be hidden by the heavy noise. In that case, when removing the noise, the details are

smoothed out too. Thus, balanced smooth coefficients are carefully chosen.

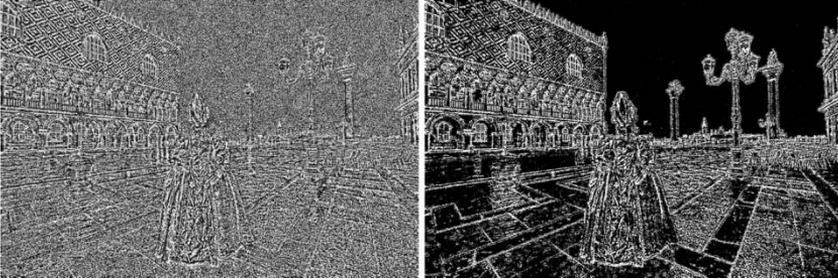


Fig. 4.8 Detail layer with (left) normal decomposition and (right) ICW-based de-noising factor

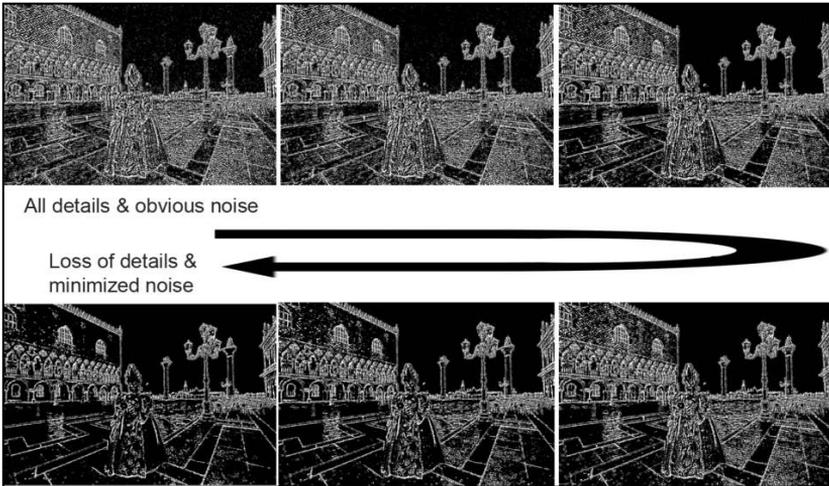


Fig. 4.9 Detail layer with different de-noising levels. The de-noising effect is increased from left to right, and from top to bottom.

We compared our solution with the approach of applying single image de-noising after tone mapping. We choose BM3D [97], which is one of the leading de-noising algorithms. The implementation was downloaded from the website provided by original authors¹. The interface is a Matlab function, but the core is implemented in C. In order to test against the full capability of

¹ Matlab code of BM3D is downloaded from <http://www.cs.tut.fi/~foi/GCF-BM3D/>

BM3D, the noise level sigma is varied from 15 to 70, in order to find the best PSNR for BM3D.

Gaussian white noise is added to the original clean HDR image with different signal-to-noise ratio vary from 10dB to 25dB. Three tone mapping solutions are tested using WLS, WLS+BM3D, and the proposed ICW. As shown in Table 4.1, on average, the result of the proposed ICW is 1dB less than the result of WLS+BM3D solution, but the processing time is significantly reduced. The big difference occurs when the noise level of the input image is high. The BM3D method is able to use the information from the neighboring blocks at the tradeoff of the processing time.

In some experiments, the proposed ICW-based solution outperforms the combination of WLS + BM3D, when the noise level of the input image is not high. A possible explanation is that the proposed ICW works directly on the full radiance map of HDR image, while BM3D works on the result of WLS, which contains less information due to compression.

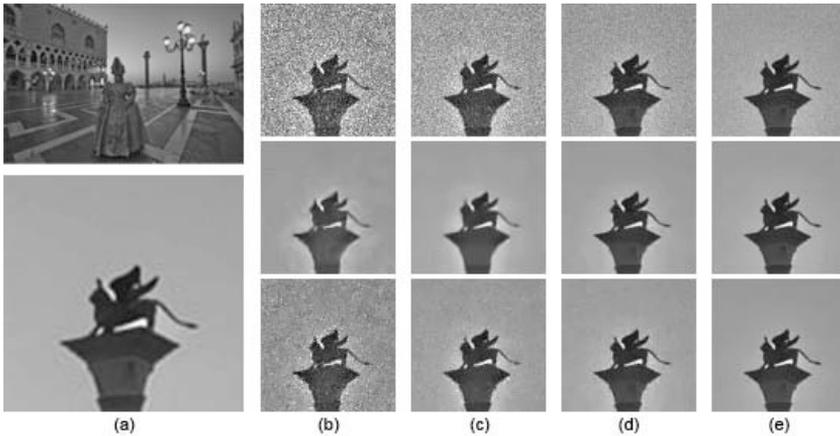


Fig. 4.10 Visual comparison with WLS+BM3D. (a) Clean image and a zoom-in object. The input image of (b-e) are added with different noise level at (b) SNR=10dB, (c) SNR=15dB, (d) SNR=20dB, and (e) SNR=25dB. In (b-e), the (top row) is WLS, the (middle row) is WLS+BM3D, and the (bottom row) is ICW proposed in [P7].

Visual comparison was conducted with five leading tone mapping algorithms based on different approaches. We selected the state-of-the-art global operator (2008) [108], and used the

implementation in Luminance HDR¹. The other four local operators were selected from a scale-decomposition-based solution (2005) [117], a bilateral-based decomposition (2007) [111], a slide-window-based direct luminance compression (2010) [118], and an edge-preserved decomposition (2008) [34]. All implementations were downloaded from the website of original authors. Visual comparison show good improvement in terms of noise reduction.

Table 4.1 Comparison with WLS+BM3D using quality metrics.

HDR image noise level	Method	WLS	WLS+BM3D	[P7]
SNR=10	PSNR	24.65	29.94	28.08
	SSIM	0.3949	0.8994	0.6869
SNR=15	PSNR	28.82	32.95	32.60
	SSIM	0.6194	0.9209	0.8913
SNR=20	PSNR	33.13	36.25	35.47
	SSIM	0.8230	0.9643	0.9456
SNR=25	PSNR	36.82	38.53	37.88
	SSIM	0.9330	0.9804	0.9710
Average Execution Time (Resolution 1200x800)		3 s	35 s	7 s

For simplicity, we applied the same method directly on the R, G, B color channels respectively, and therefore, color shift can be seen in some tone mapping result. Future work will focus on how to find in a better HDR color space for the least color distortions. Pioneer work about HDR color space is proposed in [119].

¹ Luminance HDR 2.3.1 is available at <http://qtpfsgui.sourceforge.net/>. It was previously named Qtpfsgui.

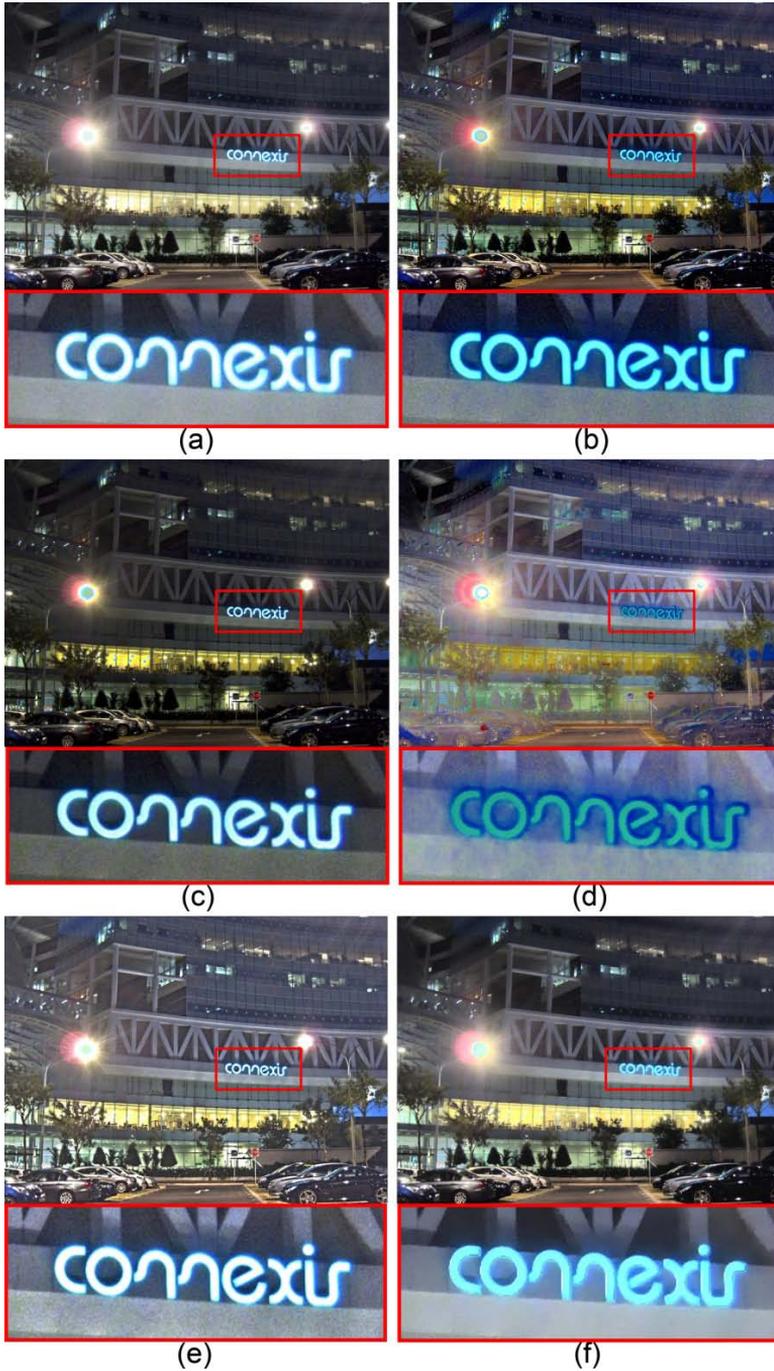


Fig. 4.11 Tone mapping visual comparison: (a) global tone curve [108], (b) scale-decomposition [117], (c) bilateral filtering decomposition [111], (d) direct luminance compression [118], (e) edge-preserved decomposition[34], and (f) noise reduced tone mapping based on ICW in [P7].

5 Image Difference Metrics

5.1 IMAGE DIFFERENCE METRICS FOR EXPOSURE BRACKETING IMAGES

Quality metrics have been well studied for both image and video processing in the LDR domain. In image quality assessment, when a pair of images is given, the measurement of their similarity is classified as *full-reference* measurement. Many intensity-based indices [120, 121] have been proposed for their simplicity and fast processing. The most straightforward method is the *peak signal to noise ratio* (PSNR) calculated from the *mean square error* (MSE). However, PSNR is not accurate in many cases as it does not consider *human visual system* (HVS) [120]. An HVS-based structural similarity (SSIM) index [120] was proposed and was applied to video coding[122].



Fig. 5.1 Images captured with moving people and vehicle using different exposures settings: (a) +1 EV, (b) 0 EV, and (c) -1 EV. Although they are captured at the same scene within a short period of time, the similarity index (SSIM) between each pair of images appears to be low at $SSIM(a,b)=0.6685$ and $SSIM(a,c)=0.4216$.

Although these indices are designed to focus on cases where the dynamic ranges and intensities of the two images are almost the same, in practical situation where processing of high dynamic range imaging requires different level of exposures [2], such indices would not serve the scenarios, as shown in Fig. 5.1. In fact, it is known that the dynamic range can also be changed

due to images or videos with different bit-depth[123]. Moreover, the images being compared are assumed to be well registered by the intensity-based indices. However, it is unavoidable that there are rotations and translations in a set of differently exposed LDR images, especially when they are captured for outdoor scenes. Thus, in order to test if an image sequence is captured at the same scene and can be used for an HDR synthesis, it is desirable to provide simple similarity indices that are robust to translation, rotation, intensity change and dynamic range between two input images.

In [P2], a structural similarity index for two LDR image by using intensity mapping and Richards' curve [124] is proposed. Due to the limitation of image un-alignment, the histogram based IMF [51] is used to unify the dynamic ranges or intensities of two images from the same scene[125]. On the other hand, when there is neither intensity nor dynamic range change between two images from the same scene, the accumulated histograms of two similar images are almost the same, and the structure of the corresponding image is kept even though it is mapped by the IMFs. The Richards' curve is adopted to design a switch on the deduction of the mean values from these two images. When two images are from the same scene, the similarity index value is calculated without the deduction of mean values from the corresponding images. As a result, the proposed index is robust to small translation and rotation between two images. When two images are from different scenes, the mean value is deducted from the corresponding images before the similarity index value is computed. As such, the proposed index is sensitive to two "similar" images from different scenes.

Different from the traditional quality metrics, the proposed image difference metrics does not contain any *mean-opinion-scores*. As such, it is designed to test if two images are from the same scene instead of measuring the image quality.

The proposed index is then extended by dividing the whole image into local windows as in[120]. The similarity of two images is detected by checking all pairs of local windows. Such

a local similarity index is insensitive to large intensity changes, dynamic range changes, translation and small rotations between images from the same scene.

The proposed image difference metrics start with mapping of two input images. Similar to the bi-directional comparison in [P4], the bi-directional mapping is written as

$$\dot{Z}_1(p) = \begin{cases} \Lambda_{1,2}(Z_1(p)), & \omega(Z_1(p)) \geq \omega(Z_2(p)) \\ Z_1(p), & \omega(Z_1(p)) < \omega(Z_2(p)) \end{cases} \quad (5.1)$$

$$\dot{Z}_2(p) = \begin{cases} \Lambda_{2,1}(Z_2(p)), & \omega(Z_2(p)) \geq \omega(Z_1(p)) \\ Z_2(p), & \omega(Z_2(p)) < \omega(Z_1(p)) \end{cases} \quad (5.2)$$

where ω is a triangle weighting function, $Z_1(p)$ and $Z_2(p)$ are the intensity value of the two input images at pixel p , Λ denotes the histogram based IMF [51], and $\dot{Z}(p)$ represents an intensity unified pixel. This intensity unified pixel is then further transformed using Richards' curve [124] as

$$\ddot{Z}_1(p) = \dot{Z}_1(p) - \left(1 - \Gamma\left(\beta(\dot{Z}_1, \dot{Z}_2)\right)\right) * \mu_{\dot{Z}_1} \quad (5.3)$$

$$\ddot{Z}_2(p) = \dot{Z}_2(p) - \left(1 - \Gamma\left(\beta(\dot{Z}_1, \dot{Z}_2)\right)\right) * \mu_{\dot{Z}_2} \quad (5.4)$$

where $\mu_{\dot{Z}_1}$ and $\mu_{\dot{Z}_2}$ are the mean values of \dot{Z}_1 and \dot{Z}_2 , respectively, $\beta(\dot{Z}_1, \dot{Z}_2)$ is the cosine of the angle between the two lexicographic order vectors of \dot{Z}_1 and \dot{Z}_2 , and $\Gamma(z)$ is the Richards' curve defined as

$$\Gamma(z) = \frac{1}{1 + \exp(\zeta_1(\zeta_2 - z))} \quad (5.5)$$

where ζ_1 and ζ_2 are the growth rate and the threshold of maximum growth. Both Γ and β are used to detect the similarity of two images. On one hand, when two images are from the same scene, the existence of intensity changes, small translation and rotation between them would appear in the value of Γ

wherein the value will be almost equal to 1. On the other hand, when two images are from different scenes, the value of Γ approaches 0.

The similarity index is defined as

$$S_s(Z_1, Z_2) = \frac{2\mu_{Z_1}\mu_{Z_2} + \varepsilon_1}{\mu_{Z_1}^2 + \mu_{Z_2}^2 + \varepsilon_1} \cdot \frac{\sum_p 2\check{Z}_1(p)\check{Z}_2(p) + \varepsilon_2}{\sum_p \check{Z}_1^2(p) + \check{Z}_2^2(p) + \varepsilon_2} \quad (5.6)$$

where ε_1 and ε_2 are two small positive constants and they are adopted to improve the robustness of the proposed index when the local signal to noise ratio is very low. Similar to SSIM index [120], the proposed similarity index is applied locally rather than globally, and the final index is the summation of all the local indices as

$$S_s^b(Z_1, Z_2) = \frac{1}{p} \sum_{i=1}^p S_s(Z_1, Z_2). \quad (5.7)$$

Experiments show similar result when the input images are of the same exposure level, while the proposed similarity index has great improvement when dealing with exposure bracketing images. Table 5.1 shows the metrics comparison based on the images in Fig. 5.1.

Table 5.1 Metrics comparison using images in Fig. 5.1. The similarity indices mark from 0 (least similar) to 1 (identical).

Image Pair	(1,2)	(1,3)
SSIM [120]	0.6685	0.4216
S_s [P2]	0.9828	0.9780
S_s^b [P2]	0.9681	0.9504

Table 5.2 Metrics comparison using images in Fig. 5.2. The similarity indices mark from 0 (least similar) to 1 (identical).

Image Pair	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
SSIM [120]	0.5064	0.4750	0.4791	0.3803	0.3048
S_s [P2]	0.9787	0.9357	0.9879	0.9674	0.9845
S_s^b [P2]	0.9220	0.8547	0.9414	0.8875	0.8706

Another example is given with the input images captured using hand held camera, shown in Fig. 5.2. The image sequence contains camera movement, moving object, and exposure changes. The propose image difference metrics show the robustness in Table 5.2.



Fig. 5.2 Sequence of images contain camera movement, moving object, and exposure difference of EV interval of 2/3. Images are marked as 1 to 6 from left to right, top to bottom.

5.2 IMAGE DIFFERENCE METRICS FOR HDR RADIANCE MAP

Due to lack of strict link between physical light radiance and image radiance map, HDR images with the same scene can be very different in terms of their direct peak signal-to-noise ratio (PSNR), as shown in Fig. 5.3.

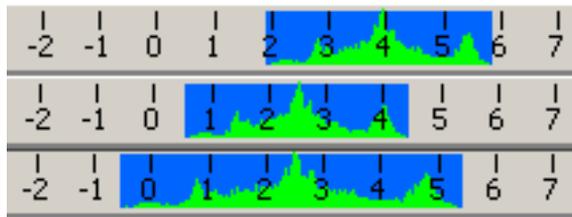


Fig. 5.3 Histogram in log radiance domain of (top) original HDR image, (mid) shifted HDR image, and (bottom) scale-invariant HDR image.

Near-identical HDR images are defined in [P3] as the images that carry the same scene information, but are shifted or have

different scale in the log radiance domain. These images are classified as

$$H = \{\text{HDRI}_q: R_q | \log(R_q) = a_q \cdot \log(R_0) + b_q\} \quad (5.8)$$

where R is the radiance of the HDR image, 0 represents the reference image defined as original, q represents the test image, a_q is the scale factor, and b_q is the shift factor. If the HDR image is composed from an LDR image set and the exposure info is known, the scale factor can be estimated as

$$a_q = \frac{\log(k_{q,1,2})}{\log(k_{0,1,2})} \quad (5.9)$$

where $k_{1,2}$ represents the exposure ratio between the first two images in the bracketing LDR image set. If the scale factor equals to 1, we call this HDR image shift-identical to the reference HDR image, and the shift factor is calculated as

$$b_q = \log(\Delta t_{0,1}) - \log(\Delta t_{q,1}) \quad (5.10)$$

where Δt denotes the exposure time of the first LDR image in the sequence. If the scale factor does not equal to 1, we name it scale-identical, where the value of the shift factor is related to the scale factor and the image radiance.

When comparing the similarity of two HDR images, they are assumed to be near-identical. Thus, the scale factor is estimated using

$$a_q = \frac{\max(h|\phi_0(h) > \varepsilon) - \min(h|\phi_0(h) > \varepsilon)}{\max(h|\phi_q(h) > \varepsilon) - \min(h|\phi_q(h) > \varepsilon)} \quad (5.11)$$

where ϕ is the histogram function and ε is a small number for the robustness to noise. The shift factor is calculated by correlating two HDR radiance maps using a sliding window. The biggest coefficient represents the best fit as

$$b_q = \frac{\kappa(h) - H}{H} \cdot (M - m) \quad (5.12)$$

where H is the total number of bins in plotting the histogram, M and m are the maximum and minimum of the input radiance map, $\kappa(h) = \{h | \sigma(h) = \max(\sigma(i)), i = 1, 2, \dots, 2H\}$, and σ is the correlation coefficient function corresponding to the starting bin number of the sliding window. Finally, the proposed HDR metrics is calculated as

$$\text{PSNR}_{\text{HDR}} = 10 \cdot \log_{10} \left(\frac{P \cdot \Phi^2}{\sum_{p=1}^P [a_q \cdot \log(R_{q,p}) + b_q - \log(R_{0,p})]^2} \right) \quad (5.13)$$

where Φ is a constant.

Table 5.3 Compose HDR images with the following parameters in six experiments.

	$\Delta t_1(s)$	ΔEV_1	ΔEV_2	ΔEV_3	ΔEV_4
Original	0.02	1	1	1	1
EXP1	0.50	1.00	0.98	0.99	0.96
EXP2	0.50	1.58	1.53	1.54	1.51
EXP3	0.50	0.32	0.93	0.49	1.58
EXP4	0.02	1.32	1.42	0.58	0.74
EXP5	0.50	0.15	0.17	3.22	3.60
EXP6	0.03	0.26	0.32	1.00	3.32

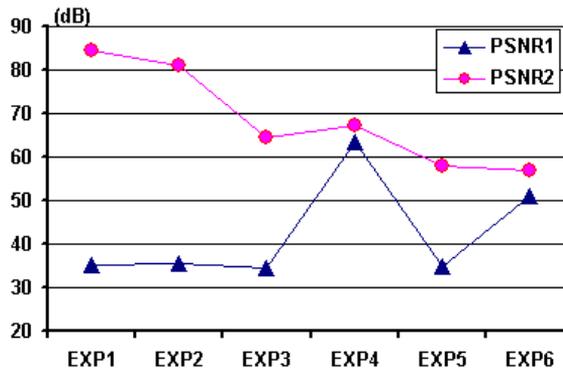


Fig. 5.4 Chart of PSNR from EXP1-6 from Table 5.1, where PSNR1 is the result of direct pixel-by-pixel comparison and PSNR2 is the result of [P3].

An example is shown with HDR images synthesized with different parameters, as shown in Table 5.3. They are then compared with original HDR image using direct comparison and the proposed metrics. Results are illustrated in Fig. 5.4. In EXP1 and EXP2, despite the difference of initial exposure time (Δt_1) and small distortion of EV intervals ($\Delta EV_i = \log_2(\Delta t_{i+1}/\Delta t_i)$), the relative relations among ΔEV_1 to ΔEV_4 remains constant. The composed HDR images are verified to have high similarity in comparison to the original image, while the direct comparison shows a very low PSNR value in Fig. 5.4. In EXP3-6, random numbers are given for all parameters. The proposed metrics find the distortion and shown a drop of PSNR at average of 20dB. On the contrary, the direct comparison method was able to recognize the distortions.

6 *Summary of Contributions*

6.1 CONTRIBUTION OF THE THESIS

[P1]: A new intensity mapping, named inter-pixel relationship function (IRF), was proposed to map in between two images with different exposures by using their spatial correlations. The IRF has a variety of applications related to HDR imaging. In this paper, a dual patching is demonstrated using IRF with 10dB improvement compared to single directional patching using camera response function (CRF).

[P2]: In this paper, a structural similarity index is proposed for identifying if two images are captured from the same scene and can be used for HDR synthesis. Different from traditional quality metrics, the proposed method is robust to different dynamic range, small moving object, as well as possibly small rotation and translation. Therefore, it is useful to find exposure bracketing sequence from a large image database for batch processing.

[P3]: In this paper, we introduce the concept of near-identical HDR images, which carry the same scene information with different radiance representations. The same set of near-identical HDR images can be transformed to each other with linear functions in log radiance domain, which is useful in displaying HDR images and HDR tone mapping. By using the log histogram of the radiance map, we manage to derive the scale and shift parameters for the near-identical HDR images, which will help to get an accurate assessment using PSNR. A simple application of how to compose near-identical HDR images using IRF [P1] is also presented in the paper.

[P4]: A real-time framework to detect and remove the ghost artifact is proposed based on IRF and bi-directional comparison. The algorithm delivers robust ghost removal in $O(n)$ time, which makes it a good candidate for mobile application.

[P5]: A robust de-ghosting algorithm is proposed to solve an extreme scenario where a moving object occupies a large area and distort the spatial correlation. The distorted spatial correlation is corrected by using pixel intensity distribution. A statistical threshold model is trained from the image database, and the key parameters are determined on the fly. Experiments show that the proposed algorithm achieves good visual quality in both day-time and night-time lighting conditions.

[P6]: In this paper, a sub-sampling based moving object detection is proposed to shorten the de-ghosting processing time for images with large resolution. Experiments show that the detection speed can be increased by 50%. Combining with this sub-sampling scheme, the real-time de-ghosting framework proposed in [P4] is capable of running on mobile device to de-ghost on three 8MB image within 700ms.

[P7]: A noise reduced tone-mapping is proposed to suppress the noise from the input HDR image. The de-noising process is embedded in the tone-mapping process based on the information content weighting. Therefore, it requires very little additional processing time. The experiments show a significant improvement in the visual quality of final tone-mapping image in terms of noise reduction.

[P8]: A 2D de-noising factor is proposed to generate noise reduced HDR image from a set of noise LDR images. Contrary to existing solutions that operates in the HDR radiance domain, the proposed de-noising factor works directly on the pixels of input LDR images, which reduces the processing time to 20% of the radiance-based method.

6.2 SUMMARY OF RESULTS

The publications [P1-P8] have covered a large portion of the HDR imaging value chain, as shown in Fig. 1.3.

Technologies described in [P1, P4-6] are implemented in our PC application and mobile application. An example of three input images with moving object is shown in Fig. 6.1. Results from two state-of-the-art algorithms, the entropy-based deghosting algorithm [82] and the patch-based iterative deghosting algorithm [126], are shown in Fig. 6.2 and Fig. 6.3 respectively. The former is good in the background sky but fails in the foreground people, while the latter performs the opposite way.

We also compare with two leading image processing applications, Photoshop and Photomatix. The results are illustrated in Fig. 6.4 and Fig. 6.5, where the ghosting artifacts are clearly visible. Our proposed method, as shown in Fig. 6.6, generates clear image without ghosting artifacts.

Except the entropy-based algorithm [82], which requires to be running on MacOS, all the rest four are tested on the same laptop. Our proposed method uses the shortest processing time of 1.5 seconds. The others vary from 2.5 seconds to 350 seconds.



Fig. 6.1 Three input images captured with moving object at different exposure levels with 2EV apart from each other. The image resolution is 2144x1424. No camera movement is involved.



Fig. 6.2 HDR image synthesized using entropy-based de-ghosting algorithm proposed in [82]. Obvious ghosting artifacts are seen in red and green blocks. The implementation was downloaded from original author's website in the binary of PhotoSphere¹. The processing time was around 5 seconds on a Mac Pro desktop. An HDR image was synthesized from the binary app, and it was then been compressed using Photomatix through tone mapping.

¹ PhotoSphere is a MacOS based binary implemented by the original authors. It is downloaded from <http://www.anywhere.com/>.



Fig. 6.3 HDR image synthesized using patch-based iterative approach in [126]. The de-ghosting performs well at foreground object in the red and green blocks. However, it fails at sky region. Obvious halo artifacts are seen around the palm tree leaves, and ghosting artifacts are seen in the blue block. The implementation was downloaded from original author's website¹ in the format of Matlab interface calling C routines. The whole processing cost 350 seconds.

¹ Source code was downloaded from author's project webpage at <http://ece.ucsb.edu/~psen/hdr>.



Fig. 6.4 HDR image synthesized using Photoshop CS5. The sky in the blue block is well kept, while some under the palm tree are distorted. The foreground people are heavily distorted as seen in the red and green block. The processing cost around 7 seconds on a Dell Precision M6700 laptop.



Fig. 6.5 HDR image synthesized using Photomatix Pro 4.2. The sky in the blue block performs well, while the foreground people suffer from ghosting artifacts, as seen in the red and green blocks. The processing time is around 2.5 seconds on a Dell Precision M6700 laptop.



Fig. 6.6 HDR image synthesized using our algorithm. A clear and sharp image is presented for both foreground people and background sky. The processing time is 1.5 seconds on a Dell Precision M6700 laptop.

State-of-the-art tone mapping algorithms usually focus on detail preservation and local contrast enhancement. We make use of information content weightings and proposed a noise-reduced tone mapping in [P7]. White Gaussian noise (SNR=20dB) is added in the original HDR image.



Fig. 6.7 Tone mapping using tone curve [108]. Implementation is found in Luminance HDR.



Fig. 6.8 Tone mapping using subbands decomposition [117]. Implementation is found from authors' website¹.

¹ Matlab source code is provided at http://www.mit.edu/~yzli/hdr_compinging.htm



Fig. 6.9 Tone mapping using edge-preserved decomposition [34]. Implementation is found from authors' website¹.



Fig. 6.10 Noise reduced tone mapping proposed in [P7].

We compare with three leading tone mapping algorithms: global tone curve [108] as shown in Fig. 6.7, subbands decomposition [117] as shown in Fig. 6.8, and edge-preserved

¹ Matlab source code is provided at <http://www.cs.huji.ac.il/~danix/epd/>

decomposition [34] as shown in Fig. 6.9. Result of the proposed method is shown in Fig. 6.10.

It is important to test if a sequence of images is from the same scene, and can be used for HDR synthesis. A useful application is the batch processing for generating HDR images from a large database. Fig. 6.11 shows five pair of images captured from the same scene with different exposures. The proposed method from [P2] demonstrates the robustness to the exposure change.

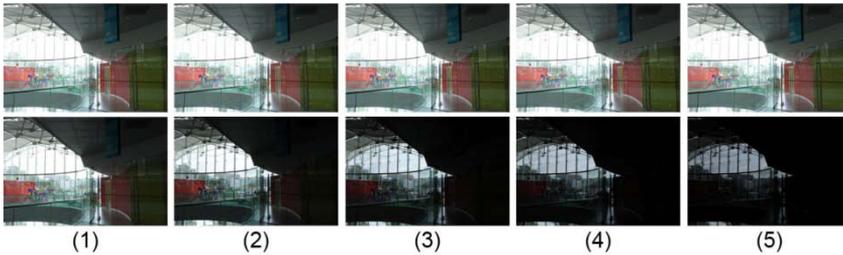


Fig. 6.11 Five image pairs captured at the same scene with different exposures. The exposure value differences for the five pair of images are (1) 2/3EV, (2) 4/3EV, (3) 2EV, (4) 8/3EV, and (5) 10/3EV.

Table 6.1 Comparison of similarity indices using image pairs from Fig. 6.11

Image Pair	(1)	(2)	(3)	(4)	(5)
SSIM [120]	0.8638	0.5502	0.3229	0.1985	0.1067
S_s [P2]	0.9999	0.9999	0.9999	0.9999	0.9999
S_s^b [P2]	0.9994	0.9972	0.9917	0.9845	0.9855

Another five pairs of images are shown in Fig. 6.12, where pair 1 shows two images with moving objects and slightly exposure difference, and pair 2 and 3 are captured using hand-held camera with camera movement. The proposed image difference metrics [P2] is robust to all these changes. Pair 4 and 5 are from two different scenes but looks similar. The proposed indices are able to identify the differences, as shown in Table 6.2.

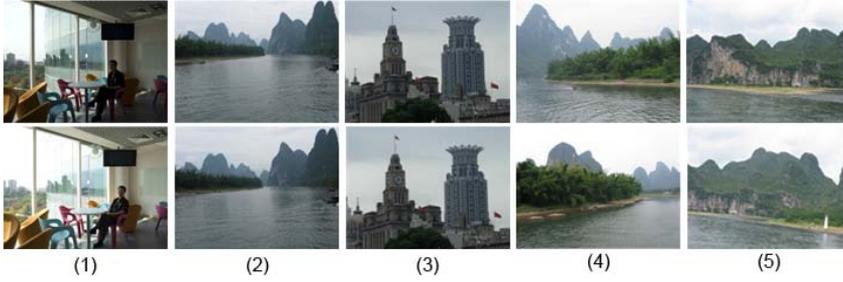


Fig. 6.12 Five image pairs of different scenarios: (1) moving object in the image; (2) camera movement; (3) camera movement; (4) similar scene; and (5) similar scene. Pairs of (1)-(3) are suitable for HDR synthesis, while (4)-(5) are not.

Table 6.2 Comparison of similarity indices using image pairs from Fig. 6.12

Image Pair	(1)	(2)	(3)	(4)	(5)
SSIM [120]	0.4750	0.5688	0.6454	0.4730	0.4725
S_s [P2]	0.9357	0.9199	0.9634	0.6466	0.7193
S_s^b [P2]	0.8547	0.9153	0.8932	0.4802	0.5030

7 Conclusion

In this thesis, we have studied a high quality software solution for HDR imaging, including intensity mapping, de-ghosting, noise-reduced HDR synthesis, noise-reduced tone mapping and simple image difference metrics.

De-ghosting is an important step to remove artifacts caused by moving object. A real-time solution is proposed for the application of mobile devices. An extended robust solution is proposed for handling moving object covering large area. Both solutions require moving object detection and correction by using intensity mapping. In the future, a model of how to detect moving object without hard threshold shall be studied.

We have proposed to incorporate de-noising during the process of HDR synthesis and tone mapping respectively. The major advantage is computational efficiency. A 2D de-noising factor based on signal-to-noise ratio is proposed in the noise-reduced HDR synthesis. An information content based de-noising factor is proposed based on human visual system in the noise-reduced tone mapping. Both two de-noising steps operate in the RGB domain. In future work, better HDR luminance color space [119] shall be considered.

An image difference metrics for measuring whether a pair of images is from the same scene is proposed to identify correct image sequence from a large image database for the synthesis of HDR images. It is robust to limited distortions, including small image un-alignment, small moving object and large luminance change due to exposure difference. Another metrics measures the similarity of two HDR images using histogram-based transform. This is still a very preliminary work for difference measures. In the future more sophisticated dynamic range independent quality metrics [127, 128] can be studied, which can measure images with different bit depth.

8 Reference

- [1] E. Reinhard, G. Ward, S. Pattanaik and P. Debevec, High dynamic range imaging: Acquisition, display and image-based lighting, San Mateo, CA: Morgan Kaufmann, 2005.
- [2] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of Annual Conference on Computer Graphics and Interactive Techniques*, 1997.
- [3] A. R. Varkonyi-Koczy, A. Rovid and T. Hashimoto, "Gradient-based synthesized multiple exposure time color HDR image," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 8, pp. 1779-1785, 2008.
- [4] F. Banterle, A. Artusi, K. Debattista and A. Chalmers, Advanced high dynamic range imaging: theory and practice, CRC Press, 2011.
- [5] J. J. McCann and A. Rizzi, The art and science of HDR imaging, John Wiley & Sons, 2011.
- [6] K. Myszkowski, R. Mantiuk and G. Krawczyk, High dynamic range video, Morgan & Claypool Publishers, 2008.
- [7] C. Bloch, The HDRI handbook: high dynamic range imaging for photographers and CG artists, Rocky Nook, 2007.
- [8] J. Tumblin, A. Agrawal and R. Raskar, "Why I want a gradient camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [9] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000.
- [10] G. Wetzstein, I. Ihrke and W. Heidrich, "Sensor saturation in fourier multiplexed imaging," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

- [11] S. K. Nayar, V. Branzoi and T. E. Boult, "Programmable imaging using a digital micromirror array," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [12] S. K. Nayar, V. Branzoi and T. E. Boult, "Programmable imaging: towards a flexible camera.," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 7-22, 2006.
- [13] S. Mann and R. W. Picard, "On being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures," in *Proceedings of IS&T*, 1995.
- [14] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Fort Collins, CO, 1999.
- [15] M. A. Robertson, S. Borman and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging*, vol. 12, no. 2, pp. 219-228, 2003.
- [16] M. Granados, B. Ajdin, M. Wand and C. Theobal, "Optimal HDR reconstruction with linear digital cameras," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.
- [17] S. B. Kang, M. Uyttendaele, S. Winder and R. Szeliski, "High dynamic range video," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, p. 319-325, 2003.
- [18] G. Impoco, S. Marsi and G. Ramponi, "Adaptive reduction of the dynamics of HDR video sequences," in *IEEE International Conference on Image Processing (ICIP)*, 2005.
- [19] S. Mangiat and J. Gibson, "Spatially adaptive filtering for registration artifact removal in HDR video," in *IEEE International Conference on Image Processing (ICIP)*, 2011.
- [20] D. Li, P. Tao and J. Wen, "High dynamic range video with synthesized gain control," in *IEEE International Conference on Emerging Signal Processing Applications (ESPA)*, 2012.
- [21] Y. Schechner and S. K. Nayar, "Generalized mosaicing: high dynamic range in a wide field of view," *International Journal of Computer Vision*, vol. 53, no. 3, p. 245-267, 2003.

- [22] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *International Journal of Computer Vision*, vol. 58, no. 1, pp. 7-17, 2004.
- [23] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes and S. K. Nayar, "Optical splitting trees for high-precision monocular imaging," *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 32-42, 2007.
- [24] M. Tocci, C. Kiser, N. Tocci and P. Sen, "A versatile HDR video production system," *ACM Transaction on Graphics*, vol. 30, no. 4, p. 41, 2011.
- [25] J. Kronander, S. Gustavson, G. Bonnet and J. Unger, "Unified HDR reconstruction from raw CFA data," in *Computational Photography (ICCP)*, 2013.
- [26] A. Pardo and G. Sapiro, "Visualization of high dynamic range images," *IEEE Transactions on Image Processing*, vol. 12, no. 6, pp. 639-647, 2003.
- [27] J. Duan and G. P. Qiu, "Fast tone mapping for high dynamic range images," in *International Conference on Pattern Recognition*, 2004.
- [28] C. H. Liu, "Image characteristic oriented tone mapping for high dynamic range images," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2008.
- [29] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 257-266, 2002.
- [30] R. Fattal, D. Lischinski and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 249-256, 2002.
- [31] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 267-276, 2002.
- [32] Z. G. Li, S. Rahardja, S. S. Yao, J. H. Zheng and W. Yao, "High dynamic range compression by half quadratic regularization," in *IEEE International Conference on Image Processing (ICIP)*, 2009.
- [33] T. Jinno, K. Mouri and M. Okuda, "HDR video tone mapping

- based on gamma blending," in *IEEE International Conference on Image Processing (ICIP)*, 2010.
- [34] Z. Farbman, R. Fattal, D. Lischinski and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, p. 67, 2008.
- [35] 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 (TOG)*, vol. 23, no. 3, pp. 760-768, 2004.
- [36] O. Bimber and D. Iwai, "Superimposing dynamic range," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5, pp. 1-8, 2008.
- [37] O. Bimber, T. Amano, A. Grundhofer, D. Kurz, D. Klock, S. Thiele and F. Hantsch, "Closed-loop feedback illumination for optical inverse tone-mapping in light microscopy," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 6, pp. 857-870, 2010.
- [38] R. Hoskinson, B. Stoeber, W. Heidrich and S. Fels, "Light reallocation for high contrast projection using an analog micromirror array," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 6, pp. 165:1-10, 2010.
- [39] A. A. Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing*, vol. 23, no. 6, pp. 611-618, 2005.
- [40] R. C. Bilcu, A. Burian, A. Knuutila and M. Vehvilainen, "High dynamic range imaging on mobile devices," in *IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2008.
- [41] T. Mertens, J. Kautz and F. V. Reeth, "Exposure fusion," in *Pacific Conference on Computer Graphics and Applications*, 2007.
- [42] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 358-365, 2012.
- [43] M. Tico, N. Gelfand and K. Pulli, "Motion-blur-free exposure fusion," in *IEEE International Conference on Image Processing (ICIP)*, 2010.

- [44] T. Kartalov, Z. Ivanovski and L. Panovski, "Fully automated exposure fusion algorithm for mobile platforms," in *IEEE International Conference on Image Processing (ICIP)*, 2011.
- [45] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *IEEE International Conference on Computer Vision (ICCV)*, 1993.
- [46] M. D. Grossberg and S. K. Nayar, "High dynamic range from multiple images: Which exposures to combine?," in *ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, 2003.
- [47] N. Barakat, A. N. Hone and T. E. Darcie, "Minimal-bracketing sets for high dynamic range image capture," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1864-1875, 2008.
- [48] K. Hirakawa and P. J. Wolfe, "Optimal exposure control for high dynamic range imaging," in *IEEE International Conference on Image Processing (ICIP)*, 2010.
- [49] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel and H. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [50] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *IEEE Transaction on Image Processing*, vol. 9, no. 8, pp. 1389-1406, 2000.
- [51] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: What is knowable?," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 25, no. 11, p. 1455-1467, 2003.
- [52] "Wikipedia Reciprocity," Wikipedia, [Online]. Available: [http://en.wikipedia.org/wiki/Reciprocity_\(photography\)](http://en.wikipedia.org/wiki/Reciprocity_(photography)).
- [53] J. C. Hung, C. H. Hwang, Y. C. Liao, N. C. Tang and T. J. Chen, "Exemplar-based image inpainting base on structure construction," *Journal of Software*, vol. 3, no. 8, pp. 57-64, 2008.
- [54] A. Criminisi, E. Perez and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE*

- Transactions on Image Processing*, vol. 13, no. 9, pp. 1200-1212, 2004.
- [55] A. Bugeau, M. Bertalmio, V. Caselles and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2634-2645, 2010.
- [56] Z. B. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1153-1165, 2010.
- [57] H. James and A. E. Alexei, "Scene completion using millions of photographs," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, p. 4, 2007.
- [58] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement," in *Proceedings of Vision Modeling and Visualization*, 2006.
- [59] J. H. Zheng, Z. G. Li, Z. J. Zhu and S. Rahardja, "A hybrid patching scheme for high dynamic range imaging," in *Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2011.
- [60] J. H. Zheng, Z. G. Li, Z. J. Zhu, S. Q. Wu and S. Rahardja, "Patching of moving objects for ghosting-free HDR synthesis," in *ACM SIGGRAPH Poster*, 2012.
- [61] Z. G. Li, Z. J. Zhu, S. Q. Wu, J. H. Zheng and S. Rahardja, "Fast correction of moving regions for differently exposed images with moving objects," in *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2012.
- [62] Z. G. Li, Z. J. Zhu, S. Q. Wu and S. Rahardja, "Fast patching of moving regions for high dynamic range imaging," in *ACM SIGGRAPH Asia 2010 Posters*, 2010.
- [63] Z. G. Li, S. Q. Wu, Z. J. Zhu, S. L. Xie and S. Rahardja, "Anti-ghost of Differently Exposed Images with Moving Objects," in *IEEE International Conference on Image Processing (ICIP)*, 2012.
- [64] H. L. Eng, J. X. Wang, A. H. K. S. Wah and W. Y. Yau, "Robust human detection within a highly dynamic aquatic environment in real time," *IEEE Transactions on Image*

- Processing*, vol. 15, no. 6, pp. 1583-1599, 2006.
- [65] L. Y. Li, W. M. Huang, I. Y. H. Gu and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459-1472, 2004.
- [66] A. Monnet, A. Mittal, N. Paragios and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *IEEE International Conference on Computer Vision (ICCV)*, 2003.
- [67] K. K. Hadziabdic, J. H. Telalovic and R. Mantiuk, "Comparison of deghosting algorithms for multi-exposure high dynamic range imaging," in *Proc. of Spring Conference on Computer Graphics*, 2013.
- [68] K. Simonson, S. Drescher and F. Tanner, "A statistics based approach to binary image registration with uncertainty analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 29, no. 1, pp. 112-125, 2007.
- [69] D. G. Lowe, "Object recognition from local scale-invariant features," in *IEEE International Conference on Computer Vision (ICCV)*, 1999.
- [70] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [71] G. Ward, "Fast robust image registration for compositing high dynamic range photographs from hand-held exposures," *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17-30, 2003.
- [72] F. M. Candocia, "On the featureless registration of differently exposed images," in *Proceedings of International Conference on Imaging Science, Systems and Technology*, 2003.
- [73] L. Cerman and V. Hlavac, "Exposure time estimation for high dynamic range imaging with hand held camera," in *Computer Vision Winter Workshop*, 2006.
- [74] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2007.

- [75] M. Gevrekci, "On geometric and photometric registration of images," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.
- [76] B. Guthier, S. Kopf and W. Effelsberg, "Histogram-based image registration for real-time high dynamic range videos," in *IEEE International Conference on Image Processing (ICIP)*, 2010.
- [77] A. Mittal and N. K. Paragios, "Motion-based background subtraction using adaptive kernel density," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [78] Y. Ren, C. S. Chua and Y. K. K. Ho, "Motion detection with non-stationary background," in *International Conference on Image Analysis and Processing*, 2001.
- [79] T. Mitsui and H. Fujiyoshi, "Object detection by joint features based on two-stage boosting," in *IEEE International Conference on Computer Vision (ICCV)*, 2009.
- [80] Y. K. Chen, T. Y. Cheng and S. T. Chiu, "Motion detection with using theory of entropy," in *IEEE International Symposium on Industrial Electronics*, 2009.
- [81] E. A. Khan, A. O. Akyuz and E. Reinhard, "Ghost removal in high dynamic range images," in *IEEE International Conference on Image Processing (ICIP)*, 2006.
- [82] K. Jacobs, C. Loscos and G. Ward, "Automatic high dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications*, vol. 128, no. 2, pp. 84-93, 2008.
- [83] Z. G. Li, Z. J. Zhu, S. L. Xie, S. Q. Wu and S. Rahardja, "Robust generation of high dynamic range images," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2010.
- [84] O. Gallo, N. Gelfand, W. C. Chen, M. Tio and K. Pulli, "Artifact-free high dynamic range imaging," in *IEEE International Conference on Computational Photography*, 2009.
- [85] "Wolfram Math World," [Online]. Available: <http://mathworld.wolfram.com/BezierCurve.html>.
- [86] "Wikipedia Bazier Curve," Wikipedia, [Online]. Available: http://en.wikipedia.org/wiki/B%C3%A9zier_curve.
- [87] A. O. Akyuz and E. Reinhard, "Noise reduction in high

- dynamic range imaging," *Journal of Visual Communication and Image Representation*, vol. 18, no. 5, pp. 366-376, 2007.
- [88] C. Liu, W. T. Freeman, R. Szeliski and S. B. Kang, "Noise estimation from a single image," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.
- [89] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 16, no. 3, pp. 267-276, 1994.
- [90] K. H. Lundberg, "Noise sources in bulk CMOS," 2002.
- [91] N. C. Gallagher and G. L. Wise, "A theoretical analysis of the properties of median filters," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, p. 1136-1141, 1981.
- [92] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *IEEE International Conference on Computer Vision (ICCV)*, 1998.
- [93] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 12, no. 7, pp. 629-639, 1990.
- [94] 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*, vol. 12, no. 11, p. 1338-1351, 2003.
- [95] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, p. 2866-2878, 2006.
- [96] C. Kervrann and J. Boulanger, "Local adaptivity to variable smoothness for exemplar-based image denoising and representation," *International Journal of Computer Vision*, vol. 79, no. 1, p. 45-69, 2008.
- [97] K. Dabov, A. Foi, V. Katkovnik and K. O. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, p. 2080-2095, 2007.
- [98] P. Chatterjee and P. Milanfar, "Is denoising dead?," *IEEE*

- Transactions on Image Processing*, vol. 19, no. 4, p. 895–911, 2010.
- [99] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1635-1649, 2012.
- [100] A. C. Bovik, *Handbook of image and video processing*, Academic Press, 2010.
- [101] W. Yao, Z. G. Li and S. Rahardja, "Intensity mapping function based weighted frame averaging for high dynamic range imaging," in *IEEE International Conference on Industrial Electronics and Applications (ICIEA)*, 2011.
- [102] A. A. Bell, C. Seiler, J. N. Kaftan and T. Aach, "Noise in high dynamic range imaging," in *IEEE International Conference on Image Processing (ICIP)*, 2008.
- [103] S. D. Neve, B. Goossens, H. Luong and W. Philips, "An improved HDR image synthesis algorithm," in *IEEE International Conference on Image Processing (ICIP)*, 2009.
- [104] B. Goossens, H. Luong, J. Aelterman, A. Pizurica and W. Philips, "Reconstruction of high dynamic range images with Poisson noise modeling and integrated denoising," in *IEEE International Conference on Image Processing (ICIP)*, 2011.
- [105] W. Yao, Z. G. Li, S. Rahardja, S. S. Yao and J. H. Zheng, "Half-quadratic regularization based de-noising for high dynamic range image synthesis," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010.
- [106] R. C. Bilcu, S. Alenius and M. Vehvilainen, "Adaptive local tone mapping of color images," in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2012.
- [107] A. O. Akyuz, R. Fleming, B. E. Riecke, E. Reinhard and H. H. Bulthoff, "Do HDR displays support LDR content?: A psychophysical evaluation," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 38:1-38:7, 2007.
- [108] R. Mantiuk, S. Daly and L. Kerofsky, "Display adaptive tone mapping," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 68:1-68:10, 2008.
- [109] E. H. Land and J. J. McCann, "Lightness and retinex theory,"

- Journal of the Optical Society of America*, vol. 61, no. 1, pp. 1-11, 1971.
- [110] D. Jobson, Z. U. Rahman and G. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing (TIP)*, vol. 6, no. 3, pp. 451-462, 1997.
- [111] J. Kuang, G. M. Johnson and M. D. Fairchild, "icam06: A refined image appearance model for hdr image rendering," *Journal of Visual Communication and Image Representation*, vol. 18, no. 5, pp. 406-414, 2007.
- [112] G. Guarnieri, S. Marsi and G. Ramponi, "High dynamic range image display with halo and clipping prevention," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1351-1362, 2011.
- [113] B. Gu, W. Li, M. Zhu and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 70-79, 2013.
- [114] H. R. Sheikh, A. C. Bovik and G. Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117-2128, 2005.
- [115] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185-1198, 2011.
- [116] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *IEEE International Conference on Image Processing (ICIP)*, 2006.
- [117] Y. Li, L. Sharan and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, p. 836-844, 2005.
- [118] Q. Shan, J. Jia and M. Brown, "Globally optimized linear windowed tone mapping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 4, pp. 663-675, 2010.
- [119] O. Pirinen, A. Foi and A. Gotchev, "Color high dynamic range

- imaging: The luminance–chrominance approach," *International Journal of Imaging Systems and Technology - Special Issue on Applied Color Image Processing*, vol. 17, no. 3, pp. 152-162, 2007.
- [120] Z. Wang, A. C. Bovik, H. Sheikh and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [121] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Prentice Hall, 2002.
- [122] A. Glantz, A. Krutz, M. Haller and T. Sikora, "Video coding using global motion temporal filtering," in *IEEE International Conference on Image Processing (ICIP)*, 2009.
- [123] K. J. Oh, S. Yea, A. Vetro and Y. S. Ho, "Depth reconstruction filter for depth coding," *IET Electronics Letter*, vol. 45, no. 6, pp. 305-306, 2009.
- [124] F. J. Richards, "A flexible growth function for empirical use," *Journal of Experimental Botany*, vol. 10, no. 2, pp. 290-301, 1959.
- [125] Z. G. Li, W. Yao, S. L. Xie and S. Rahardja, "Robust image similarity indices via intensity mapping functions and image partition," in *APSIPA ASC*, Singapore, 2010.
- [126] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, pp. 203:1-203:11, 2012.
- [127] R. Mantiuk, K. J. Kim, A. G. Rempel and W. Heidrich, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 14, pp. 40:1-40:14, 2011.
- [128] T. O. Aydin, R. Mantiuk, K. Myszkowski and H. P. Seidel, "Dynamic range independent image quality assessment," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 69:1-69:10, 2008.

Paper P1

Z. J. Zhu, Z. G. Li, S. Rahardja, and P. Fränti, "Recover real world scene: high quality image inpainting using multi-exposed references", *Electronics Letters*, pp. 1310-1312, Vol. 45, Issue 25, Dec 2009.

RECOVER REAL WORLD SCENE: HIGH QUALITY IMAGE INPAINTING USING MULTI-EXPOSED REFERENCES

Zijian Zhu, Zhengguo Li, Susanto Rahardja

Pasi Fränti

Institute for Infocomm Research

University of Joensuu

ABSTRACT

A novel method of high quality image inpainting for recovering original scene of degraded image using reference images of different exposures is proposed. It consists of a new inter-pixel relationship function and the respective refinement to synthesize missing pixel from existing spatially co-related pixels, and a dual patching to minimize the noise caused by dynamic range lost. Experiment on the method was conducted and the results demonstrate the reliability of the proposed method.

1. INTRODUCTION

Images of same scene can be captured with different exposures and combined with computing power to synthesize image that overcomes limitation of conventional cameras. However, useful data can be lost due to camera shake, especially when capturing by a hand held device, which generates noticeable artifact in the synthesized image. In other words, different from the traditional image inpainting [1] [2] and image completion [3], which generate only photorealistic patches, the degraded image inpainting in digital photography requires true luminance value of real world scene. Therefore, the challenge of patching is to find useful relations between missing pixels and the remaining pixels. The Camera Response Function (CRF) based fixing method [4] [5] uses only inter-image relationship. Unfortunately, the patched pixel is just a luminance shift from the reference pixel, and cannot represent the pixel value at the correct exposure of the degraded image. Motivated by these observations, we propose a new method using refined Inter-pixel Relationship Function (IRF) with both inter-image and intra-image correlations to recover the real scene luminance reliably, and a dual patching to reduce the dynamic range lost that further enhance the inpainting accuracy.

2. INTER-PIXEL RELATIONSHIP FUNCTION

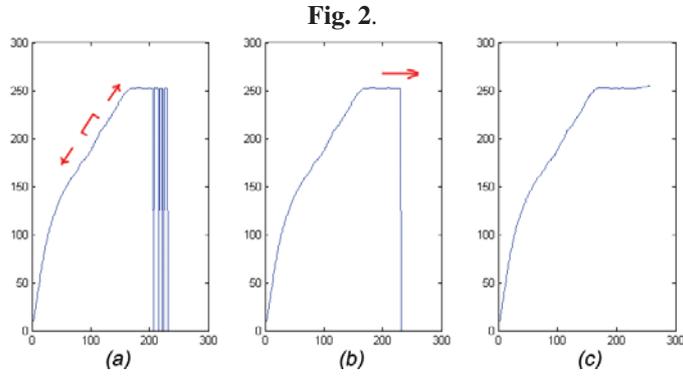
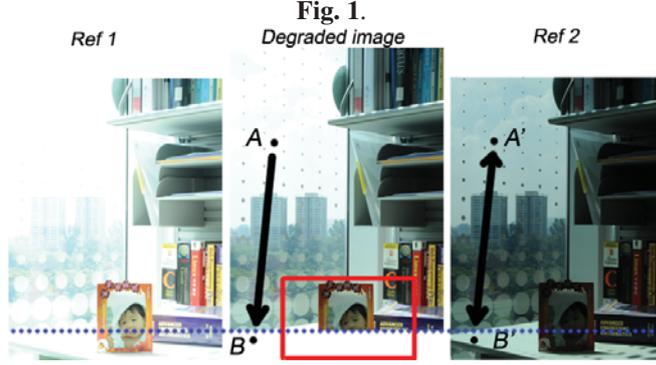
As shown in Fig. 1, the pixel A' and B' have the same intensity in the reference image. According to photography reciprocity, when the exposure time changes, the pixel values of A' and B' change correspondingly. Intuitively, the missing value of B can be copied from A in the degraded image. However, during the image capturing process, sensor noise, sampling noise and compression noise are commonly generated. Thus, it is more accurate to find all the pixels with the same intensity as B' in the reference image (\hat{Z}) and calculate their co-location values in the degraded image (Z) using mean average. We define the IRF as

$$\psi_c(\hat{Z}_c(x, y)) = \frac{\sum_{(\hat{x}, \hat{y}) \in \Omega(\hat{Z}_c(x, y))} Z_c(\hat{x}, \hat{y})}{|\Omega(\hat{Z}_c(x, y))|}, \quad (1)$$

where c is the color channel, and $|\Omega(\hat{Z}_c(x, y))|$ is the cardinality of the spatially co-related pixels set $\Omega(\hat{Z}_c(x, y))$, which is defined as

$$\Omega(\hat{Z}_c(x, y)) = \left\{ (\hat{x}, \hat{y}) \mid \hat{Z}_c(\hat{x}, \hat{y}) = \hat{Z}_c(\vec{x}, \vec{y}) \right\}. \quad (2)$$

The IRF has three useful characteristics inherited from the physical camera response. *Char1*: The IRF is a monotonically increasing function. *Char2*: The pixels located at left end (dark pixels) and right end (bright pixels) are highly compressed due to dynamic range limit. *Char3*: When choosing different reference images, shorter exposure time leads to smaller slope at the left end and bigger slope at the right end.



An "empty value" problem is raised in the raw IRF when the reference image does not span the whole dynamic range as shown in Fig. 2. Using *Char1*, the refinement has two steps. First, a median filter is adopted starting from the middle of the valid values towards left and right separately. The median filter corrects the monotonic errors, and recovers the empty values between the valid values. The second step extends two ends of the curve by using the neighborhood slope. The refined IRF is defined as

$$\Psi_c(z) = \text{Extend}(\text{Median}(\psi_c(z))), z \in [0, 255] \quad (3)$$

3. DUAL PATCHING

In order to increase the accuracy, the reference image is selected to have the smallest exposure difference with the degraded image. However, the dynamic range lost caused by *Char2* is still inevitable.

If the reference image has shorter exposure time than the degraded image, then as can be seen from *Char3*, the dark pixels in the reference image are mapped from a big dynamic range to a small one. In other words, it is a compressing process mapping multiple values to one, which in turn makes the IRF in this area reliable. On the contrary, a highly compressed bright pixel in the reference image is mapped into multiple values in the degraded image, which causes inaccuracy due to the dynamic range lost. Thus, multiple reference images, with longer and shorter exposure time respectively, can be adopted to recover the lost dynamic range and enhance the patching accuracy. The missing pixel intensity is calculated by

$$P_c = \frac{\dot{\Psi}_c(\dot{Z}) \cdot W(\max(\dot{Z}_R, \dot{Z}_G, \dot{Z}_B)) + \ddot{\Psi}_c(\ddot{Z}) \cdot W(\max(\ddot{Z}_R, \ddot{Z}_G, \ddot{Z}_B))}{W(\max(\dot{Z}_R, \dot{Z}_G, \dot{Z}_B)) + W(\max(\ddot{Z}_R, \ddot{Z}_G, \ddot{Z}_B))} \quad (4)$$

where \dot{Z} and \ddot{Z} are the intensities of two reference images at the same co-location, c is the color channel ($c = R, G, B$) and W is the weighting function defined as

$$W(z) = \begin{cases} \log(z + 1), & EV(Ref) > EV(Degrade) \\ \log(256 - z), & EV(Ref) < EV(Degrade) \end{cases}, z \in [0, 255] \quad (5)$$

4. RESULT

As shown in Fig. 3, the degraded area destroys the integrity of the original image composition. Clearly, the exemplar-based inpainting [2] algorithm works well on simple texture, such as the table top. However, obvious errors can be seen at complex contents, like the baby's face and the title of the book, due to short of reference. The CRF method [4] recovers the contents by the luminance shift from the reference pixels, where obvious artifact can be seen at the border. Our algorithm restores the original scene effectively. We tabulate the Peak Signal-to-Noise Ratio (PSNR) of each method in Table 1.

Fig. 3.



Table 1.

Method	Exemplar-based inpainting [2]	CRF method [4]	Our method
PSNR (dB)	11.75	20.54	33.66

In addition, we test our algorithm in the High Dynamic Range (HDR) image synthesis [6]. The border artifact generated by the degraded image in Fig. 3(e) is completely removed after patching using our algorithm in Fig. 3(f).

5. CONCLUSION

This Letter describes a new image inpainting method to patch the degraded image in an exposure set. As it uses all the relations of the valid pixels with refined IRF and reconstructs the missing area by dual patching, it demonstrates better quality than other algorithms. Experimental results with the HDR image synthesis further verify the efficiency of the proposed method.

6. AUTHORS' AFFILIATIONS

Z.J. Zhu, Z.G. Li, S. Rahardja (Department of Signal Processing, Institute For Infocomm Research, A*STAR, 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632), and P. Fränti (Department of Computer Science, University of Joensuu, P.O.Box 111, Joensuu, Finland 80101)

Email: zhuzj@i2r.a-star.edu.sg, ezgli@i2r.a-star.edu.sg, rsusanto@i2r.a-star.edu.sg, and franti@cs.joensuu.fi

7. REFERENCES

- [1] Hung J. C., Hwang C. H., Liao Y. C., Tang N. C., and Chen T. J.: 'Exemplar-based Image Inpainting Base on Structure Construction', *Journal of Software* 2008, 3(8), pp. 57-64.
- [2] Criminisi A., Pérez E., and Toyama K.: 'Region Filling and Object Removal by Exemplar-Based Image Inpainting', *IEEE Transaction on Image Processing* 2004, 13(9), pp. 1200-1212.
- [3] James H., and Alexei A. E.: 'Scene Completion Using Millions of Photographs', *SIGGRAPH Proceedings* 2007, Vol. 26, No. 3.
- [4] Grosch T.: 'Fast and Robust High Dynamic Range Image Generation with Camera and Object Movement', *Proceedings of Vision Modeling and Visualization* 2006, pp. 277-284.
- [5] Mann S.: 'Comparametric equations with practical applications in quantigraphic image processing', *IEEE Transaction on Image Processing*, Aug. 2000, Vol. 9, No. 8, pp. 1389-1406.
- [6] Debevec P., and Malik J.: 'Recovering High Dynamic Range Radiance Maps From Photographs', *SIGGRAPH Proceedings* 1997, pp. 130-135.

Paper P2

Z. G. Li, Z. J. Zhu, and S. Rahardja, “Structural similarity indices for high dynamic range imaging”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1145-1148, Prague, Czech Republic, May 2011.

STRUCTURAL SIMILARITY INDICES FOR HIGH DYNAMIC RANGE IMAGING

Zhengguo Li, Zijian Zhu and Susanto Rahardja

Signal Processing Department, Institute for Infocomm Research, 1 Fusionopolis Way, Singapore

ABSTRACT

In this paper, a structural similarity index is first proposed for two images with possibly different dynamic ranges and intensities as well as possibly small rotation and translation. The proposed index is then extended by dividing two images into local windows, and the similarity is detected by checking all pairs of local windows. It is shown by experimental results that the proposed indices are more robust to large intensity and dynamic range changes of two images from the same scene than the structural similarity (SSIM) index in [4].

1. INTRODUCTION.

Many real scenes possess significant higher dynamic ranges than the dynamic range that can be captured by digital cameras. In those scenes, a single shot low dynamic range (LDR) image usually turns out to be underexposed and/or overexposed in certain regions of the image. Because of this reason, a single shot does not have full dynamic range and one way to overcome this is to capture a set of differently exposed LDR images [1, 2, 3]. Such a technology is called “high dynamic range (HDR) imaging” which is becoming more and more popular. In fact, there are already many commercial HDR software, such as, Photomatix, Photoshop CS5, and so on. Several companies, such as, Pentax, Canon, Sony and Apple, have provided cameras/handphones to support HDR imaging. It can be expected that there are more and more LDR images with different exposures. As such, the quality metrics of differently exposed images will become an important R&D topic in the field of image processing.

The quality metrics have been well studied for both image and video processing. Many intensity-based indices were proposed to assess the similarity of a pair of images. Their similarity is determined by comparing the corresponding pixel intensities [4, 5]. Due to their simplicity, they are adopted in different applications. For example, the structural similarity (SSIM) index [4] was recently applied to video coding [6]. Although these indices are designed to focus on cases where the dynamic ranges and intensities of the two images are almost the same, in practical situation where processing of high dynamic range imaging requires different level of exposures, such indices would not work optimum [3]. In fact, it is known that the dynamic range can also be changed due to images or videos with different bit-depth [7]. Moreover, the images being compared are assumed to be well registered by the intensity-based indices. In addition, it is unavoidable that there are rotations and translations in a set of differently exposed LDR images, especially when they are captured for outdoor scenes. It is thus desirable to provide simple similarity indices that are robust to translation, rotation, intensity change and dynamic range between two input images.

In this paper, we first propose a structural similarity index for two LDR images by using intensity mapping functions (IMFs) and the Richards’ curve [11]. The IMFs between two images are identified by using their accumulated histograms [9]. It can be shown that when there exists either intensity or dynamic range change between two images, the IMFs can be used to unify the dynamic ranges or intensities of two images from the same scene [10]. On the other hand, when there is neither intensity nor dynamic range change between two images from the same scene, the accumulated histograms of two similar images are almost the same, and the structure of the corresponding image is kept even though it is mapped by the IMFs. Due to the function of IMFs, the proposed index is robust to the dynamic range and intensity changes between two images. The Richards’ curve [11] is adopted to design a switch on the deduction of the mean values from these two images. When two images are from the same scene, the similarity index value is calculated without the deduction of mean values from the corresponding images. As a result, the proposed index is robust to small translation and rotation between two images. When two images are from different scenes, the mean value is deducted from the corresponding images before the similarity index value is computed. As such, the proposed index is sensitive to two “similar” images from different scenes. The proposed index is then extended by dividing the whole image into local windows as in [4]. The similarity of two images is detected by checking all pairs of local windows. Such a local similarity index is insensitive to large intensity changes, dynamic range changes, translation and small rotations between images from the same scene.

The rest of this paper is organized as follow. The proposed similarity indices are presented in Section 2. Experimental results are provided in Section 3 to illustrate the efficiency of the proposed indices. Finally, concluding remarks are listed in Section 4.

2. IMF BASED STRUCTURAL SIMILARITY INDICES

2.1. Mapping of Two Input Images

Assume that two images being compared are denoted as \mathbf{Z}_1 and \mathbf{Z}_2 , respectively. $\Lambda_{1,2}(z)$ and $\Lambda_{2,1}(z)$ are the intensity mapping functions (IMFs) from \mathbf{Z}_1 to \mathbf{Z}_2 and vice versa, respectively. \mathbf{Z}_1 and \mathbf{Z}_2 are first mapped by using the IMFs $\Lambda_{1,2}(z)$ and $\Lambda_{2,1}(z)$ as

$$\tilde{Z}_1(p) = \begin{cases} \Lambda_{1,2}(Z_1(p)); & \text{if } Z_1(p) \text{ is more reliable} \\ Z_1(p); & \text{otherwise} \end{cases}, \quad (1)$$

$$\tilde{Z}_2(p) = \begin{cases} \Lambda_{2,1}(Z_2(p)); & \text{if } Z_2(p) \text{ is more reliable} \\ Z_2(p); & \text{otherwise} \end{cases}, \quad (2)$$

where $p=(x,y)$ represents a pixel point. The relative reliability is defined according to two typical scenarios as follows:

- 1) Both \mathbf{Z}_1 and \mathbf{Z}_2 are two LDR images that are captured from

the same scene with different exposure times. Pixel $Z_1(p)$ is more reliable if $w(Z_1(p)) \geq w(Z_2(p))$ where the weighting function $w(z)$ is a triangular function as [3]:

$$w(z) = \begin{cases} (z+1)/128; & \text{if } z \leq 127 \\ (256-z)/128; & \text{otherwise} \end{cases}.$$

2) Both \mathbf{Z}_1 and \mathbf{Z}_2 are LDR images that are captured from the same scene at different bit depths [7]. $Z_1(p)$ is more reliable if the bit depth of image \mathbf{Z}_1 is higher than that of image \mathbf{Z}_2 .

The main objective of IMFs is to unify the intensities and dynamic ranges of \mathbf{Z}_1 and \mathbf{Z}_2 which was verified in [10]. $\tilde{\mathbf{Z}}_1$ and $\tilde{\mathbf{Z}}_2$ are further mapped to $\check{\mathbf{Z}}_1$ and $\check{\mathbf{Z}}_2$ as

$$\check{Z}_1(p) = \tilde{Z}_1(p) - (1 - \Gamma(\beta(\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2))) * \mu_{\tilde{\mathbf{Z}}_1}, \quad (3)$$

$$\check{Z}_2(p) = \tilde{Z}_2(p) - (1 - \Gamma(\beta(\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2))) * \mu_{\tilde{\mathbf{Z}}_2}, \quad (4)$$

where $\mu_{\tilde{\mathbf{Z}}_1}$ and $\mu_{\tilde{\mathbf{Z}}_2}$ are the mean values of $\tilde{\mathbf{Z}}_1$ and $\tilde{\mathbf{Z}}_2$, respectively. $\Gamma(z)$ is the Richards' curve [11]:

$$\Gamma(z) = \frac{1}{1 + \exp(\zeta_1(\zeta_2 - z))}, \quad (5)$$

ζ_1 and ζ_2 are the growth rate and the threshold of maximum growth. $\beta(\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2)$ is the cosine of the angle between the two lexicographic order vectors of $\tilde{\mathbf{Z}}_1$ and $\tilde{\mathbf{Z}}_2$. Both Γ and β are used to detect the similarity of two images. On one hand, when two images are from the same scene, the existence of intensity changes, small translation and rotation between them would appear in the value of Γ wherein the value will be almost equal to 1. On the other hand, when two image are from different scenes, the value of Γ approaches 0.

2.2. The Proposed Similarity Indices

Consider a pair of images in the spatial domain, $\mathbf{Z}_i = \{z_i(p) | 1 \leq x \leq N_1, 1 \leq y \leq N_2\}$ ($i = 1, 2$). An intensity based similarity index is defined as

$$S_s(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{2\mu_{\tilde{\mathbf{Z}}_1}\mu_{\tilde{\mathbf{Z}}_2} + \epsilon_1 \sum_p 2\check{Z}_1(p)\check{Z}_2(p) + \epsilon_2}{\mu_{\tilde{\mathbf{Z}}_1}^2 + \mu_{\tilde{\mathbf{Z}}_2}^2 + \epsilon_1 \sum_p [\check{Z}_1^2(p) + \check{Z}_2^2(p)] + \epsilon_2}, \quad (6)$$

where ϵ_1 and ϵ_2 are two small positive constants and they are adopted to improve the robustness of the proposed index when the local signal to noise ratios are very low.

Besides the robustness with respect to intensity changes, translation and rotation, an image similarity index should also be sensitive to two images that are captured from two different scenes. In other words, it is also important not to classify two different images into the same group. The global structure of two images from different scenes could be very similar. Although the global structure of two images from different scenes could be very similar, but it can be observed that the local structures of two images from different scene are different. Based on this observation, the proposed index $S_s(\mathbf{Z}_1, \mathbf{Z}_2)$ is applied locally rather than globally for image quality assessment which is similar to the SSIM index [4]. For simplicity, we assume that the number of local windows is P and the contents of \mathbf{Z}_1 and \mathbf{Z}_2 at the i th window are denoted as $\mathbf{Z}_{1,i}$ and $\mathbf{Z}_{2,i}$, respectively. The value of $S_s(\mathbf{Z}_{1,i}, \mathbf{Z}_{2,i})$ are computed for all pixels in the i th local window by using Equation (6).

The corresponding local similarity index, $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$, is defined as the mean of all $S_s(\mathbf{Z}_{1,i}, \mathbf{Z}_{2,i})$'s, i.e.,

$$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2) = \frac{1}{P} \sum_{i=1}^P S_s(\mathbf{Z}_{1,i}, \mathbf{Z}_{2,i}). \quad (7)$$

3. EXPERIMENTAL RESULTS

In this section, we shall first show that the IMFs do not change the structure of the input images. The value of $\Gamma(\beta(\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2))$ in Equations (3) and (4) is fixed as 0. The tested images are from the VQEG [12] and they contain different quantization noises as shown in Fig. 1. Five pairs of image are tested with one image is fixed as the first image. It is shown from Table 1 that the values of SSIM and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ are almost the same.



Fig. 1. An image sequence from the VQEG

Table 1. Comparison of SSIM and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ for the image sequence in Fig. 1

Pair	1	2	3	4	5
SSIM	0.9478	0.9083	0.8771	0.8582	0.8103
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9477	0.9085	0.8778	0.8584	0.8125

We then compare the SSIM [4], $S_s(\mathbf{Z}_1, \mathbf{Z}_2)$ and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ in the case that the exposures of two input images are different [3]. Suppose that $L_{\mathbf{Z}_i}$ and $U_{\mathbf{Z}_i}$ are the lower and upper bounds of image \mathbf{Z}_i , respectively. Define a constant ϖ as $\min_{i=1,2}\{U_{\mathbf{Z}_i} - L_{\mathbf{Z}_i}\}$. The values of ϵ_1 , ϵ_2 , ζ_1 , and ζ_2 are empirically determined as $(0.01 * \varpi)^2 N_1 N_2 / P$, $(0.03 * \varpi)^2 N_1 N_2 / P$, 1024, and 0.975 respectively. Two static scenes are chosen, namely Fusionopolis and Memorial. The Fusionopolis contains both the indoor and outdoor scenes of a building but there are no moving objects in this sequence as shown in Fig. 2. The other, "Memorial", includes the indoor scene of a church [3]. The experimental results are shown in Tables 2 and 3, respectively. We now study two scenes with moving objects. The "Pantry" is captured in a pantry and the human subject moves his head in the sequence as demonstrated in Fig. 3. The other, "Street", is captured outdoor and there are several moving cars and people in the sequence as illustrated in Fig. 4. The experimental results for these two image sequences

are illustrated in Tables 4 and 5, respectively. It is shown in Tables 2-5 that the gaps between 1 and $S_s(\mathbf{Z}_1, \mathbf{Z}_2)$ and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ are usually smaller than the gap between 1 and the SSIM [4]. Thus, $S_s(\mathbf{Z}_1, \mathbf{Z}_2)$ and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ are usually more robust when two images are captured from the same scene, especially when there are no moving objects in the scene. We also test the combined perturbation of intensity, translation and rotation by testing “Pantry” with camera movement as shown in Fig. 5. The rotation and translation values are also unknown. The experimental results are demonstrated in Table 6. Obviously, $S_s(\mathbf{Z}_1, \mathbf{Z}_2)$ and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ are more robust than the SSIM [4] with respect to rotation and translation.

Table 2. “Memorial” with nonlinear intensity change

$\frac{\Delta t_1}{\Delta t_2}$	2	4	8	16	32
SSIM	0.863	0.6979	0.512	0.3756	0.3112
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9983	0.9985	0.9986	0.999	0.9992
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9973	0.9975	0.997	0.9972	0.9971



Fig. 2. Sequence of “Fusionopolis” with different exposures

Table 3. “Fusionopolis” with nonlinear intensity change

$\frac{\Delta t_1}{\Delta t_2}$	2	4	8	16	32
SSIM	0.8638	0.5502	0.3229	0.1985	0.1067
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9999	0.9999	0.9999	0.9999	0.9999
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9994	0.9972	0.9917	0.9845	0.9855

Finally, we compare these three indices in the case that the exposures of two input images are same. We consider the case that there are rotation and translation between two input images. We test five pairs of images with scenes that are depicting typical daily life which is shown in Fig. 6. The rotation and translation values are unknown. The rotation is very small while the translation could be large. It is demonstrated in Table 7 that the proposed indices are applicable to a pair of conventional images with the same exposure. We also study five pairs of images that are captured from different scenes but look somewhat similar as in Fig. 7. The experimental results are shown in Table 8. Both the SSIM and $S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$ are sensitive to images from different scenes.



Fig. 3. Sequence of “Pantry” with different exposures

Table 4. “Pantry” with nonlinear intensity change

$\frac{\Delta t_1}{\Delta t_2}$	1.6	1.6 ²	1.6 ³	1.6 ⁴	1.6 ⁵
SSIM	0.8737	0.7605	0.5348	0.4031	0.4389
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9983	0.9968	0.9945	0.9948	0.9931
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.994	0.9777	0.9689	0.9571	0.9617

4. CONCLUSION

In this paper, an intensity based similarity index which uses intensity mapping functions (IMFs) between two images is proposed. The concept is extended by dividing the images into local windows and checking the local windows. The proposed indices are robust to intensity and dynamic range changes as well as translation and rotation of two images from the same scene. They are also sensitive to two “similar” images from different scenes.

References

- [1] C. W. Wyckoff, “An experimental extended response film,” *SPIE Newslett.*, pp.16-20, June-July 1962.
- [2] S. Mann, “Comparametric equations with practical applications in quantigraphic image processing,” *IEEE Trans. on Image Processing*, vol. 9, no. 8, pp. 1389-1406. Aug. 2000.
- [3] P. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photograph,” In *Proceedings SIGGRAPH 1997*, pp.369-378, 1997.
- [4] Z. Wang, A. C. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp.600-612, Apr. 2004.
- [5] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Prentice Hall, 2002.
- [6] A. Glantz, A. Krutz, M. Haller, and T. Sikora, “Video coding using global motion temporal filtering,” In *2009 IEEE International Conference on Image Processing*, pp. 1053-1056, Egypt, Nov. 2009.
- [7] K.-J. Oh, S. Yea, A. Vetro, and Y.-S. Ho, “Depth reconstruction filter for depth coding,” *IET Electronics Letter*, vol. 45, no.6, pp. 305-306, Mar. 2009.

- [8] S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using Delaunay triangulation," *IEEE Trans. on Image Processing*, vol. 11, no. 12, pp. 1427-1441, Dec. 2002.
- [9] M. Grossberg and S. Nayar, "Determining the camera response from images: what is knowable?" *IEEE Trans. on pattern analysis and machine intelligence*, pp. 1455-1467, Vol. 25, No. 11, Nov. 2003.
- [10] Z. G. Li, W. Yao, S. L. Xie, and S. Rahardja, "Robust image similarity indices via intensity mapping functions and image partition", In *APSIPA ASC 2010*, 14-17 Dec. 2010, Singapore.
- [11] F. J. Richards, "A flexible growth function for empirical use," *Journal of Experimental Botany*, vol. 10, no. 2, pp. 290-301, Feb. 1959.
- [12] <http://www.its.bldrdoc.gov/vqeg/>



Fig. 4. Sequence of "Street" with different exposures

Table 5. "Street" with nonlinear intensity change

$\frac{\Delta t_1}{\Delta t_2}$	1.6	1.6^2	1.6^3	1.6^4	1.6^5
SSIM	0.6685	0.4216	0.4193	0.2589	0.2799
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9828	0.978	0.9399	0.9555	0.9833
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9681	0.9504	0.9331	0.9085	0.9499



Fig. 5. Sequence of "Pantry" with camera movement and different exposures

Table 6. "Pantry" with both intensity change and camera movement

$\frac{\Delta t_1}{\Delta t_2}$	1.6	1.6^2	1.6^3	1.6^4	1.6^5
SSIM	0.5064	0.475	0.4791	0.3803	0.3048
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9787	0.9357	0.9879	0.9674	0.9845
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.922	0.8547	0.9414	0.8875	0.8706



Fig. 6. Five pairs of images from the same scenes

Table 7. Five pairs of images with camera movement

Pair	1	2	3	4	5
SSIM	0.8584	0.5688	0.6454	0.7545	0.5069
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9757	0.9199	0.9634	0.979	0.9864
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.9622	0.9153	0.8932	0.8695	0.939



Fig. 7. Five pairs of images from different scenes

Table 8. Five pairs of images from different scenes

Pair	1	2	3	4	5
SSIM	0.4855	0.473	0.4725	0.7	0.4535
$S_s(\mathbf{Z}_1, \mathbf{Z}_2)$	0.7327	0.6466	0.7193	0.5059	0.3576
$S_s^b(\mathbf{Z}_1, \mathbf{Z}_2)$	0.4627	0.4802	0.503	0.6309	0.3672

Paper P3

Z. J. Zhu, S. Rahardja, Z. G. Li, P. Fränti, "Detecting and composing near-identical HDR images without exposure information", *IEEE International Conference on Image Processing*, pp. 3337-3340, Hong Kong, Sep 2010.

DETECTING AND COMPOSING NEAR-IDENTICAL HDR IMAGES WITHOUT EXPOSURE INFORMATION

Zijian Zhu, Susanto Rahardja, Zhengguo Li

Signal Processing Department
Institute for Infocomm Research
1 Fusionopolis Way, #21-01, Singapore 138632

Pasi Fränti

School of Computing
University of Eastern Finland
P.O.Box 111, Joensuu, Finland 80101

ABSTRACT

In high dynamic range (HDR) imaging, one essential problem is to verify whether two images are from the same scene. We name such a pair of HDR images as near-identical images. Even though the images carrying the same information, their similarity cannot be detected by pixel-wise comparisons. To solve this problem, we propose a new radiance alignment by using their log histogram similarity. We also develop a new algorithm to compose near-identical HDR image from unprocessed images without any prior information about their exposures. According to experiments, our detection method is able to identify near-identical HDR images effectively, and our synthesis algorithm is able to recover the correct exposure ratios and compose near-identical images.

Index Terms— High dynamic range, near-identical, image synthesis, image comparison

1. INTRODUCTION

In order to overcome dynamic range limitation of typical *low dynamic range* (LDR) images, a radiance map is defined to represent the whole range of visible light intensities and stored as a *high dynamic range* (HDR) image. Nowadays, the radiance map is commonly recovered from a set of differently exposed LDR images [1][4][5][6]. Due to the differences in synthesis algorithms and input parameters, the same set of LDR images can generate different HDR images carrying the same scene information with unequal radiance presentation. We classify them as near-identical. On the other hand, some HDR images are generated by estimated exposure times, which are not accurate. Although they look identical with human eyes, many pixels carry inaccurate values. Thus, it is important to detect whether two images are near-identical, and also to recover the missing exposures to prevent the generating of inaccurate HDR images.

Similar HDR images can be produced in many ways. For example, if we capture two sets of LDR images at the same scene with the same camera settings except a change in film speed (ISO), two identical image sets (when ISO noise is small) with different exposures will be generated. The HDR

images composed by using these two input sets carry the same scene information. However, they are near-identical images with shift in the radiance domain. In another example, if the original exposure ratio of the input LDR images has been changed proportionally, such as changing the exposure time from [1/2s, 1/4s, 1/8s] to [1/2s, 1/8s, 1/24s], the two synthesized HDR image are near-identical with scale-invariant. This situation is worse when exact exposure information is not available. If the exposure values are given randomly, the *camera response function* (CRF) [1][3][4][5] is not estimated exactly, and the composed HDR images do not carry the same scene information. However, since the differences are usually gradual color changes at particular intensity values, it can hardly be detected by human eyes. As we can see, the near-identical images and not identical images can not be differentiated by human eyes. More difficultly, if we compare those images pixel-by-pixel as LDR image comparison usually does, none of them can be recognized as being the same.

On the other side, some methods have been introduced to recover the missing exposures. A radiometric model [6] was introduced to estimate the exposure ratio with an initial value set by user and an iterative process to fine-tune it. The results vary due to different initial settings and the order of the polynomial model used. Another rough approximation was obtained by using *intensity mapping function* (IMF) [2] between two images. However, as realized by the authors themselves, there are a lot of restrictions and assumptions involved. Thus, they used the approximated exposure ratio as the initial value to the radiometric model [6] in their experiments.

In this paper, we propose two new methods to detect and compose the near-identical HDR images without exposure information respectively. Our detection method verifies the similarity of two HDR images based on a radiance alignment. Instead of finding individual exposure ratio, we present an accurate algorithm to recover the rate in between the different exposure ratios by using *inter-pixel relationship function* (IRF) [7] in the image composition. We verify that the synthesized images carry the same scene information in the experiments.

2. NEAR-IDENTICAL HDR IMAGE

2.1. Definition of Near-Identical HDR Image

If two HDR images are exactly identical, their *peak signal-to-noise ratio* (PSNR) calculated from pixel-by-pixel comparison shall be infinite, or at least a very large number when tolerating small noises due to compression or sampling. However, it is not true for two near-identical images, as they store the same information at different locations. The fundamental factor of such mismatch is due to lack of strict link between physical light radiance and image radiance map. For intuitive understanding, we plot histograms of three near-identical HDR images in log radiance domain, as shown in Fig. 1.

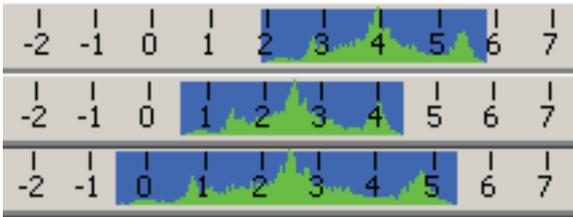


Fig. 1. Histogram in log radiance domain of (top) original HDR image, (mid) shifted HDR image, and (bottom) scale-invariant HDR image.

We define near-identical HDR images as images that carry the same scene information, but are shifted or have different scale in the log radiance domain. In other words, we classify them as

$$\Pi = \{HDRI_q : R_q | \log(R_q) = a_q \cdot \log(R_0) + b_q\}, \quad (1)$$

where R is the radiance of the HDR image, 0 represents the reference image defined as original, q represents the test image, a_q is the scale factor, and b_q is the shift factor. If the HDR image is composed from an LDR image set and the exposure info is known, the scale factor can be estimated as $a_q = \log(k_{q,1,2}) / \log(k_{0,1,2})$, where $k_{1,2}$ represents the exposure ratio between the first two images in the LDR image set. If the scale factor equals to 1, we call this HDR image *shift-identical* to the reference HDR image, and the shift factor is calculated as $b_q = \log(\Delta t_{0,1}) - (\Delta t_{q,1})$, where Δt_1 denotes the exposure time of the first LDR image. If the scale factor does not equal to 1, we name it *scale-identical*, where the value of the shift factor is related to the scale factor and the image radiance.

Near-identical HDR images are usually synthesized from the same set of LDR images using different algorithms or different parameters. For example, when we use the CRF based method [1][4] to generate HDR images, an assumption of CRF (g) is usually given as $g(Z_{mid}) = 0$, which does not have any physical meaning. In practice, Z_{mid} is equal to 128. However, we can replace it to any other pixel values, such

as 60, 100 or 200. Such changes lead to a shift in CRF and eventually cause the shift in radiance.

Not only different algorithms can generate near-identical HDR images, but more commonly, the change of input exposure time leads to the same consequence. For instance, the exposure times of a set of three LDR images are [1/125s, 1/250s, 1/500s]. If we replace them to [1/2s, 1/4s, 1/8s], where the ratios between the images are the same as the original values, the synthesized HDR image is shift-identical to the original HDR image. If we replace them to [1/2s, 1/8s, 1/32s], where the ratios between the images are changed proportionally, the synthesized image is scale-identical.

In all the above cases, the HDR images recovered from the same LDR image set carry the same scene information, and can be displayed identically on LDR monitor with some parameters change in viewing software. Furthermore, the tone mapping result of the shift-identical HDR image is usually the same as the result from the original image for most tone mapping algorithms. Even the scale-identical HDR image can have the same tone mapping result if the range info is not used in the tone mapping algorithm.

2.2. Detecting Two Near-Identical HDR Images

The histogram of the log radiance is calculated from the minimum log radiance (m) to the maximum log radiance (M) in both images, with m and M being defined as

$$\begin{cases} m = \min_p (\log(R_{q,p}), \log(R_{0,p})) \\ M = \max_p (\log(R_{q,p}), \log(R_{0,p})) \end{cases}, \quad p = 1, 2, \dots, P, \quad (2)$$

where R_q denote the radiance value of HDR image q , and P denotes the total number of pixels in all color channels. The total number of bins (H) in plotting the histogram is an adjustable parameter, where we use $H = 1000$ in our experiments. The scale factor is then calculated by

$$a_q = \frac{\max(h|\psi_0(h) > \epsilon) - \min(h|\psi_0(h) > \epsilon)}{\max(h|\psi_q(h) > \epsilon) - \min(h|\psi_q(h) > \epsilon)}, \quad (3)$$

where ψ is the histogram function, and ϵ is a small number of pixels for noise reduction. We use $\epsilon = P/H/30$ in our experiments.

We choose one image as a reference image and the other as a test image. If the scale factor does not equal to 1, the test image will first be multiplied by the scale factor, and then its histogram is recalculated in the new scale. After that, we fill in the test histogram with an all zero vector (length of H) at both sides. Correlation coefficients are computed in between the test histogram and the reference histogram by sliding a window (of size H) one bin at a time on the padded test histogram for $2H$ comparisons. The biggest coefficient represents the best fit of the two HDR images. The shift factor is calculated as

$$b_q = \frac{\kappa(h) - H}{H} \cdot (M - m), \quad (4)$$

where $\kappa(h) = \{h | \sigma(h) = \max(\sigma(i)), i = 1, 2, \dots, 2H\}$, and σ is the correlation coefficient function corresponding to the starting bin number of the sliding window.

Fig. 2 gives an example of the radiance alignment, where the highest correlation is detected at bin 850 indicating the distance for shifting the test image is 150 units.

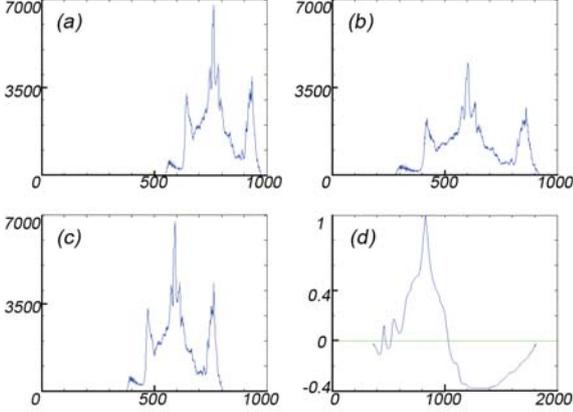


Fig. 2. Histogram of (a) reference image, (b) test image, and (c) test image after scaling, where X axis is bin number, and Y axis is number of pixels. (d) is the correlation results from scaled test image to the reference image with shift of one bin each time, where X axis is the starting bin number of the sliding window, and Y axis is the correlation coefficient.

The PSNR of the two HDR images is calculated as

$$PSNR = 10 \cdot \log_{10} \left(\frac{P \cdot \Phi^2}{\sum_{p=1}^P [\log(\hat{R}_{q,p}) - \log(R_{0,p})]^2} \right), \quad (5)$$

where \hat{R}_q is the aligned radiance map calculated by Equation (1), and Φ is a constant and it is chosen as $\Phi = \log(2^{127}) - \log(\frac{0.5}{256} \cdot 2^{-128})$. Two HDR images are said to be near-identical if their PSNR is greater than a predefined threshold. The threshold is set as 70dB in this paper.

3. SYNTHESIZE WITHOUT EXPOSURE INFO

As mentioned earlier, the exposure ratios between input images are essential in composing HDR images. Here, we recover the exposure ratio by using their EV relations. Since the EV interval is defined by $EV_i = \log_2(\Delta t_i / \Delta t_{i+1})$, it reflects the exposure ratio directly. EV unit is defined as the EV difference between two images proportional to the real EV interval. For any three images in an exposure set, we select one image as reference. The EV unit in between the second test

image and the reference image is calculated by

$$\delta_{r,t,\hat{t}} = \frac{\sum_{z=\hat{Z}_{r,\hat{t},c}}^{\hat{Z}_{r,\hat{t},c}} [g_{r,t,c}(z) - g_{r,t,c}(\varphi_{r,\hat{t},c}(z))]}{\hat{Z}_{r,\hat{t},c} - \hat{Z}_{r,\hat{t},c} + 1}, \quad (6)$$

where r and t represent the reference image and the first test image in creating the CRF [1][4] mapping g , \hat{t} is the second test image used in IRF [7] mapping φ , and valid region boundaries are defined as

$$\begin{cases} \hat{Z}_{r,\hat{t},c} = \max(\alpha, \varphi_{r,\hat{t},c}^{-1}(\alpha)) \\ \hat{Z}_{r,\hat{t},c} = \min(\beta, \varphi_{r,\hat{t},c}^{-1}(\beta)) \end{cases}. \quad (7)$$

where α and β are two constants, and they are chosen as 10 and 250 respectively. For simplicity, we only use R channel in our experiments for this calculation. When the first and second test images are the same, a base EV unit is estimated. The unknown exposure ratio is then calculated by

$$\tilde{k}_{i,i-1} = \left(\frac{\delta_{i-2,i-1,i}}{\delta_{i-2,i-1,i-1}} - 1 \right) \cdot \tilde{k}_{i-1,i-2}, \quad i = 3, \dots, n. \quad (8)$$

If there is no prior exposure knowledge, assumptions of the exposure time of the first LDR image and the exposure ratio between the first two LDR images are given to synthesize a near-identical HDR image using our algorithm. If the assumption of initial exposure time is correct, a shift-identical HDR image is created. If both assumptions are not accurate, a scale-identical HDR image is composed.

4. EXPERIMENTAL RESULTS

Six experiments (EXP1-6) are shown in Table 1, based on a set of five LDR images captured using bracket function with EV interval of 1. In the first two experiments, the HDR images are synthesized using our composing algorithm with different initial assumptions, as if we do not know the exposure info. We assume that the initial exposure time (Δt_1) equals to 0.5s and the initial EV interval (EV_1) equals to 1 in EXP1.

Table 1. Compose HDR image without exposure info by using (EXP1-2) initial assumptions and EV recovered using our algorithm (bold), and (EXP3-6) all random exposure values.

	Δt_1 (s)	EV_1	EV_2	EV_3	EV_4
Original	0.02	1	1	1	1
EXP1	0.50	1.00	0.98	0.99	0.96
EXP2	0.50	1.58	1.53	1.54	1.51
EXP3	0.50	0.32	0.93	0.49	1.58
EXP4	0.02	1.32	1.42	0.58	0.74
EXP5	0.50	0.15	0.17	3.22	3.60
EXP6	0.03	0.26	0.32	1	3.32

Since the assumption of the initial EV interval is correct, as we can see, the rest of the EV intervals computed from our algorithm are almost the same as the original camera setting. Thus, a shift-identical HDR image is synthesized. In EXP2, We change the assumption of the initial EV interval to 1.58, which is different from the camera setting. The computed EV intervals proportionally equal to the original values, and a scale-identical HDR image is generated.

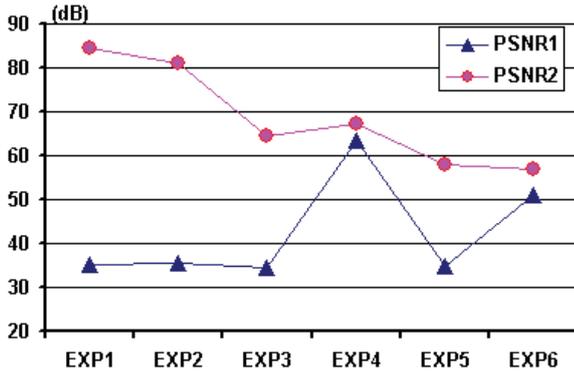


Fig. 3. Chart of PSNR from EXP1-6, where PSNR1 and PSNR2 are calculated using direct pixel-by-pixel comparison and our detection method respectively.

We then perform direct pixel-by-pixel comparison and our comparison using near-identical detection algorithm on the generated HDR images. As plotted in Fig. 3, the pixel-wise comparisons indicate that both test sets are different by having very small PSNR values of around 35dB, whereas our detection method verifies the similarity of the images by giving PSNR value greater than 80dB. We further test the robustness of our detection algorithm by using artificial random exposures in the next four experiments. When the initial exposure time is close to the original setting, as shown in EXP4 and EXP6, the direct comparison gives relative high PSNR. In contrary to this, our detection method detects bigger PSNR values in EXP3 and EXP4 compare to those in EXP5 and EXP6, which connects to the changes of the exposure ratios correctly.

Table 2. EV correction with known initial exposure ratio

LDR image	1	2	3	4	5
True bracketing EV interval	1	1	1	1	N.A.
Exposure time (s)	1	1/2	1/3	1/6	1/13
EV interval by exposure time	1	0.59	1	1.12	N.A.
EV interval by our algorithm	1 (given)	0.92	0.94	0.88	N.A.

Table 2 demonstrates that our ratio recovery algorithm can

efficiently fix the not accurate exposure time caused by human or camera errors. Five images are captured using bracket function with EV interval of 1, and the exposure time is automatically given by the camera. Obviously, the calculated EV difference is not the same as the actual settings, especially with the third LDR image, where the difference is higher than 40%. However, our algorithm is able to correct the EV intervals with the given initial ratio.

5. CONCLUSION

In this paper, we introduce a new concept of near-identical HDR images, which carry the same scene information with different radiance representations. The same set of near-identical HDR images can be transformed to each other with linear functions in log radiance domain, which is useful in displaying HDR images and HDR tone mapping. We also propose a method to detect whether two HDR images carry the same scene information and an algorithm to synthesize near-identical HDR images without exposure information. Experiments show the effectiveness of our algorithms.

6. REFERENCES

- [1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High dynamic range imaging: acquisition, display and image-based lighting*, Morgan Kaufmann, 2005.
- [2] M. Grossberg, and S. Nayar, “Determining the camera response from images: what is knowable?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 11, pp. 1455-1467, Nov. 2003.
- [3] M. Grossberg, and S. Nayar, “Modeling the space of camera response functions”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1272-1282, Vol. 26, No. 10, Oct. 2004.
- [4] P. Debevec, and J. Malik, “Recovering high dynamic range radiance maps from photographs”, *SIGGRAPH Proceedings*, pp. 130-135, 1997.
- [5] S. Mann, “Comparametric equations with practical applications in quantigraphic image processing”, *IEEE Transaction on Image Processing*, Vol. 9, No. 8, pp. 1389-1406, Aug. 2000.
- [6] T. Mitsunaga, and S. Nayar, “Radiometric self calibration”, *Computer Vision and Pattern Recognition*, pp. 380, Vol. 1, Jun. 1999.
- [7] Z.J. Zhu, Z.G. Li, S. Rahardja, and P. Franti, “Recover real world scene: high quality image inpainting using multi-exposed references”, *Electronics Letters*, pp. 1310-1312, Vol. 45, Issue 25, Dec. 2009.

Paper P4

Z. J. Zhu, S. Rahardja, S.Q. Wu, and P. Fränti, "Real-time ghost removal for composing high dynamic range images", *IEEE Conference on Industrial Electronics and Applications*, pp. 1627-1631, Taichung, Taiwan, Jun 2010.

Real-time Ghost Removal for Composing High Dynamic Range Images

Zijian Zhu, Shiqian Wu, Susanto Rahardja
Signal Processing Department
Institute for Infocomm Research
Singapore 138632
Email: {zhuzj, shiqian, rsusanto}@i2r.a-star.edu.sg

Pasi Fränti
School of Computing
University of Eastern Finland
P.O.Box 111, Joensuu, Finland 80101
Email: franti@cs.joensuu.fi

Abstract—It is known that high dynamic range image can be recovered by a set of conventional images shooting at the same scene but with different exposure levels. However, ghost artifacts are commonly seen due to the movement of the object between different captures. In this paper, we propose a real-time algorithm to detect and remove the ghost artifact based on four simple procedures: 1) Inter-pixel Relationship Function estimation for mapping differently exposed images into the same exposure level; 2) bi-directional comparison for further enhancing the detection accuracy, especially at the underexposed and saturated areas; 3) moving object classification with edge artifacts removal; 4) moving object correction for fixing the missing dynamic range. The new algorithm delivers robust ghost removal in $O(n)$ time. To further reduce the processing time, the ghost removal process can be executed in parallel with the photo capturing process.

I. INTRODUCTION

A *High Dynamic Range* (HDR) image stores a greater range of tonal detail than a conventional camera can capture in a single photo. The HDR image can be recovered from a set of differently exposed *Low Dynamic Range* (LDR) images [1][2][3][4], and good results have been reported in case of static scene. However, when an object, such as a walking person, changes its position in different LDR images, the ghost artifact is generated during the HDR image synthesis. Various methods [5][6][7][8] have addressed this problem, and the biggest challenge is how to detect the moving object since all the images are in different exposure levels, and cannot be compared directly [9].

Khan et al [5] and Pedoen and Heikkilä [6] tackle the problem by calculating the probability of a pixel belonging to the background or the moving object iteratively. Those methods are computationally expensive since they need multiple rounds of slow iterative processes. Faster solutions aim at detecting the moving object directly, and the key is to find the common attributes in connecting the images captured with different exposures. Grosch [7] links the differently exposed images using the *Camera Response Function* (CRF) and replaces the moving object with the CRF mapping values from the reference image. As a trade-off, it introduces noise due to inaccurate camera response estimation, mainly caused by the compression at underexposed and saturated pixels. Furthermore, although the moving object pixels are fixed by the CRF, they are only luminance shift from the reference image, and as such the dynamic range of the moving object is still lost. Entropy based

statistical process [8] uses local histogram around a small region of each pixel to distinguish the moving object from the background. However, as realized by the authors themselves, the algorithm fails when two images have similar structures. In other words, even big intensity difference will not be detected if the entropy of the co-located pixels are similar in the two images. There are also some commercial software [10][11][12] available in the market, which include ghost removal functions in the package. However, the results are not very satisfying, and obvious ghost artifacts can still be seen.

The objective of this paper is to propose a new algorithm that provides ghost free synthesis of HDR image and actual real-time implementation. Specifically, the main contribution is the computational inexpensive ghost removal process, consisting of *Inter-pixel Relationship Function* (IRF) estimation, bi-directional comparison, moving object classification and moving object correction, as shown in Fig. 1. The first three steps detect the moving object using both co-location connections in between images and the spatial co-relations inside the target image. They remove the edge artifact and increase the detection accuracy, especially at underexposed and saturated pixels. The moving object correction recovers the missing dynamic range by fixing the moving object pixels using IRF before HDR composing.

Besides the synthesis quality, the computational complexity of each step is also minimized to $O(n)$. In addition, the whole moving object detection and correction process works independently in between two images. Thus, the ghost removal process of the k th LDR image can be executed in parallel with the capturing of the $(k+1)$ th image, which can further reduce the processing time.

The rest of the paper is organized as follows. In Section 2, a new framework of moving object detection and correction is formulated. In Section 3, the composition of HDR image is described. Experimental results and conclusion are presented in Section 4 and Section 5 respectively.

II. MOVING OBJECT DETECTION AND CORRECTION

Normally, the luminance of each pixel in the synthesized HDR image is reconstructed by the co-located pixels in the whole set of LDR images with a certain weighting scheme. However, when a moving object appears, the pixels at the

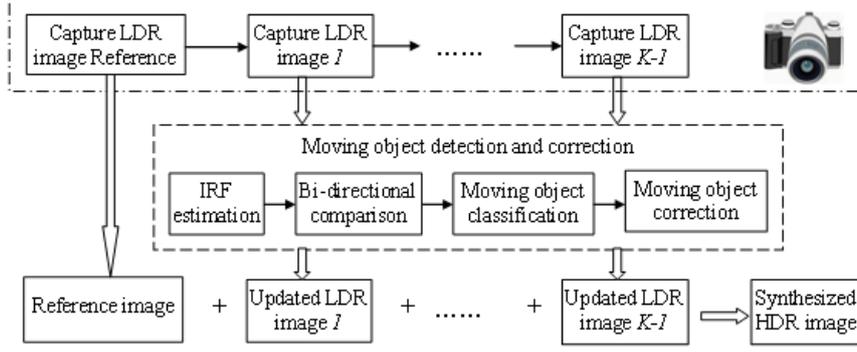


Fig. 1. System diagram of the proposed real-time HDR image synthesis

same co-location represent different objects, and an unexpected luminance value is computed which then causes ghost artifact. In order to remove the ghosts, a reference image is selected to be the background image. All other LDR images are compared with the reference image to recognize the pixels belonging to the moving object. Those pixels in each LDR image are replaced by the background pixels before the HDR synthesis will take place.

The details of the procedures are discussed in the following subsections.

A. Inter-pixel Relationship Function Estimation

The *Inter-pixel Relationship Function* (IRF) is designed to map differently exposed images to the same exposure level. In one image, if two pixels, neither under-exposed nor saturated, have the same intensity value, their represented locations on the light sensor receive the same amount of photons. We call them *spatially co-related*. When the exposure time changes, the number of photons that fall into the same light sensor changes accordingly, which result in a change of intensity values, known as *reciprocity law*. The co-location intensity values of these two pixels will be equal to each other in the new exposure. In a static scene, we can find all the spatially co-related pixels throughout the dynamic range and map them to a new exposure. Hence, with the assumption that the size of the moving object is relatively small compared to the background, the IRF is defined by

$$\Phi_{k,c}(\hat{Z}_c(x,y)) = \frac{\sum_{(u,v) \in \Omega(\hat{Z}_c(x,y))} Z_{k,c}(u,v)}{\Omega(\hat{Z}_c(x,y))}, \quad (1)$$

where k is the LDR image number, c is the color channel, Z_k and \hat{Z} represent the intensity values of the k th image and the reference image, respectively, $|\Omega(\hat{Z}_c(x,y))|$ is the cardinality of the set $\Omega(\hat{Z}_c(x,y))$ and $\Omega(\hat{Z}_c(x,y))$ is the set of spatially co-related pixels in the reference image:

$$\Omega(\hat{Z}_c(x,y)) = \{(u,v) | \hat{Z}_c(u,v) = \hat{Z}_c(x,y)\}. \quad (2)$$

The pixel values are smoothed by IRF using all the spatially co-related pixels. This minimizes the negative effect of the

sensor noise, compressing noise, sampling noise and moving object interference generated during the capturing process. The IRF estimation requires only $O(n)$ time. It is also independent from CRF and can be computed using any two images. As a result, the moving object detection can be processed in parallel with the next image capture.

B. Bi-directional Comparison

The reciprocity law fails at underexposed and saturated pixels, as they are cut-off values due to the limitation of the dynamic range of the camera. Thus, the IRF values of those regions are unreliable.

As shown in Fig. 2, three zones are defined in the reference image as *dark zone* $[Z_{min}, Z_{min} + \delta)$, *normal zone* $[Z_{min} + \delta, Z_{max} - \delta]$ and *bright zone* $(Z_{max} - \delta, Z_{max}]$, where δ is a small predefined number ($\delta = 5$ in our experiments) and $[Z_{min}, Z_{max}]$ represents the full dynamic range of the conventional image. A difference matrix is calculated only for the reliable IRF values as

$$\Gamma_{k,c}(x,y) = \begin{cases} |Z_{k,c}(x,y) - \Phi_{k,c}(x,y)|, & \hat{Z}_c(x,y) \in \Psi \\ 0, & \hat{Z}_c(x,y) \notin \Psi \end{cases}, \quad (3)$$

where Ψ represents the normal zone interval.

In order to cover the underexposed (dark zone) and saturated (bright zone) pixels, a second direction of reverse mapping and comparison is performed. The reverse IRF $\hat{\Phi}_{k,c}$ is used to map the k th image to the same exposure of the reference image, and the reverse difference matrix $\hat{\Gamma}_{k,c}$ is derived afterwards. The complete difference map of the k th image is then calculated by the maximum value of all the channels in both directions:

$$\Gamma_k(x,y) = \max \{ \Gamma_{k,c}(x,y), \hat{\Gamma}_{k,c}(x,y) \}, \quad c = R, G, B. \quad (4)$$

The long exposure image will cover the dark zone of the short exposure image, and the short exposure image will cover the bright zone of the long exposure image. If a pixel does not belong to normal zone in either of the two images, its value is insignificant in the HDR synthesis and it is not expected to cause ghost artifact. The bi-directional comparison enhances

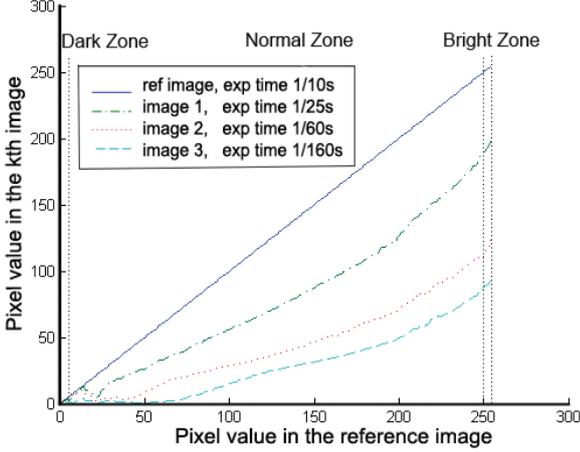


Fig. 2. Graph of the Inter-pixel Relationship Function (IRF) mapping from the reference image to other exposure levels.

the detection accuracy, especially for the underexposed and saturated pixels at the cost of $O(n)$ computation.

C. Moving Object Classification

The moving object pixels can be classified using the complete difference map:

$$\Pi_k = \{(x, y) | \Gamma_k(x, y) > \epsilon\}, \quad (5)$$

where ϵ is a threshold. As a nominal bound, ϵ needs to be kept small. However, in order to tolerate the value changes due to the different camera response at different exposure levels, it cannot be too small. We use $\epsilon = 0.1$ in our experiments.

In digital photography, images are stored in pixels and each pixel value is sampled within the dynamic range. Therefore, the border of two objects where the transition of the intensity happens, has mid intensity in between the adjacent pixels caused by digitizing. For this reason, the outline of a moving object, width of 1 pixel, can be detected erroneously with the given threshold in Equation (5). Unfortunately, those pixels contribute highly to the edge ghost artifacts. Thus, the outline of the pre-classified moving object is re-evaluated using a smaller threshold,

$$\Pi_k = \{(x, y) | D_k(x, y) > \epsilon/m, \text{ if } \exists (x \pm 1, y \pm 1) \in \Pi_k\} \cup \Pi_k, \quad (6)$$

where we use $m = 4$ in our experiments.

The outline re-evaluation takes at most $O(n)$ time because it is computed only on the contour of the pre-classified moving object.

D. Moving Object Correction

The classified moving object must be removed before HDR synthesis to prevent ghost artifacts. However, such a process loses the dynamic range information, especially when a moving object pixel is captured with right exposure. To fix the lost dynamic range, we use IRF to correct the moving object pixels in all LDR images as follows:

$$I_{k,c}(x, y) = \begin{cases} Z_{k,c}(x, y), & (x, y) \notin \Pi_k \\ \Phi_{k,c}(Z_{k,c}(x, y)), & (x, y) \in \Pi_k \end{cases}, \quad (7)$$

As described before, the IRF ($\Phi_{k,c}$) is computed using both co-location relations between the images and spatial correlations inside the target LDR image. In other words, the pixel values of the moving object are fixed using the spatially related pixels in the target LDR image with the required exposure. It helps to preserve the dynamic range of each pixel in its corresponding exposure with a small time cost less than $O(n)$ time.

E. Composing HDR Image

The final HDR image is synthesized using the corrected LDR images. Since all the images have been classified and corrected in the ghost removal process, the HDR synthesis is performed in the same way as in a static scene. In addition, the sample pixels for CRF estimation are guaranteed to be from the same object, which reduces the possible interference brought from the moving object. Similar to [1][2][4], the image radiance map is calculated in each color channel by

$$E_c(x, y) = \frac{\sum_{k=1}^K \omega(I_{k,c}(x, y)) \cdot \left(\frac{g^{-1}(I_{k,c}(x, y))}{t(k)} \right)}{\sum_{k=1}^K \omega(I_{k,c}(x, y))}, \quad (8)$$

where g is the CRF mapping function, t is the exposure time, ω is a triangle weighting function, and K denotes the total number of LDR images.

III. EXPERIMENTAL RESULT

Fig. 3 shows the HDR images composed from a set of five LDR images, in which a baby is waving his hand. The saturated area, which reflects the sun light as boxed in the bottom left image, are synthesized wrongly using uni-direction comparison. In contrast, bi-directional comparison recovers those areas.

The edge artifact, as shown in the bottom left image of Fig. 4, is the contour of the human's head. It is completely removed after the enhanced moving object classification, the outline re-evaluation.

We plotted the moving object detection result with different algorithms in Fig. 5. In the entropy approach [8], the histogram of the correspondent window (m) was calculated for each pixel in the image (n), which has the computational complexity of $O(mn)$. If the local pattern of the moving object was similar to the background, such as the hair and face in Fig. 5(a), even obvious colour difference cannot be detected. Grosch [7] estimated the reference image using CRF curve, and $O(mn)$ is needed for decoding the estimated radiance value to pixel value. Wrong detection can be made at the underexposed and overexposed pixels, such as the shoulder in Fig. 5(b), where the CRF noise was propagated. Our algorithm provides better quality and a faster solution at $O(n)$ in each of the processing

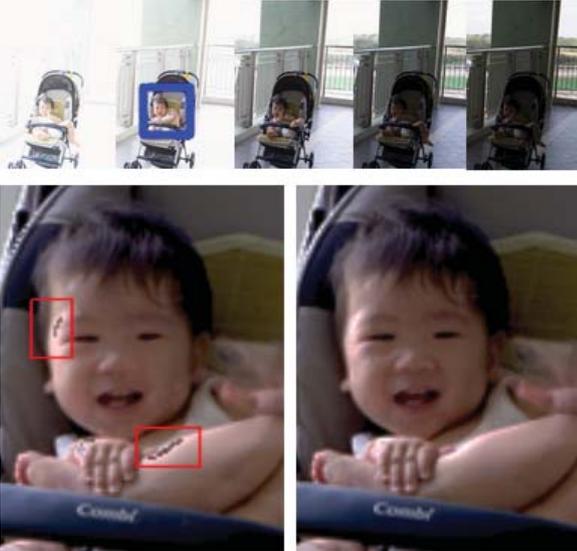


Fig. 3. Original LDR images (top), and the HDR images composed with uni-direction (bottom left) and bi-direction (bottom right) comparison result



Fig. 4. Original LDR images (top), and the HDR images synthesized without outline re-evaluation in moving object classification process (bottom left) and with outline re-evaluation (bottom right)

steps. At the same time, it detects the moving object more reliably.

In Fig. 6, we compared the HDR images synthesized from a set of five photos of a baby playing on a playground using three different software, FDRTools [10], Photomatix [11], and Qtpfsgui [12]. Among them, Photomatix delivers the best result with some background appears in the hair, and Qtpfsgui, performing worst, shows the busy baby with very blurred ghosting artifacts. The experiments suggest that our algorithm deliver better ghost free HDR image than any of the commercial software compared.

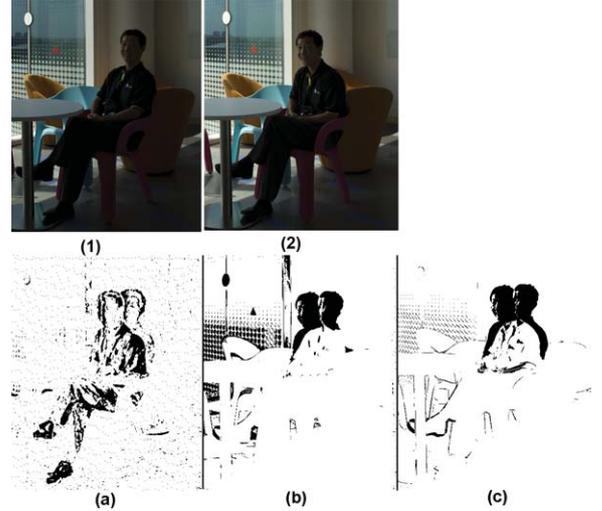


Fig. 5. Moving object detection result using (a) entropy approach, (b) CRF estimation and (c) our method, from LDR images of (1) exposure time 1/40s and (2) exposure time 1/25s.

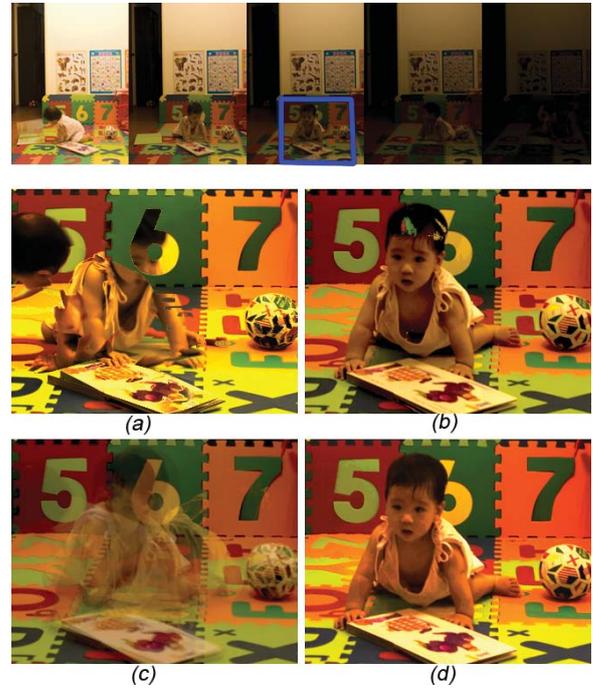


Fig. 6. HDR images synthesized by five LDR images (top) with ghost removal function enabled using (a) FDRTools, (b) Photomatix, (c) Qtpfsgui, and (d) our method.

IV. CONCLUSION

In this paper, a real-time ghost removal algorithm is presented using four important steps, IRF estimation, bi-directional comparison, moving object classification and moving object correction. With this scheme, the ghost removal process minimizes the time complexity to $O(n)$ and the dynamic range of the moving object is preserved well. In addition, the ghost removal process is independent from CRF and can be executed in parallel with the image capturing process.

Therefore, our algorithm has the potential to be embedded into image capturing devices where computation power is low.

As indicated earlier, the IRF estimation is based on the assumption that the moving object is relatively small compared to the image. Thus, it is very challenging to provide a solution when moving object occupies very large part of the image. Same concern was noted in [1] where each color channels are considered separately. It is also challenging to provide a solution where all color channels can be considered at the same time. We will study them in our future research.

REFERENCES

- [1] P. Debevec, and J. Malik, 'Recovering high dynamic range radiance maps from photographs', *SIGGRAPH Proceedings*, pp. 130-135, 1997.
- [2] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High dynamic range imaging: acquisition, display and image-based lighting*, Morgan Kaufmann, 2005.
- [3] A. R. Várkonyi-Kôczy, A. Rôvid, and T. Hashimoto, 'Gradient-based synthesized multiple exposure time color HDR image', *IEEE Transactions on Instrumentation and Measurement*, vol. 57, No. 8, Aug 2008.
- [4] S. Mann, 'Comparametric equations with practical applications in quantitative image processing', *IEEE Transaction on Image Processing*, Vol. 9, No. 8, pp. 1389-1406, Aug. 2000.
- [5] E. A. Khan, A. O. Akyuz, and E. Reinhard, 'Ghost removal in high dynamic range images'. *IEEE International Conference on Image Processing*, 2006.
- [6] M. Pedone, and J. Heikkilä, 'Constrain propagation for ghost removal in high dynamic range images'. *International Conference on Computer Vision Theory and Applications*, 2008
- [7] T. Grosch, 'Fast and robust high dynamic range image generation with camera and object movement'. *Proceedings of Vision Modeling and Visualization*, 2006, pp. 277-284.
- [8] K. Jacobs, C. Loscos, and G. Ward, 'Automatic high dynamic range image generation for dynamic scenes'. *IEEE Computer Graphics and Applications*, 28(2), pp. 84-93, Mar-Apr, 2008.
- [9] M. Grossberg, and S. Nayar, 'Determining the camera response from images: what is knowable?' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 11, pp. 1455-1467, Nov. 2003.
- [10] HDR software FDRTools Advanced, available <http://www.fdrtools.com>
- [11] HDR software Photomatix Pro, available <http://www.hdrsoft.com>
- [12] HDR software Qtpfsgui, available <http://qtpfsgui.sourceforge.net>

Paper P5

Z. J. Zhu, Z. G. Li, S. Rahardja, and P. Fränti, "De-ghosting of HDR images with double-credit intensity mapping", *IEEE International Conference on Image Processing*, pp. 1361-1364, Brussels, Belgium, Sep 2011.

DE-GHOSTING OF HDR IMAGES WITH DOUBLE-CREDIT INTENSITY MAPPING

Zijian Zhu, Zhengguo Li, Susanto Rahardja

Institute for Infocomm Research
1 Fusionopolis Way, #21-01, Singapore 138632

Pasi Fränti

University of Eastern Finland
P.O.Box 111, Joensuu, Finland 80101

ABSTRACT

Ghosting artifacts are usually caused by moving object when composing a high dynamic range image from multiple differently exposed conventional images. In this paper, a robust de-ghosting algorithm is proposed based on a double-credit intensity mapping function (IMF) and an adaptive threshold model derived from statistical training. The double-credit IMF is estimated using both pixel intensity distribution and spatial correlation. A statistical threshold model is trained from the image database, and the key parameters are determined on the fly with variance vector calculated during the IMF estimation to adapt to different scenarios. Optimal bi-directional comparison is used for further improves the detection accuracy. The experiments show the effectiveness of the proposed de-ghosting method.

Index Terms— High dynamic range, de-ghosting, intensity mapping function, adaptive threshold

1. INTRODUCTION

A *high dynamic range* (HDR) image is commonly generated from a set of differently exposed *low dynamic range* (LDR) images to overcome the dynamic range limitation of existing camera sensors. However, ghosting artifacts usually appear due to moving object, such as moving people, trees waving in the wind, etc. To solve this, de-ghosting is applied before the HDR synthesis. Typical de-ghosting process includes detecting of moving object pixels, removing them and synthesizing new pixels using only the background pixels.

There are three major challenges in the de-ghosting process. Firstly, LDR images are taken with different exposures and cannot be compared directly. It is difficult to find a good criterion to bring different exposed images to the same comparable scale. Secondly, moving object pixels are classified by a threshold, but other factors, such as capturing parameters (ISO, shutter speed), capturing hardware and light conditions, affect the pixel co-relationship and militate against the selection of a static threshold. The third challenge is how to recover the missing dynamic range information.

Various methods were proposed to tackle these problems. An iterative approach was proposed in [1] by detecting the

probability of a pixel belonging to moving object or background. The algorithm was time consuming. Local entropy was proposed in [2] by matching local pixel value distributions. However, as realized by the authors themselves, it cannot detect even large intensity differences when two co-located areas have similar local structure. Grosch [3] proposed a method using *camera response function* (CRF). However, with moving object, it is very difficult to estimate CRF accurately. Gallo et al. [4] used a similar idea with assumption that log exposure can be mapped to a 45 degree straight line, which in practice may be different due to different capturing hardware. In order to minimize the influence of moving object and capturing device, Li et al. [5] proposed a method using *intensity mapping function* (IMF). The calculation of IMF [6] involves only two images without exposure information. However, if the moving object changes statistical information of the image, the IMF will not be accurate. All the above methods [1-5] used empirical thresholds, and only the method in [5] uses multiple images to compensate the missing dynamic range with IMF. Again, if the IMF is not accurate, color differences will be still visible in the synthesized HDR image.

In this paper, we propose a de-ghosting algorithm to solve these problems. To accurately map two differently exposed images, a double-credit IMF estimation is introduced to use both pixel intensity distribution and spatial correlation information. A new threshold model is derived statistically and the key parameters are determined directly from the variance vector calculated during the IMF estimation. The missing dynamic range information is then patched using the double-credit IMF.

2. HDR DE-GHOSTING

2.1. Double-credit IMF Estimation

Histogram is commonly used to calculate IMF [6], which maps the pixel value from one image to the other (with different exposure) based on the intensity distribution. However, such a process loses position information. On the other hand, IMF calculated by spatial correlation [7] preserves position information, but can easily be distorted by moving objects. In order to generate an accurate IMF, a double-credit IMF is

proposed here by combining intensity distribution with spatial correlation.

First, reference points are detected as the points whose histogram IMF (ϕ_H) and spatial correlation IMF (ϕ_S) differ less than a given threshold δ ,

$$|\phi_H(z) - \phi_S(z)| < \delta, \quad z = 0, \dots, 255. \quad (1)$$

We use $\delta = 5$ in our experiments.

Then, a guidance curve (ϕ_G) is derived by connecting the reference points. The most effective way is to find the point (peak point) with the biggest distance to the 45 degree line. The values in between the origin and the peak point are interpolated using straight line, and the values in between the peak point to the target point (255, 255) are derived using quadratic Bezier curve as,

$$\phi_G(z) = \begin{cases} \frac{y_0}{x_0} \cdot z, & z \leq x_0 \\ B\left(\frac{255-z}{255-x_0}\right), & z > x_0 \end{cases} \quad (2)$$

where $P_0(x_0, y_0)$ is the peak point, P_2 is the target point, P_1 is (255, 255 - |x_0 - y_0|) when mapping bright image to dark image or (255 - |x_0 - y_0|, 255) otherwise, and B is defined as $B(t) = (1-t)^2 \cdot P_0 + 2(1-t) \cdot t \cdot P_1 + t^2 \cdot P_2, t \in [0, 1]$.

The guidance curve is only an approximation, based on which the final double-credit IMF (Φ) will be fine-tuned by calculating the spatial correlation using only the reliable pixels as,

$$\Phi(z) = \frac{\sum_{i \in \Omega(z)} Z_t(i) + \phi_G(z) \cdot \beta}{|\Omega(z)| + \beta}, \quad z = 0, \dots, 255, \quad (3)$$

where $\Omega(z)$ is the set of pixels with the same intensity of z in the reference image, $|\Omega(z)|$ represents the cardinality of this set, Z_t denotes the pixel value of test image, $\beta = 40$ is the initial weight given to the guidance curve, and the reliable pixels are defined by

$$\{i \mid |\phi_G(z) - Z_t(i)| < t_G, \quad i \in \Omega(z)\}. \quad (4)$$

Here, we use $t_G = 10$ to filter out potential moving object pixels. At the same time, a variance vector is derived together with the fine tune step. It will be used to calculate the key parameters for the threshold model.

An example of estimating double-credit IMF is shown in Fig. 1. We can see that both statistical information and spatial information are well utilized in the final IMF. The process involves scanning image three times for calculating histogram IMF, spatial correlation IMF, and then fine-tuning the result. In fact, each image can be sub-sampled into multiple classifications, and each scan can be simplified by using only one classification. Based on our database of 312 images, experiments show that if a classification contains around 6000-8000 pixels, it can represent the original intensity correlations well. Thus, in our de-ghosting process, if the image resolution is larger than 240x100, we divide it into three classifications by sub-sampling at $3k$, $3k + 1$, and $3k + 2$ ($k = 1, 2, \dots$).

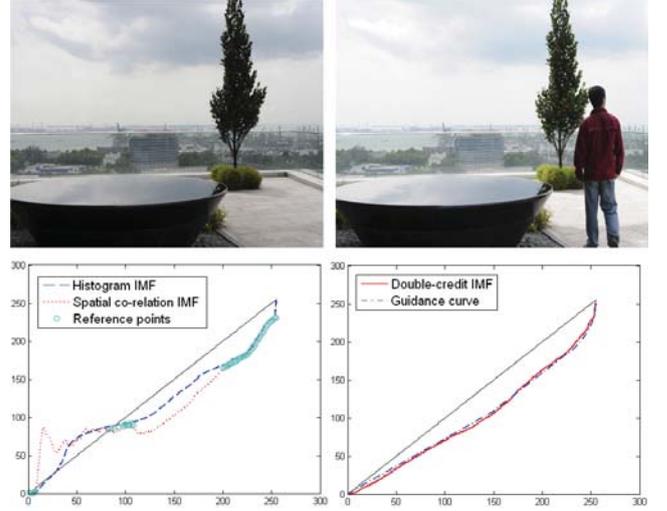


Fig. 1. An example of double-credit IMF estimation. Two input images with different exposures (Top); extraction of reference points (Bottom-left); guidance curve and the fine-tuned double-credit IMF (Bottom-right).

2.2. Optimum Bi-directional Comparison

A moving object can be detected by mapping the reference image to the test image or vice versa. Bi-directional comparison is introduced to map only the more reliable pixel in two mapping directions [5]. The reliability is represented in a weighting function by calculating the difference to a central point, where was fixed as 128 in [5]. Here, an optimum bi-directional reliability weighting is proposed to be determined by the IMF curve. An example is shown in Fig. 2, where the central point is in the middle of the two peak pixels (x_{peak1}, x_{peak2}) calculated as,

$$M = x_{peak1} \cdot \gamma + x_{peak2} \cdot (1 - \gamma). \quad (5)$$

Here $\gamma = 0.75$ if x_{peak1} is more close to 128, otherwise, $\gamma = 0.25$. Bi-direction reliability weighting is then calculated as,

$$\omega(z) = |z - M|. \quad (6)$$

2.3. A Statistical Threshold Model

A data driven training is conducted based on images captured from tripod cameras with no moving object in the scene. The training is aimed to find a threshold model suitable for different scenarios. A threshold vector is calculated based on pixel intensities (from 0 to 255) to classify 95% of the pixels as background. An example is shown at the bottom graph of Fig. 2. The image database includes more than 300 images from different cameras including Nikon D3, D300, Canon EOS-1, IXUS850, IXUS900, and images downloaded from

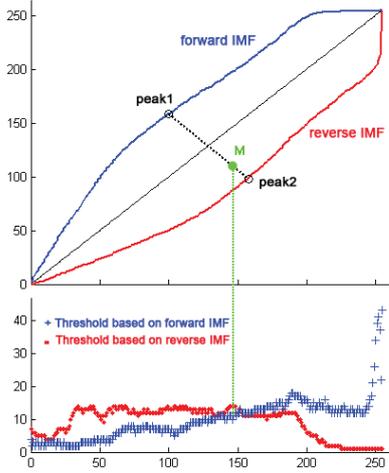


Fig. 2. The top graph plots the forward IMF and the reverse IMF of two static images. The bottom graph plots the thresholds, with which 95% of the pixels are classified as background. The forward IMF gives better performance on the left side of the central point M , and the reverse IMF minimizes the threshold value on the right.

Internet. The images have been captured using different capturing parameters at different lighting conditions to ensure extensive coverage.

From the training, we conclude the followings. First, some capturing hardware generate large noise at low intensity area. Second, the threshold can be minimized using bi-directional comparison, where the biggest threshold appears around the peak point of IMF. Third, in the same bracketing sequence, brighter image pair requires a smaller threshold than dark image pair. Obviously, a fixed threshold is not enough to generate effective result, and thus, we calculate an adaptive threshold model as

$$T(z, M) = \begin{cases} a, & 0 \leq z \leq 30 \\ a + \frac{b-a}{M-30} \cdot (z-30), & 30 < z \leq M \\ b + \frac{c-b}{255-M} \cdot (z-M), & M < z \leq 255 \end{cases}, \quad (7)$$

where z represents an intensity value, M represents the optimum bi-directional comparison central point, and a , b , c are the key parameters derived from the variance vector (V) generated during the fine-tuning of the double-credit IMF estimation. We use $a = 0.8 \cdot \sqrt{\frac{1}{30} \sum_{z=1}^{30} V(z)}$, $b = 0.8 \cdot \sqrt{\frac{1}{30} \sum_{z=M-15}^{M+15} V(z)}$, and $c = 5$.

2.4. Moving Object Detection

The moving object pixels, to be inpainted before HDR synthesis [5], are classified as,

$$\{i \mid \Re(Z_r(i), Z_t(i)) > \tilde{T}(Z_r(i), Z_t(i))\}, \quad (8)$$

where Z_r denotes the pixel intensity of the reference image, and the comparison function and the threshold are defined as

$$\Re(z_1, z_2) = \begin{cases} |\Phi(z_1) - z_2|, & \omega(z_1) > \omega(z_2) \\ |\hat{\Phi}(z_2) - z_1|, & \omega(z_1) \leq \omega(z_2) \end{cases}, \quad (9)$$

and

$$\tilde{T}(z_1, z_2) = \begin{cases} T(z_1, M), & \omega(z_1) > \omega(z_2) \\ T(z_2, M), & \omega(z_1) \leq \omega(z_2) \end{cases}, \quad (10)$$

where Φ is the IMF mapping from z_1 to z_2 , and $\hat{\Phi}$ is the reverse IMF mapping from z_2 to z_1 .

3. EXPERIMENTAL RESULTS

We compare the mean of signal-to-noise ratio (SNR) of the IMF calculated using histogram [6], spatial correlation [7] and the proposed method from 60 image pairs with known IMF, as shown in Table 1. In SET1, the moving object (small size) appears only in one image. In SET2, it appears in both images at different positions. The proposed double-credit IMF shows an average of 7 dB improvement.

Table 1. SNR comparison.

SNR	Histogram	Spatial	Proposed
SET1	20.32 dB	21.79 dB	27.50 dB
SET2	28.49 dB	18.39 dB	33.05 dB

To evaluate the performance of the proposed de-ghosting algorithm, different scenarios with and without moving object are tested together with the state of the art commercial software, such as PhotoMatix and Photoshop. A daylight scene with a moving subject is shown in Fig. 3. The challenge of this sequence is the dark hair, where the pixel values are small. It is difficult to differentiate the dark hair from the dark background in the small exposure shot. Both Photomatix and Photoshop show obvious artifacts in this region. A night scene is shown in Fig. 4, and the challenge here is the multiple light sources. When an object is highlighted by direct light source, the pixel intensities near the light source change very fast. It is common that parts of the object are detected as moving object and the rest are mis-detected as background. The de-ghosting results of Photomatix and Photoshop show obvious artifacts due to this reason. Compared to these, the proposed method generates high quality ghost-free HDR images.

4. CONCLUSION

In this paper, a robust de-ghosting method is proposed with double-credit IMF estimation and adaptive threshold model. The double-credit IMF combines both intensity distribution and spatial correlation information, which improves the accuracy of the mapping function. The threshold model derived

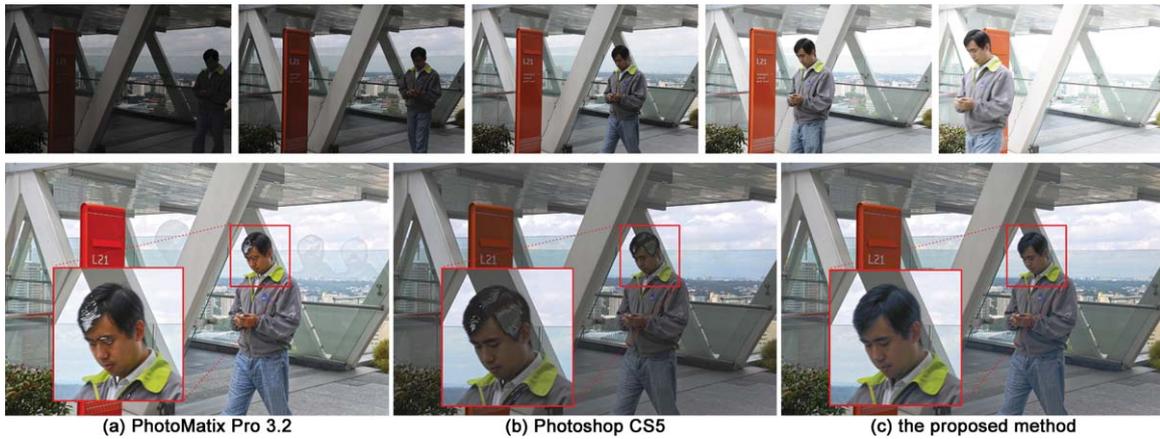


Fig. 3. Daylight scene de-ghosting with (Top) input images, using (a) PhotoMatix, (b) Photoshop, and (c) the proposed method.



Fig. 4. Night scene de-ghosting with (Top) input images, using (a) PhotoMatix, (b) Photoshop, and (c) the proposed method.

from statistical training is also reliable by testing against more than 300 pairs of images. The key parameters of the threshold model is derived on the fly from the variance vector calculated during the double-credit IMF estimation. The optimum bi-directional comparison further improves the accuracy in detecting moving object. The experimental results verify the robustness of the proposed algorithm.

5. REFERENCES

- [1] E.A. Khan, A.O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images", *IEEE International Conference on Image Processing*, pp. 2005-2008, Oct. 2006.
- [2] K. Jacobs, C. Loscos, and G. Ward, "Automatic high dynamic range image generation for dynamic scenes", *IEEE Computer Graphics and Applications*, Vol. 128, No. 2, pp. 84-93, Feb. 2008.
- [3] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement", *Vision, Modeling and Visualization*, pp. 277-284, 2006.
- [4] O. Gallo, N. Gelfand, W.C. Chen, M. Tio, and K. Pulli, "Artifact-free high dynamic range imaging", *IEEE International Conference on Computational Photography*, Apr. 2009.
- [5] Z.G. Li, S. Rahardja, Z.J. Zhu, S.L. Xie, and S.Q. Wu, "Movement detection for the synthesis of high dynamic range images", *IEEE International Conference on Image Processing*, pp. 3133-3136, Hong Kong, Sep. 2010.
- [6] M. Grossberg, and S. Nayar, "Determining the camera response from images: what is knowable?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 11, pp. 1455-1467, Nov. 2003.
- [7] Z.J. Zhu, Z.G. Li, S. Rahardja, and P. Franti, "Recover real world scene: high quality image inpainting using multi-exposed references", *Electronics Letters*, pp. 1310-1312, Vol. 45, Issue 25, Dec. 2009.

Paper P6

Z. G. Li, Z. J. Zhu, and S. Rahardja, "Fast movement detection for high dynamic range imaging", *IEEE International Conference on Image Processing*, pp. 373-376, Brussels, Belgium, Sep 2011.

FAST MOVEMENT DETECTION FOR HIGH DYNAMIC RANGE IMAGING

Zhengguo Li, Zijian Zhu and Susanto Rahardja

Signal Processing Department, Institute for Infocomm Research, 1 Fusionopolis Way, Singapore

ABSTRACT

When a high dynamic range image is synthesized by using a set of differently exposed low dynamic range (LDR) images, it is important to detect moving objects so as to remove ghosting from the final HDR image. A pixel level movement detection scheme was recently proposed in [7]. It included a pixel level similarity index for differently exposed LDR images, an adaptive threshold for the classification of pixels and an approach that utilizes intensity mapping (IMF) function for patching invalid regions. In this paper, we first propose a new adaptive threshold and a new patching approach to improve the scheme in [7]. Then, a sub-sampling method is introduced to simplify the improved movement detection scheme. Experimental results show that the improved movement detection scheme indeed outperforms the scheme in [7]. In addition, the speed is significantly improved by the proposed fast movement detection scheme.

1. INTRODUCTION.

There are many real scenes which have higher dynamic ranges than those that can be captured by digital cameras. In such scenes, a single shot low dynamic range (LDR) image usually turns out to be underexposed and/or overexposed in certain regions of the image. Because of this reason, a single shot does not have full dynamic range and one way to overcome this is to capture a set of differently exposed LDR images [1, 2]. There are two alternatives to integrate the desired information of all input images into one image. One is to synthesize a high dynamic range (HDR) image [3]. The HDR image is then compressed via a tone mapping such that it can be displayed by conventional digital devices [4]. Both HDR and LDR images are produced and it is applicable to devices with different dynamic ranges. The other is to fuse the LDR images directly into an LDR image which is much simpler [5]. The alternative is more suitable for real time applications, especially for those handheld devices with limited computational resource, such as mobile phones or digital cameras.

When an HDR/LDR image is synthesized for an outdoor scene by using multiple differently exposed LDR images, there are two challenging problems to be addressed. One is to detect moving objects in order to remove ghosting artifacts from the final HDR/LDR image. The other is to align the LDR images such that the synthesized HDR/LDR image is not blurred [6]. As with the popular HDR acquisition approach in [3], all input LDR images are assumed to be perfectly aligned, possibly using the registration algorithm in [6]. To remove ghosting artifacts due to moving objects in the scene, the pixels of all LDR images are required to be properly classified into valid or invalid, and only valid pixels are used to generate the HDR/LDR image. Recently, a pixel level movement detection scheme was proposed in [7]. This scheme was shown

to remove ghosting artifacts significantly better than existing commercial software [8] which is also demonstrated in Fig. 1. The paper has three major contributions. The first contribution is a new pixel level similarity index for differently exposed LDR images. The second is an adaptive threshold for the classification of pixels into valid or invalid. The threshold is adaptive to the values of a pixel to be detected and its reference pixel and the exposure times of two images. The third is an intensity mapping function (IMF) based approach for patching invalid pixels in a detected image by using the co-located pixels in the reference image. Since the IMFs are computed by using all pixels of these two images, the IMF based patching approach can be regarded as a global one. All pixels are exhaustively detected by using the scheme in [7] which makes the scheme complex, especially for real time applications on digital cameras and mobile phones. It is thus desirable to reduce the complexity of the scheme in [7].



Fig. 1. Comparison of different movement detection schemes.

In this paper, we first improve the movement detection scheme in [7], particularly on the latter two contributions. Besides being adaptive to the values of two co-located pixels and the exposure times of two images, the new threshold is also adaptive to the ISO value and the average exposure value of all LDR images. The patching scheme is also improved by involving local information of under-exposed/saturated pixels in the reference image and introducing a cross-image smoothing method for invalid pixels. Experimental results show that by using the improved movement detection scheme, the performance is indeed improved. We then propose a sub-sampling based method to simplify the improved movement detection scheme. The proposed method is based on an observation that small portions of differently exposed images belong to moving objects. Instead of detecting moving objects by checking all pixels as in [7], a sub-sampling based method is presented to detect moving objects by only checking part of pixels. If a pixel is detected as a background one, its neighboring pixels are skipped. Otherwise, its neighboring pixels are further checked by using a sub-sampling method. Experimental results verify that

the complexity of the movement detection scheme in [7] can be significantly reduced.

The rest of this paper is organized as follows. An improved movement detection scheme is provided in Section 2. A sub-sampling based method is proposed in Section 3 to simplify the improved movement detection method. Experimental results are provided in Section 4 to show the efficiency of the proposed scheme. Concluding remarks are given in Section 5.

2. AN IMPROVED MOVEMENT DETECTION SCHEME

Let $Z_{k,l}(p)$ denote the intensity of the l th color channel at position p when the k th LDR image is captured, i.e., p is a spatial position, l indexes over color channels of red, green and blue, and k indexes over exposure time Δt_k . Such a set of LDR images is known as a Wyckoff set [1]. Let \hat{Z}_k be the reference image of \mathbf{Z}_k for the classification of pixels in \mathbf{Z}_k as valid or invalid. An image is selected as the initial reference image for the movement detection. For simplicity, the image is denoted as k_0 and its value is set as the middle one. All pixels in \mathbf{Z}_{k_0} are marked as valid. Let p denote a coordinate (x, y) . A pixel $Z_k(p)$ ($k \neq k_0$) is marked as valid if it is similar to its co-located pixel $\hat{Z}_k(p)$ [7], i.e.,

$$S(Z_k(p), \hat{Z}_k(p)) > \text{Thr}_k(p). \quad (1)$$

Otherwise, it is marked as invalid.

2.1. A Bi-directional Similarity Index

The function $S(Z_k(p), \hat{Z}_k(p))$ in Equation (1) is given by

$$S(Z_k(p), \hat{Z}_k(p)) = \frac{\sum_{l=1}^3 2\Phi_{k,l}(p)\Psi_{k,l}(p) + 1}{\sum_{l=1}^3 [\Phi_{k,l}^2(p) + \Psi_{k,l}^2(p)] + 1}, \quad (2)$$

where $\Phi_{k,l}(p)$ and $\Psi_{k,l}(p)$ are constructed by using a bi-directional mapping method as [7]

$$\Phi_{k,l}(p) = \begin{cases} \Lambda_{l,\pi(k),k}(\hat{Z}_{k,l}(p)); & \text{if } w(Z_{k,l}(p)) \leq w(\hat{Z}_{k,l}(p)) \\ Z_{k,l}(p); & \text{otherwise} \end{cases},$$

$$\Psi_{k,l}(p) = \begin{cases} \Lambda_{l,k,\pi(k)}(Z_{k,l}(p)); & \text{if } w(Z_{k,l}(p)) > w(\hat{Z}_{k,l}(p)) \\ \hat{Z}_{k,l}(p); & \text{otherwise} \end{cases}.$$

Here, the weighting function $w(z)$ is defined as [3], and $\pi(k)$ corresponds to the exposure time of image \hat{Z}_k [7], $\Lambda_{l,k,\pi(k)}$ and $\Lambda_{l,\pi(k),k}$ are two IMFs [9], $\Lambda_{l,k,\pi(k)}$ maps intensity values in $\mathbf{Z}_{k,l}$ into $\hat{\mathbf{Z}}_{k,l}$ and $\Lambda_{l,\pi(k),k}$ vice versa.

2.2. An Improved Adaptive Threshold

The value of $\text{Thr}_k(p)$ in Equation (1) is adaptive to the values of $Z_k(p)$, $\hat{Z}_k(p)$, Δt_k and $\Delta t_{\pi(k)}$, the ISO value G for capturing the images and the average exposure value of all LDR images $\bar{E}V$. It is given as

$$\text{Thr}_{k,\pi(k)}(p) = \frac{2(1 - \xi_{k,\pi(k)}(p))}{1 + (1 - \xi_{k,\pi(k)}(p))^2}, \quad (3)$$

where the value of $\xi_{k,\pi(k)}(p)$ is computed as

$$\xi_{k,\pi(k)}(p) = \left(\frac{1}{32} + \max_{1 \leq l \leq 3} \{ \epsilon(Z_{k,l}(p)), \epsilon(\hat{Z}_{k,l}(p)) \} \right) h\left(\frac{\Delta t_a^M}{\Delta t_{k,\pi(k)}^m}, \frac{\Delta t_{k,\pi(k)}^M}{\Delta t_{k,\pi(k)}^m}, \bar{E}V, G \right),$$

the intensity factor $\epsilon(z)$ is defined as

$$\epsilon(z) = \begin{cases} \frac{1}{8} \left(1 - \frac{2z}{255} \right) \left(\frac{z}{10} \right)^{32}; & \text{if } z < 128 \\ 0; & \text{otherwise} \end{cases},$$

Δt_a^M is the maximal value of all exposure times. $\Delta t_{i,j}^M$ and $\Delta t_{i,j}^m$ are the maximal and minimal values of two exposure times Δt_i and Δt_j , respectively. $\bar{E}V$ is the average exposure value of all LDR images and is defined as $\log_2(100\omega^2/(G\Delta\bar{t}))$. Here, ω and $\Delta\bar{t}$ are the aperture value and the geometrical mean value of all exposure times, respectively. The function $h(x_1, x_2, x_3, x_4)$ is defined as

$$h(x_1, x_2, x_3, x_4) = x_1^{\frac{1}{32}} \max\left\{ 1, \frac{1}{\log_2(x_2)} \right\} \left(\frac{10}{x_3} \right)^{\frac{1}{4}} \left(\frac{x_4}{100} \right)^{\frac{1}{32}}.$$

2.3. An Improved Patching Scheme

As the correlation among two successive images is the strongest, $\hat{\mathbf{Z}}_{k_0+1}$ and $\hat{\mathbf{Z}}_{k_0-1}$ are chosen as \mathbf{Z}_{k_0} . $\hat{\mathbf{Z}}_k$ ($|k - k_0| > 1$) is updated by

$$\hat{\mathbf{Z}}_k(p) = \begin{cases} \vec{\mathbf{Z}}_{\pi(k)}(p); & \text{if } Z_{\pi(k)}(p) \text{ is invalid} \\ Z_{\pi(k)}(p); & \text{otherwise} \end{cases}.$$

It is very important to properly select the values of $\vec{\mathbf{Z}}_{\pi(k)}(p)$ to fill in the invalid regions of $\mathbf{Z}_{\pi(k)}$. A new pixel, $\vec{\mathbf{Z}}_{\pi(k)}(p)$, is synthesized by using $\hat{\mathbf{Z}}_{\pi(k),l}(p)$ and $\Lambda_{\pi(\pi(k)),\pi(k),l}(z)$ ($1 \leq l \leq 3$) as

$$\vec{\mathbf{Z}}_{\pi(k),l}(p) = \Lambda_{l,\pi(\pi(k)),\pi(k)}(\hat{\mathbf{Z}}_{\pi(k),l}(p)); \quad 1 \leq l \leq 3. \quad (4)$$

If none of $\hat{\mathbf{Z}}_{\pi(k),l}(p)$ ($1 \leq l \leq 3$) is saturated or under-exposed, $\vec{\mathbf{Z}}_{\pi(k)}(p)$ is then used to replace $Z_{\pi(k)}(p)$. Otherwise, $\vec{\mathbf{Z}}_{\pi(k)}(p)$ is first filtered by a locally weighted averaging filter as

$$\vec{\mathbf{Z}}_{\pi(k),l}(p) = \frac{\vec{\mathbf{Z}}_{\pi(k),l}(p) + \sum_{p' \in \Omega(p,\rho)} g_{\pi(k),l}(p') \hat{\mathbf{Z}}_{\pi(k),l}(p')}{1 + \sum_{p' \in \Omega(p,\rho)} g_{\pi(k),l}(p')}, \quad (5)$$

where $\Omega(p,\rho) = \{p' = (x', y') \mid |x - x'| \leq \rho, |y - y'| \leq \rho\}$ with ρ being a predefined parameter. $g_{\pi(k),l}(p')$ is the weighting factor of $\hat{\mathbf{Z}}_{\pi(k),l}(p)$. Since a pixel p' nearby a moving region has a high possibility to belong to a moving object, $g_{\pi(k),l}(p')$ is a monotonically increasing function of the distance between pixels p and p' . Meanwhile, it is a monotonically decreasing function of the absolute difference between $\hat{\mathbf{Z}}_{\pi(k),l}(p')$ and $\hat{\mathbf{Z}}_{\pi(k),l}(p)$. In this paper, the value of ρ is selected as 7, and $g_{\pi(k),l}(p')$ is chosen as

$$g_{\pi(k),l}(p') = \|p' - p\|^2 \exp^{-4(\hat{\mathbf{Z}}_{\pi(k),l}(p') - \hat{\mathbf{Z}}_{\pi(k),l}(p))^2} V_{\pi(k)}(p'),$$

$$V_{\pi(k)}(p') = \begin{cases} 1; & \text{if } Z_{\pi(k)}(p') \text{ is valid or updated} \\ 0; & \text{otherwise} \end{cases}$$

$\tilde{Z}_{\pi(k)}(p)$ is then used to replace $Z_{\pi(k)}(p)$. After all pixels are updated, all invalid pixels are finally smoothed by using its eight neighboring pixels as

$$\hat{Z}_{k,l}(p) = \frac{\gamma_{\pi(k),l}(p)\hat{Z}_{k,l}(p) + \sum_{p'} \Gamma_{\pi(k),l}(p')\hat{Z}_{k,l}(p')}{\gamma_{\pi(k),l}(p) + \sum_{p'} \Gamma_{\pi(k),l}(p')} \quad (6)$$

$$\Gamma_{\pi(k),l}(p') = \exp \left(-\frac{16(\hat{Z}_{\pi(k),l}(p') - \hat{Z}_{\pi(k),l}(p))^2}{2^{\lfloor \log_2(\hat{Z}_{\pi(k),l}(p)+1) + 0.5 \rfloor}} \right), \quad (7)$$

where the value of $\gamma_{\pi(k),l}(p)$ is 1 if $\hat{Z}_{\pi(k)}(p)$ is under-exposed or saturated, otherwise 8. $\lfloor a \rfloor$ is the largest integer less than a . It is shown from Equations (6) and (7) that the weighting factor $\Gamma_{\pi(k)}(p')$ is determined by $\hat{Z}_{\pi(k)}(p')$ instead of $\tilde{Z}_{\pi(k)}(p')$. This is based on an observation that $\tilde{Z}_{\pi(k)}(p')$ is less reliable than $\hat{Z}_{\pi(k)}(p')$. The proposed smoothing method can thus be called “a cross-image smoothing”.

3. A SUB-SAMPLING BASED FAST MOVEMENT DETECTION SCHEME

The improved movement detection scheme is simplified by using a sub-sampling based method in this section.

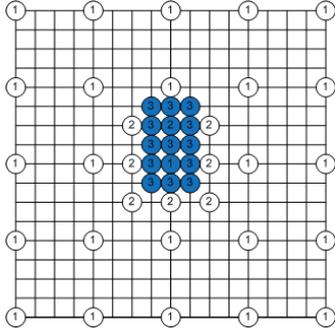


Fig. 2. A sub-sampling based movement detection scheme.

An example with a sub-sampling factor being selected as 4 ($= 2^{3-1}$) is adopted to illustrate the proposed fast movement detection scheme. The example is shown in Fig. 2. The pixels that are labeled with 1 are checked at the first round, and they are marked by the blur color if they are invalid. The neighboring pixels of a blue pixel labeled with 1 are then checked by using a sub-sampling method. In other words, all neighboring pixels that are labeled by 2 are checked at the second round. They are also marked by the blur color if they are invalid. Finally, the eight neighboring pixels of a blue pixel labeled with 2, i.e., those pixels are labeled by 3, are checked in the final round. Only 1/16 of all pixels are detected at the first round. Since only a small portion of pixels in an LDR image belongs to moving objects, the second and third rounds of detections are only conducted for a small amount of pixels in the LDR image. As such, the complexity of the improved movement detection scheme can be significantly reduced. Meanwhile, since many pixels are in the neighborhoods

of two pixels, a flag is attached to each pixel so as to indicate whether it has been detected. With the flag, each pixel will only be detected once.

Besides simplifying the proposed movement detection scheme, another important issue is to design a parallel simplified method. This can be achieved by selecting the sub-sampling factor as a power of 3. An example is demonstrated in Fig. 3 with the sub-sampling factor being chosen as 9. All pixels that are marked by 1 are detected at the first round. The neighboring pixels of a blur pixel labeled by 1 are further detected by using a sub-sampling method. It is demonstrated in Fig. 3 that the sub-sampling method can run in parallel.

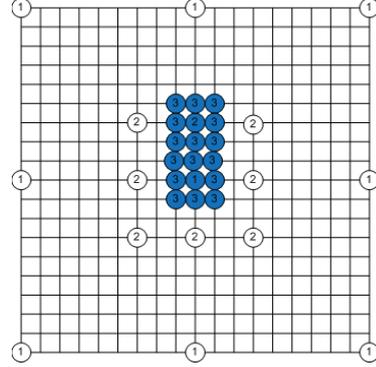


Fig. 3. A parallel sub-sampling based movement detection scheme.

4. EXPERIMENTAL RESULTS

In this section, we first compare the improved movement detection scheme with the one in [7] by testing two sets of differently exposed images. The first set is composed of 3 differently exposed images as demonstrated in Fig. 4. The second set is composed of five images as illustrated in Fig. 5. The initial reference images are selected as the second image and the first image for the first set and the second set, respectively. It is shown in Figs. 6 and 7 that the improved scheme can be adopted to produce better HDR images.



Fig. 4. A set of three differently exposed LDR images with waving leaves.

We then test the proposed fast movement detection scheme with the sub-sampling factor being selected as 4. Both image sequences in Figs. 4 and 5 are tested. 47.18% and 76.37% of pixels are detected for two input images of the former, respectively.



Fig. 5. A set of five differently exposed LDR images with a moving subject.



Fig. 6. The final LDR images for the first set.

24.04%, 35.69%, 50.92% and 54.89% of pixels are checked for four input images of the latter, respectively. Besides these two sequences, another LDR image sequence that is shown in Figs. 8 is also tested. 39.66%, 41.35%, 51.14%, 55.4% and 77.22% of pixels are detected for five images of the sequence in Fig. 8, respectively. Overall, the number of detected pixels is reduced by up to 58.61%. It should be noted that the speed could be further improved by choosing a larger sub-sampling factor such as 8 or 9.

5. CONCLUSION

An improved movement detection scheme has been provided by introducing a new adaptive threshold and a new patching approach. A sub-sampling based method has also been proposed to simplify the improved movement detection scheme. By using the movement detection scheme introduced in this paper, the quality of the final high dynamic range images is improved. The speed is also



Fig. 7. The final LDR images for the second set.



Fig. 8. A set of six differently exposed LDR images with a moving subject.

significantly increased by the proposed fast movement detection scheme.

References

- [1] C. W. Wyckoff, "An experimental extended response film," *SPIE Newslett.*, pp.16-20, June-July 1962.
- [2] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *IEEE Trans. on Image Processing*, vol. 9, no. 8, pp. 1389-1406. Aug. 2000.
- [3] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photograph," In *Proceedings SIGGRAPH 1997*, pp.369-378, 1997.
- [4] Z. G. Li, S. Rahardja, S. S. Yao, J. H. Zheng, and W. Yao, "High dynamic range compression by half quadratic regularization," In *2009 IEEE International Conference on Image Processing*, pp. 3169-3172, Cairo, Egypt, Nov. 2009.
- [5] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion," In *Pacific Graphics 2007*, pp. 382-390, Maui, USA, Oct. 2007.
- [6] K. Jacobs, C. Loscos, and G. Ward, "Automatic high dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications*, Vol. 128, No.2, pp.84-93, Feb. 2008.
- [7] Z. G. Li, S. Rahardja, Z. J. Zhu, S. L. Xie, and S. Q. Wu, "Movement detection for the synthesis of high dynamic range images," In *IEEE International Conference on Image Processing*, September 2010, Hongkong, China.
- [8] FDRTools, Photomatix and Qtpfsgui softwares: http://fdrtools.com/front_e.php, <http://www.hdrsoft.com/>, <http://qtpfsgui.sourceforge.net/>, 2009.
- [9] M. Grossberg and S. Nayar, "Determining the camera response from images: what is knowable?" *IEEE Transactions on pattern analysis and machine intelligence*, pp. 1455-1467, Vol. 25, No. 11, Nov. 2003.

Paper P7

Z. J. Zhu, Z. G. Li, S. Q. Wu, J. H. Zheng, W. Yao, P. Fränti, and S. Rahardja, "Noise reduced high dynamic range tone mapping base on information content weighting", Manuscript

Noise Reduced High Dynamic Range Tone Mapping base on Information Content Weighting

Zijian Zhu¹, Zhengguo Li¹, Shiqian Wu¹, Jinghong Zheng¹, Wei Yao¹, Fränti Pasi², and Susanto Rahardja¹

¹Institute for Infocomm Research, Singapore

²University of Eastern Finland, Finland

In this paper, we proposed a noise reduced tone mapping method based on information content weighting, where the perceptually unimportant pixels are smoothed during the decomposition in two steps. First, a base layer information content weighting is introduced to give high fidelity to the data term based on the ratio of the local pixel power and the overall noise power. Then, the detail layer is subtract based on the mutual information between the original image luminance and the clean base layer. Experiments show the effectiveness in terms of noise reduction and visual quality improvement. Compare to the state-of-the-art weighted-least-squares tone mapping, the proposed method spends 10-15% more in processing time and improves the quality by 1-4dB on the average. The proposed method can also be used for the noise reduction of conventional low dynamic range images.

Index Terms—high dynamic range, tone mapping, edge-preserving decomposition, de-noising

I. INTRODUCTION

DYNAMIC range of a real world scene is defined as the ratio between the largest and the smallest light intensities in the scene. Due to hardware limitation, an image captured using conventional camera is not enough to keep the full dynamic range. Therefore, a high dynamic range (HDR) image is usually reconstructed using either new designed sensors [1], [2] or synthesized using multiple differently exposed images [3]. Unfortunately, an HDR image cannot be shown directly on a conventional display device due to hardware limitations. Although HDR-solution-based monitor [4] and projector [5] have been proposed, they are still not widely used due to quality and cost issues. Thus, compression from an HDR image into a display-able image are studied known as HDR tone mapping.

Studies show that the perceptual of human visual system (HVS) is more sensitive to the high frequency components than the low frequency components [6]. Thus, in order to keep the high dynamic range content in a low dynamic range image, the low frequency components are compressed while the high frequency components are retained. Previous tone mapping methods can be classified into two major categories: global operators and local operators.

The global operators map the intensity values of a large range into a small range in a spatially invariant way [7]. In order to increase the scene contrast, most of these operators are non-linear [8]. There are two major advantages of the global operators. First, the relative contrast is well preserved, which means the brightness correlation of any two pixels in the HDR image will not be reversed in the tone mapping result. Second, they are simple and computational efficient, and therefore, they are good candidates for generating HDR video sequences [9]. The problem of the global operators is that they are spatially invariant, which makes them difficult to keep small local structures.

To improve the visual quality, local operators are studied

using not only the tone curve, but also spatially related neighbouring pixels. Reinhard et al. [10] treated the tone mapping process as a traditional photo development process. They proposed an auto dodge-and-burning operator using circularly symmetric Gaussian function with a locally adaptive scale factor. Another approach divided the input image into small overlapping windows [11]. Then, it reduced the overall contrast by suppressing contrast in each window, while keep the relations among the windows by using a guidance map. Both two methods directly attenuate the image luminance map, which avoid the problems associated with layer decomposition. However, their quality is highly relied on the choosing of the parameters.

Following the HVS model, most tone mapping methods decompose an HDR image into multiple layers and process them separately. Gradient-decomposition-based method manipulated the gradient field by attenuating the magnitudes of large gradients [12]. The compressed image was obtained by solving a Poisson equation on the modified gradient field. Another major approach used luminance decomposition based on Retinex theory [13], where an image was regarded as a product of an illuminance component and a reflectance component. Gaussian filtering method for decomposition was proposed in [14], and soon be replaced by bilateral filtering [15], [16] for its better edge preservation. However, halo artifact is usually seen in the compressed image due to edges leaking to the reflectance component.

To solve this, more sophisticated edge-preserving operators are proposed to decompose an HDR image into a base layer and a detail layer. Farbman et al. [17] proposed a weighted least squares (WLS) framework with progressive base layer decomposition. Other methods focused on the smoothness constraint [18], [19] and the localized data term [6] for better visual quality and fast processing.

Most tone mapping methods focused on how to keep fine details. Unfortunately, noise in an HDR image can be easily treated as the fine detail and retained in the final image. For

example, the gradient-decomposition-based method magnifies the small magnitude to review the fine details. Noise, if not carefully treated, can be magnified and become more obvious in the tone mapping result.

In order to reduce noise, Lee et al. proposed a scale-decomposition-based method [20]. It used a discrete Haar wavelet transform to decompose an HDR image into four subbands. A noise reduction step was introduced by filtering the subband with the lowest frequency using bilateral filter, and smoothing the rest subbands using soft-thresholding. The problem of using multiscale techniques is that the original signal may be distorted at the composition stage which generates the halo artifacts [21], if the parameters are not carefully selected.

In this paper, we proposed a noise reduced tone mapping algorithm based on edge-preserving operators, as they represent state-of-the-art quality [6]. The proposed method incorporates the information content weightings (ICW), and works in two steps. First, a base layer ICW is introduced for a better fidelity between the base layer and the original HDR image based on the information theory of how much information can be received after passing through a noise channel. It pushes the noise out of the based layer during the edge-preserving decomposition. In the second step, the detail layer is derived using another ICW, where the mutual information between the input image and the derived clean base image are given higher weights. The mutual information weighting is derived from the information fidelity criterion [22]. With the parameters carefully chosen, the proposed method reduces the noise significantly without any halo artifacts.

The rest of the paper is organized as follows. Section II reviews some background in the field of edge-preserving tone mapping and information fidelity criterion. Section III describes the proposed noise reduced tone mapping algorithm. Experimental results, comparison and discussion are provided in Section IV. And the paper is concluded in Section V.

II. BACKGROUND

A. Edge-Preserving Tone Mapping

In the Retinex theory [13], an image (I) is regarded as a product of two components ($I = L \cdot R$): a illuminance component (L) which contains large luminance variance, and a reflectance component (R) which contains intrinsic information.

The fundamental of the edge-preserving decomposition is based on the Retinex theory, where an HDR image is decomposed into a base layer (B) with large luminance variance and a detail layer (D) with fine details. Then, the base layer is compressed, and the detail layer is either kept or enhanced, before they are re-composed into a low dynamic range image. The decomposition is an ill-posed problem, and the most widely used methods are generalized as

$$\iint \left((B - I)^2 + \lambda \Phi(B, I) \right) dx dy, \quad (1)$$

where I denotes the log luminance of the HDR image, λ is a smoothing coefficient and Φ represents a smooth term. The

first part of the equation is a data term in L2 norm striving to minimize the distance from the base layer to the original luminance. The smooth term (Φ) defines the local smoothness of the base layer during the decomposition. And λ is used to balance between the data term and the smooth term for a good tradeoff between fidelity and smoothness. The bigger value of λ generates smoother illumination in the base layer and richer contrast in the detail layer. Here, B , D and I are all defined in log luminance domain, and therefore, the original product is rewrite as

$$D = I - B. \quad (2)$$

Many researches have been conducted on different smooth terms. In the weighted least squares (WLS) framework [17], the smooth term is defined as the partial derivatives of the base layer as

$$\Phi(B, I) = \left(a_x(I) \left(\frac{\partial B}{\partial x} \right)^2 + a_y(I) \left(\frac{\partial B}{\partial y} \right)^2 \right), \quad (3)$$

where a_x and a_y are spatially varying weights that control the smoothness of the partial derivatives. A similar approach is introduced by Guarnieri et al. [18] as

$$\Phi(B, I) = \omega |\nabla B|^2, \quad (4)$$

where ω is a spatial varying coefficient that is inversely to gradient norm of the original illuminance. Li et al. proposed another smooth term based on half quadratic regularization [19] as

$$\Phi(B, I) = (\varphi(b_x, \nabla I_x) + \varphi(b_y, \nabla I_y)), \quad (5)$$

where φ presents the coefficients for the half quadratic regularization.

B. Information Fidelity Criterion

In image quality assessment, when the reference image is present, the measurement of the fidelity from the test image to the reference image defines the quality of the test image. There are many fidelity terms, such as the widely used peak signal-to-noise ratio (PSNR) based on mean square error (MSE). However, the correlation between the MSE and the human perceptual of image quality is not very tight. Therefore, an information fidelity criterion (IFC) [22] is proposed based on natural scene statistics [23], which correlates to the sensitivity of human visual system (HVS).

The IFC is derived from an information-theoretic framework using Gaussian scale mixtures (GSM) model in the wavelet domain. Each subband of the wavelet decomposition is modeled as a GSM random field (C), which is a product of two independent stationary random field as

$$C = S \cdot U = \{S_i \cdot U_i : i \in I\}, \quad (6)$$

where S is a random filed of positive scalars, U is a random field of Gaussain scalar with mean zero and variance σ_U^2 , and I denotes the set of spatial indices for the random field. Similarly, the distortion is modeled as a simple signal attenuation (G) and an additive Gaussian noise (V) as

$$D = G \cdot C + V = \{g_i \cdot C_i + V_i : i \in I\}. \quad (7)$$

The IFC is defined as the mutual information between the source and the distorted images as

$$IFC(C^N; D^N | S^N = s^N) = \frac{1}{2} \sum_{i=1}^N \log_2 \left(1 + \frac{g_i^2 s_i^2 \sigma_U^2}{\sigma_V^2} \right), \quad (8)$$

where s^N denotes a realization of S^N , and σ_V^2 denotes the variance of the additive noise.

Wang et al. [24] further extended the IFC by combining it with the similarity index (SSIM) [25]. Concurrently, they also made an interesting discovery that with the information content weighting, even the widely criticized PSNR can be converted to a competitive perceptual quality measurement [24]. A saliency-based pooling strategy is proposed with a saliency map constructed using local variance based on the information theory of how information is received through a noise channel [26] as

$$\omega = \frac{1}{2} \log_2 \left(1 + \frac{P_{source}}{P_{channel}} \right), \quad (9)$$

where ω represents the amount of the received information, and P_{source} and $P_{channel}$ denotes the source power and the channel noise power, respectively.

III. NOISE REDUCED TONE MAPPING

The proposed noise reduced tone mapping is designed to be working with the state-of-the-art edge-preserving tone mapping. We choose the weighted least square (WLS) framework [17] in our implementation for its simplicity and fast processing. The proposed method consists of two weighting factors: a base layer information content weighting (ICW) based on local variance and a detail layer ICW based on information fidelity criteria. Details are given in the following subsections.

A. Noise Reduced Base Layer Decomposition

In edge-preserving decomposition, the data term defines the fidelity between the based layer and the original luminance. It is proposed in an L2 norm for easy calculation. At the same time, it also represents a mean square difference between two components, which does not link to the HVS tightly [22], [24]. Although the smooth term helps refine the decomposition, it is important to add an information content weighting directly on the data term as

$$\iint \left(\omega_b \cdot (B - I)^2 + \lambda \Phi(B, I) \right) dx dy, \quad (10)$$

where ω_b is the spatial varying ICW based on the perceptual of HVS. The principle is that a higher weight is given to the pixels that are perceptually more sensitive in assessing the image quality, and therefore, the base layer will be more close to the original image. On the contrary, when processing the pixels that are less sensitive in the human perceptual, commonly low frequency components, the decomposition is bias towards the smooth term.

Inspired by Equation 9, the base layer ICW is defined as

$$\omega_b(p) = \frac{1}{2} \log_2 \left(1 + \frac{\sigma^2(p)}{\sigma_c^2} \right), \quad (11)$$

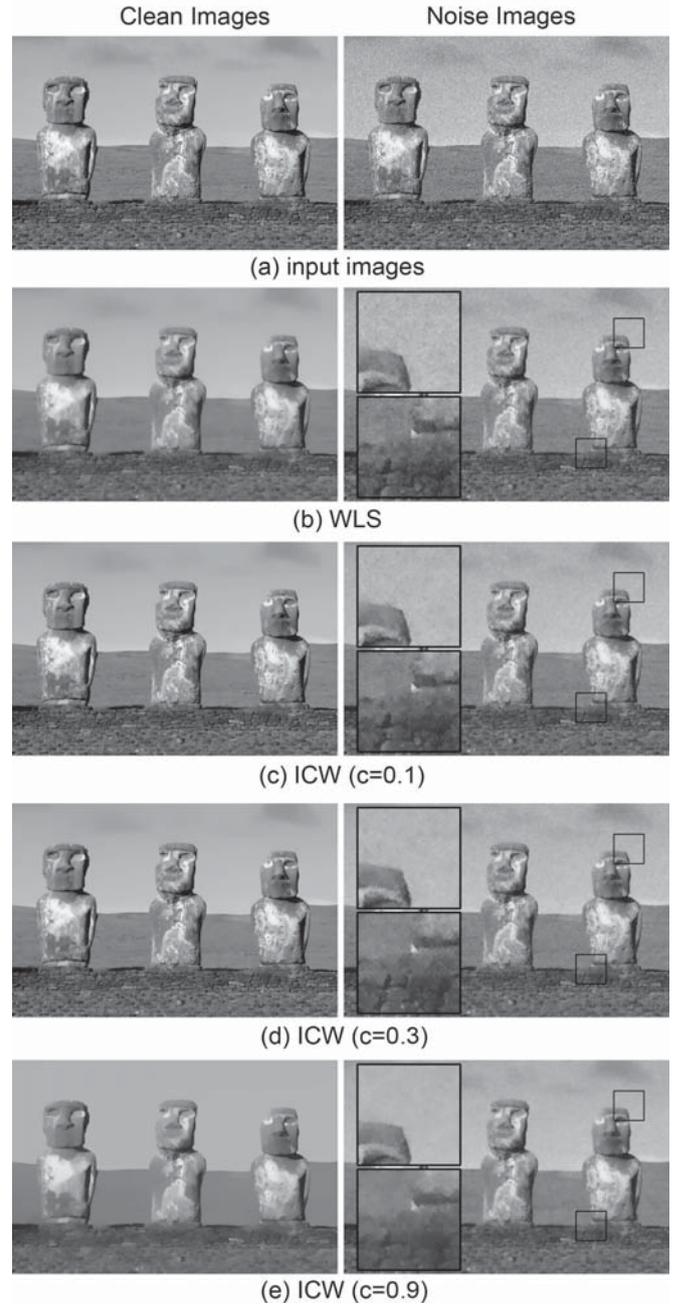


Fig. 1. The behavior of different saliency coefficient (c) on a clean image (left column) and a noise image (right column). The noise image is generated by adding a zero mean Gaussian noise with variance of 0.01. The bigger the saliency coefficient, the smoother the base layer. $c = 0.3$ is chosen as a balanced value in our implementation.

where $\sigma^2(p)$ denotes the local variance at each pixel p with a small window, and σ_c^2 is a constant represents the channel noise power ($\sigma_c^2 \in \Gamma(p) = \{\sigma^2(p), p \in P\}$). Saliency coefficient (c) defines the level of channel noise as

$$\sigma_c^2 = \vec{\Gamma}(c \cdot P), \quad (12)$$

where P denotes the total number of pixels in the image, and $\vec{\Gamma}$ is the ascending sort of Γ . For example, $c = 0.5$ represents the median value of Γ .

The saliency coefficient is an important factor that controls

the saliency of the base layer ICW. A small saliency coefficient indicates a small channel noise, and results a high fidelity between base layer and the original image. A large saliency coefficient indicates a large channel noise, and results a smooth base layer. If the saliency coefficient is too big, the base layer can be over smoothed, as shown in Fig. 1 (e), where the cloud is completely removed when there is no noise. In our implementation, $c = 0.3$ is chosen as a balanced value. However, if the noise level of the input HDR image is very high, a bigger saliency coefficient generates better denoising result.

Combined with the smooth term defined in Equation (3), the full objective function can be rewritten using matrix notation as

$$(B-I)^T \omega_b (B-I) + \lambda (B^T D_x^T A_x D_x B + B^T D_y^T A_y D_y B), \quad (13)$$

where A_x and A_y are diagonal matrices containing the smoothness weights a_x and a_y , and D_x and D_y denote discrete differentiation operators. A linear system is derived by minimizing the objective function as

$$(I_m + \lambda \omega_b^{-1} \Psi) B = I, \quad (14)$$

where $\Psi = D_x^T A_x D_x + D_y^T A_y D_y$, and I_m denotes the identity matrix. The solution of the linear system is the same as the WLS-based decomposition, and therefore, it shares the similar frequency response [17]. Thus, we keep the same smoothing coefficient (λ) values for the HDR tone mapping.

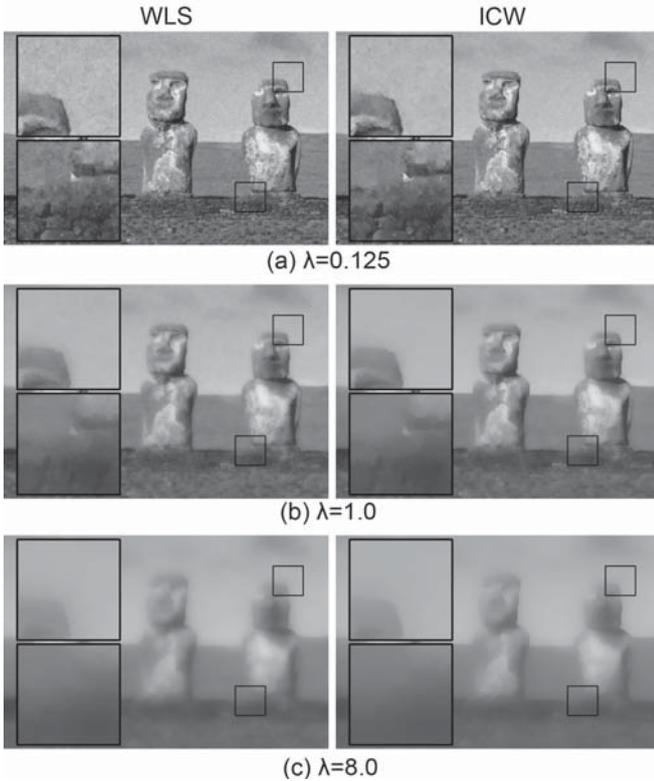


Fig. 2. Comparison of WLS-based (left column) and ICW-based (right column) decompositions with different smoothing coefficient (λ). The input image is the same noise image presented in Fig. 1. The smoothing coefficient corresponds to the global smoothness for coarsening, while the ICW controls the local smoothness for noise reduction. The values of λ is chosen to be the same as the WLS-based tone mapping [17].

Although mathematically the base layer ICW can be represented inside the smoothness term, the real motivation is to control the fidelity of the data term based on HVS. It is different from the smoothing coefficient too, as shown in Fig. 2. The smoothing coefficient controls the overall smoothness of the whole base layer, while the ICW controls the local smoothness for noise reduction.

The ICW-based decomposition reduces the base layer noise effectively compare to the WLS-based decomposition when λ is small. The bigger λ results a smoother base layer and noise are regarded as details and pushed to the detail layer. Thus, in our implementation, the base layer ICW is usually only adopted during the first decomposition where the λ is small. However, when noise is very heavy in the input HDR image, the ICW will be used in more levels of base layer decompositions for better noise reduction.

B. Noise Reduced Detail Layer Decomposition

Base on the Retinex theory [13], the detail layer is the difference between the base layer and the original image luminance, as seen in equation (2). During the subtraction, noise from the input image will be kept in the detail layer. The smoother the base layer, the more details, as well as the more noise, are kept in the detail layer. Thus, the detail layer ICW (ω_d) is proposed to construct a clean detail layer as

$$D = \omega_d(p) \cdot (I - B). \quad (15)$$

The detail layer ICW is defined as the mutual information from a clean image to a noise image. Here, we regard the base layer as a clean reference, and therefore, the mutual information between this clean reference and the original noise image are derived from

$$\omega_d(p) = \frac{1}{\alpha} \log_2 \left(1 + \beta \cdot \left(\frac{\sigma_{B \cdot I}(p)}{\sigma_B^2(p)} \right)^\gamma \right), \quad (16)$$

where $\sigma_B^2(p)$ denotes the local variance of the base layer in a small window centred at p , $\sigma_{B \cdot I}(p)$ denotes the covariance between the base layer and the original luminance, β and γ are two constants control the effectiveness of the weighting function, and α is a normalization factor which fulfills the constraints of $\omega_d \in [0, 1]$. If ω_d is bigger than 1, the detail layer is amplified. It is commonly used in detail enhancement. Meanwhile, it comes with the risk of amplify the noise value.

The proposed denoising factor is a simplified information fidelity criteria from equation (8). Since our clean image, the base layer, is derived from the original luminance, we simply assume the ratio of $\frac{\sigma_B^2}{\sigma_V^2}$ as a constant. The original g_i is adopted as the main contribution of detail layer ICW, which is estimated by linear regression as $g_i = Cov(C, D)Cov(C, C)^{-1}$ [22]. The s_i is obtained by localized sample variance estimation, which has the similar effect as the base layer ICW. The product of g_i and s_i generates a very strict mutual information weighting, which leads to a over smoothed detail layer. In our experiments, we treat s_i as a constant. An example of the proposed detail layer ICW is shown in Fig. 3.

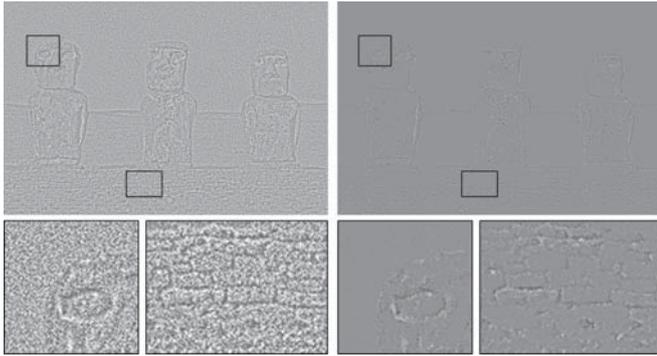


Fig. 3. Detail layer retrieved using (left) WLS-based decomposition, and (right) ICW-based decomposition. The input image is the same noise image presented in Fig. 1.

IV. RESULTS AND COMPARISON

A. Compare with WLS

We implemented the proposed ICW-based noise reduced tone mapping based on the Matlab code of the WLS tone mapping [17] using Farbman et al.'s original implementation downloaded from their website. With the new ICW factors, the processing time is increased by 10-15%. Experimental results on visual comparison of the luminance channel are shown in Fig. 4, where Gaussian noise (SNR=15dB) was added in the clean HDR images. The proposed method improves the quality of WLS tone mapping by 3dB on the average, as shown in Table I, where the reference image is generated from the clean HDR image using original WLS-based method. More tests have been conducted at different noise levels, and the proposed ICW-based tone mapping improves the WLS-based tone mapping method by 1-4dB.

TABLE I
COMPARISON WITH WLS.

Input image	Quality metrics	WLS	ICW
Lamp	PSNR (dB)	30.08	33.26
	SSIM	0.6392	0.9113
Memorial	PSNR (dB)	24.62	27.92
	SSIM	0.5601	0.7649
Leaves	PSNR (dB)	26.43	28.10
	SSIM	0.7120	0.8186
Desk	PSNR (dB)	30.49	34.74
	SSIM	0.6884	0.9398

B. Compare with BM3D

Noise reduction on conventional images has been well studied. Although these method cannot be used directly on the HDR image, they can be adopted on the tone mapping result.

We compared the proposed method with BM3D [27], the state-of-the-art single image denoising method. The BM3D

algorithm is applied on the tone mapping result from the WLS-based method. The visual quality of different noise levels are shown in Fig. 5, and the quality metrics are illustrated in Table II. In some cases, the proposed ICW method is very close to BM3D, especially when the noise level is low. Although in some of our experiments, the ICW outperforms the BM3D, on average, it is 1dB less than the BM3D in noise reduction. The major difference occurs when the noise level is high. The BM3D uses neighbouring blocks to average and smooth out the noise, while the proposed ICW only uses local neighbouring information. As a tradeoff, the proposed ICW executes significantly faster than BM3D using less than 1/5 of the processing time based on our C implementation.

TABLE II
COMPARISON WITH BM3D

Input noise (dB)	Quality metrics	WLS	BM3D	ICW
SNR=10	PSNR (dB)	24.65	29.94	28.08
	SSIM	0.3949	0.8994	0.6869
SNR=15	PSNR (dB)	28.82	32.95	32.60
	SSIM	0.6194	0.9209	0.8913
SNR=20	PSNR (dB)	33.13	36.25	35.47
	SSIM	0.8230	0.9643	0.9456
SNR=25	PSNR (dB)	36.82	38.53	37.88
	SSIM	0.9330	0.9804	0.9710

C. Compare with Other TM Methods

We compared the proposed ICW-based tone mapping method with other 5 most representative tone mapping algorithms, a global tone mapping operator [9], a subbands-based scale decomposition [21], a bilateral-filtering-based decomposition [16], a direct luminance compression [11], and the edge-preserving WLS [17]. Except the global tone mapping operator, which is implemented in an open source project Luminance HDR, the implementation of the other methods are provided by their authors.

The input HDR image of Fig. 6 is a clean HDR image added with Gaussian white noise, while the input HDR image of Fig. 7 is an HDR image synthesized from a sequence of differently exposed images captured using normal camera. Different tone mapping algorithms give different visual experiences, which is very subjective in terms of right or wrong. However, the noise from the HDR image is kept in the tone mapping results from all the 5 tone mapping algorithms. Compare to them, the proposed ICW generates a clean display-able image.

Adapting ICW on color HDR image is different from other tone mapping methods, where usually only the luminance channel is calculated, and R, G, B channels are derived proportionally to the luminance channel. In our implementation, due to lack of HDR color model, all R, G, B color channels are processed separately, which may cause color shift in some area.

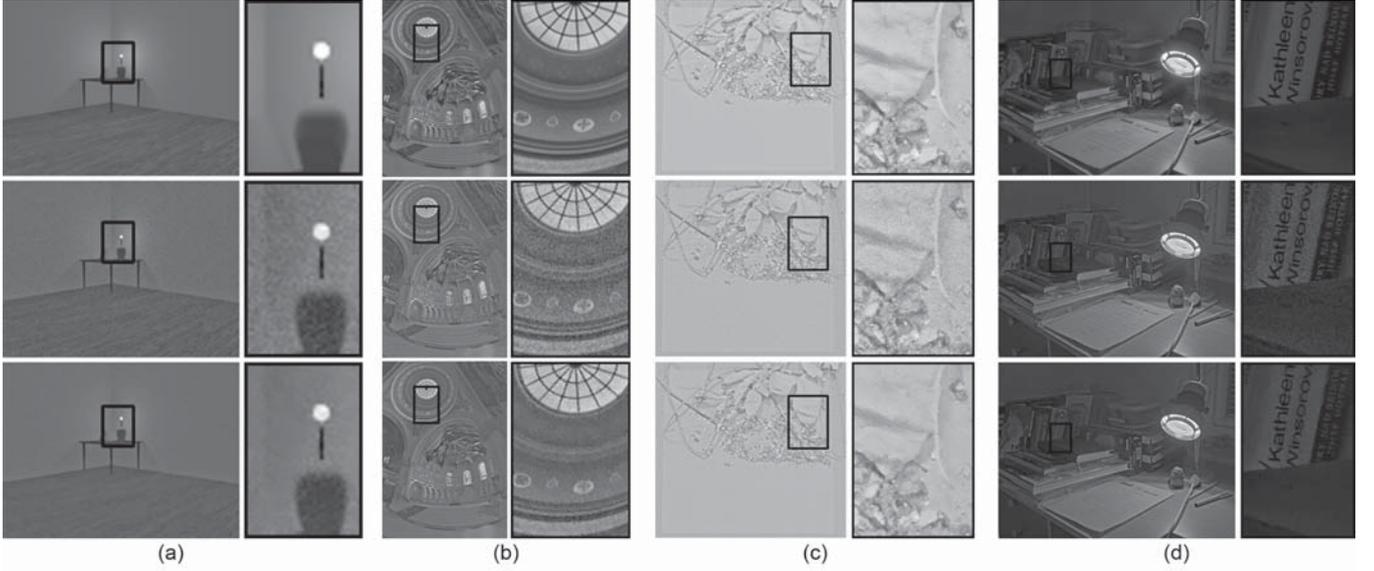


Fig. 4. Visual comparison of the luminance component generated using WLS (mid row) and ICW (bottom row). The clean image is specified in the top row

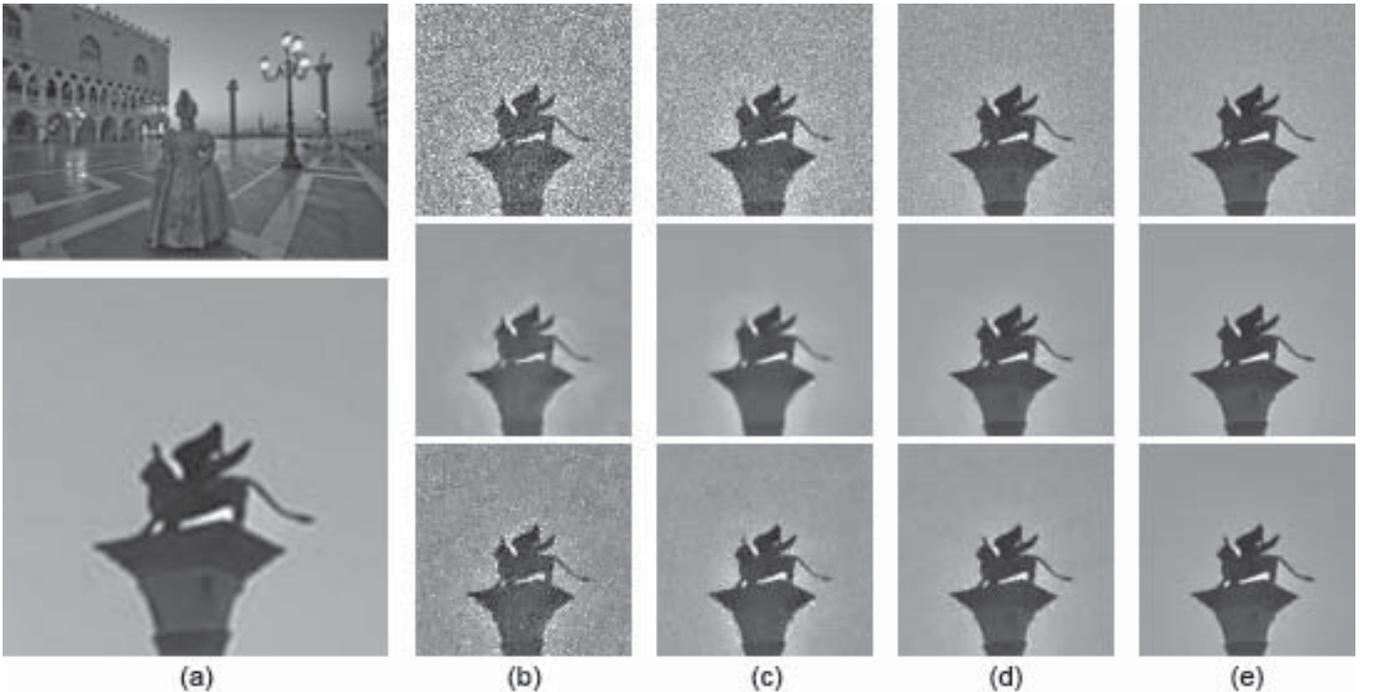


Fig. 5. Visual comparison with WLS+BM3D. (a) Clean image and a zoom-in object. The input image of (b-e) are added with different noise level at (b) SNR=10dB, (c) SNR=15dB, (d) SNR=20dB, and (e) SNR=25dB. In (b-e), the (top row) is WLS, the (middle row) is WLS+BM3D, and the (bottom row) is the proposed ICW.

D. Single Image Noise Reduction

Although the proposed ICW is designed to be used for noise reduced HDR tone mapping, it can also be used directly on the conventional image. Fig. 8 shows an example, in which BM3D outperforms the proposed ICW by 1.7dB. When BM3D is applied on the tone mapping result from WLS, the image has been compressed. Compare with it, the ICW works directly on the HDR image, which access to more information. Therefore, the ICW is close to BM3D. However, when purely on a single

image. The only advantage of ICW is the speed.

V. CONCLUSION

In this paper, we presented a noise reduced tone mapping method based on information content weighting. The proposed method works in two steps and reduces the noise on both base layer and detail layer. The experiments show that the proposed method effectively reduces the noise compared to the other state-of-the-art tone mapping algorithms. Additionally, the processing time of the proposed ICW is very fast.

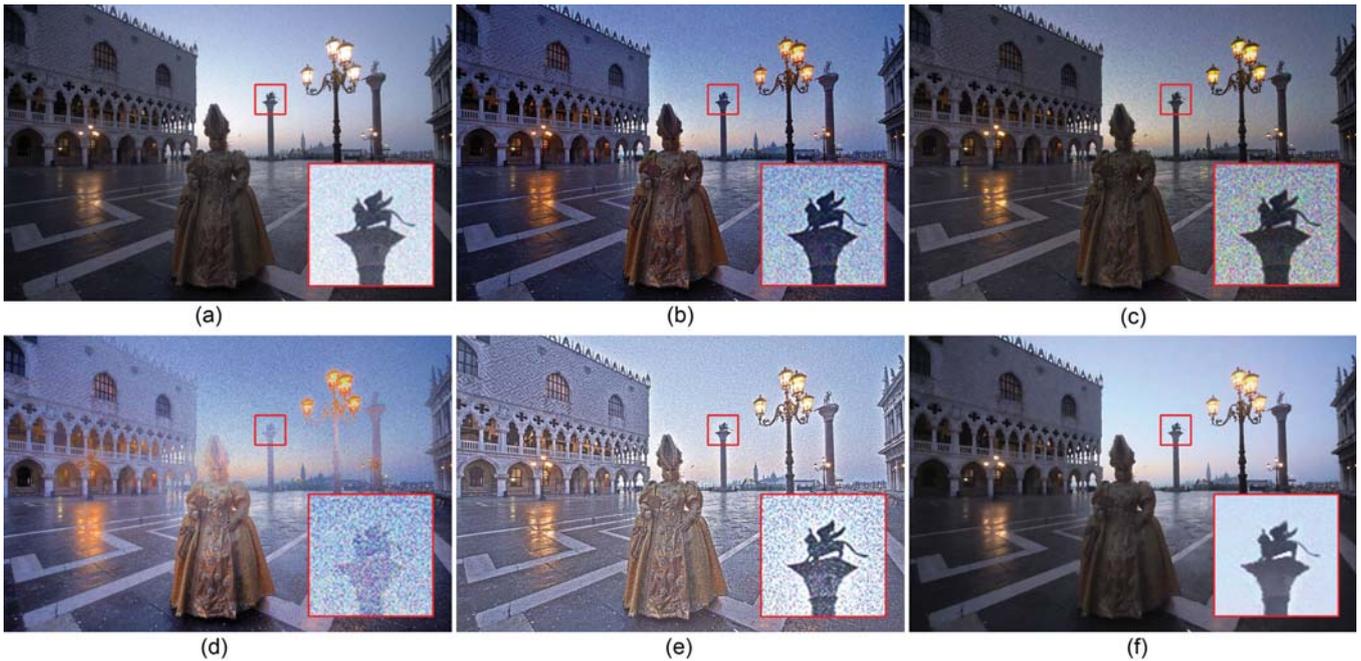


Fig. 6. Visual comparison of different tone mapping algorithms: (a) global tone curve [9]; (b) scale decomposition [21]; (c) bilateral filtering decomposition [16]; (d) direct luminance compression [11]; (e) WLS [17]; and (f) proposed ICW

We are continuing efforts into improving the denoising method on color HDR images by studying color space that is suitable for HDR images.

REFERENCES

- [1] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000, pp. 472–479.
- [2] G. Wetzstein, I. Ihrke, and W. Heidrich, "Sensor saturation in fourier multiplexed imaging," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 545–552.
- [3] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '97, 1997, pp. 369–378.
- [4] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 760–768, Aug. 2004.
- [5] R. Hoskinson, B. Stoeber, W. Heidrich, and S. Fels, "Light reallocation for high contrast projection using an analog micromirror array," *ACM Trans. Graph.*, vol. 29, no. 6, pp. 165:1–165:10, Dec. 2010.
- [6] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 70–79, 2013.
- [7] G. Ward, "A contrast-based scalefactor for luminance display," in *Graphics gems IV*, 1994, pp. 415–421.
- [8] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.
- [9] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 68:1–68:10, Aug. 2008.
- [10] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.
- [11] Q. Shan, J. Jia, and M. Brown, "Globally optimized linear windowed tone mapping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 4, pp. 663–675, 2010.
- [12] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 249–256, Jul. 2002.
- [13] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 1–11, Jan 1971.
- [14] D. Jobson, Z.-u. Rahman, and G. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, 1997.
- [15] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, Jul. 2002.
- [16] J. Kuang, G. M. Johnson, and M. D. Fairchild, "icam06: A refined image appearance model for hdr image rendering," *J. Vis. Commun. Image Represent.*, vol. 18, no. 5, pp. 406–414, Oct. 2007.
- [17] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67:1–67:10, Aug. 2008.
- [18] G. Guarnieri, S. Marsi, and G. Ramponi, "High dynamic range image display with halo and clipping prevention," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1351–1362, 2011.
- [19] Z. Li, S. Rahardja, S. Yao, J. Zheng, and W. Yao, "High dynamic range compression by half quadratic regularization," in *IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 3169–3172.
- [20] J. W. Lee, R.-H. Park, and S. Chang, "Noise reduction and adaptive contrast enhancement for local tone mapping," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 578–586, 2012.
- [21] Y. Li, L. Sharan, and E. H. Adelson, "Compressing and companding high dynamic range images with subband architectures," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 836–844, Jul. 2005.
- [22] H. Sheikh, A. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [23] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, Jan. 2003.
- [24] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [25] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [26] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *IEEE International Conference on Image Processing*, 2006, pp. 2945–2948.
- [27] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by

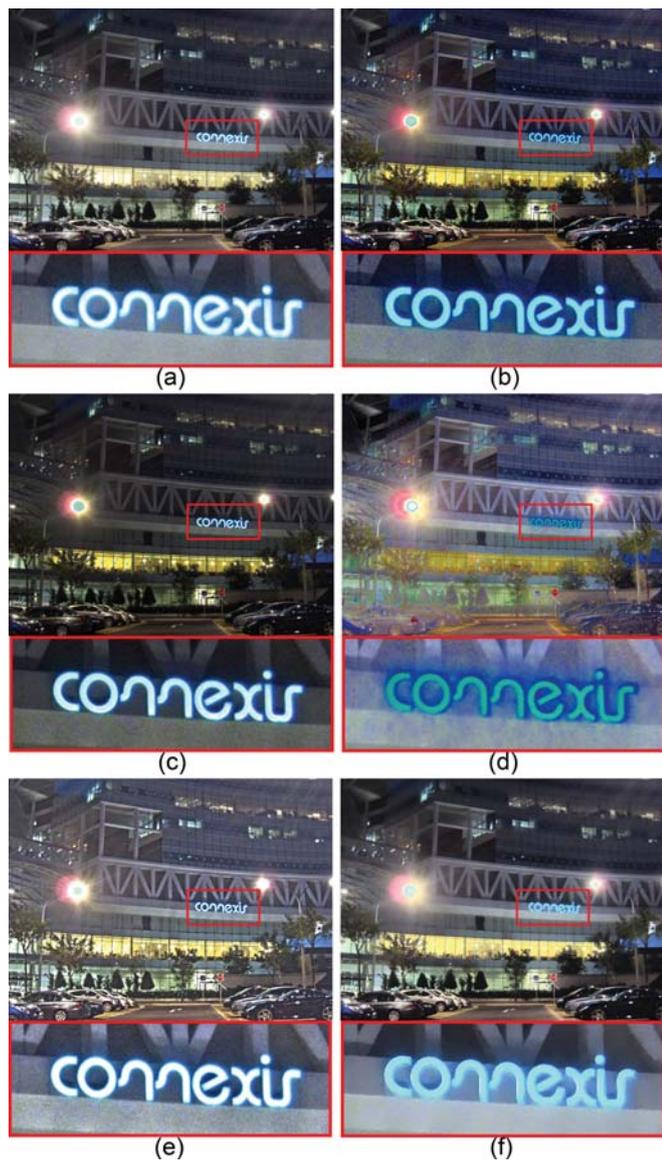
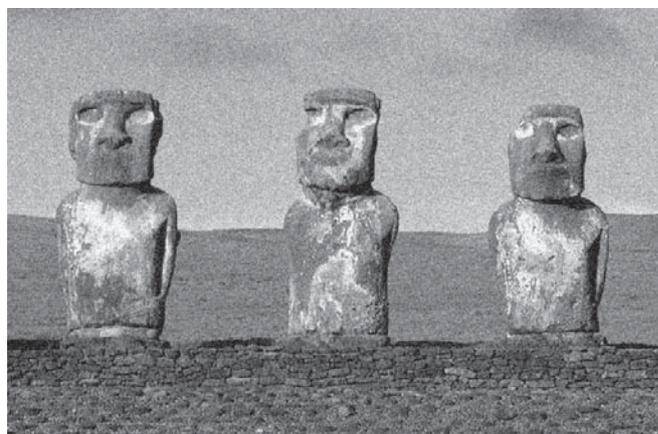


Fig. 7. Connexis, (a) global tone curve [9]; (b) scale decomposition [21]; (c) bilateral filtering decomposition [16]; (d) direct luminance compression [11]; (e) WLS [17]; and (f) proposed ICW

sparse 3-d transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.



(a)



(b)



(c)

Fig. 8. Single image noise reduction. (a) input noise image (Gaussian noise with variance of 0.01); (b) denoising using BM3D (PSNR=28.30dB); and (c) denoising using ICW (PSNR=26.59dB)

Paper P8

Z. J. Zhu, Z. G. Li, S. Rahardja, and P. Fränti, "2D Denoising Factor for High Dynamic Range Imaging," *ACM SIGGRAPH*, Poster, Los Angeles, US, Aug 2012.

2D Denoising Factor for High Dynamic Range Imaging

Zijian Zhu, Zhengguo Li, and Susanto Rahardja*
Institute for Infocomm Research

Pasi Fränti†
University of Eastern Finland



Figure 1: HDR image generated by using (a) Debevec and Malik 1997, (b) Yao et al. 2010, (c) Akyuz and Reinhard 2007, and (d) the proposed method which not only generates better result than method (c) but also uses only 1/5 of its processing time.

1 Introduction

The presence of noise in a high dynamic range (HDR) synthesis poses a serious degradation to the HDR image especially when the input images are captured at low light condition or with high sensitivity settings. Thus, a two-dimensional (2D) denoising factor is proposed to assign higher weight to a pixel with less noise based on both pixel luminance and image exposure. This pure temporal denoising factor is controlled by two key coefficients and can preserve edge and fine detail without blurring artifact. In addition, both memory and computation time are significantly reduced compare to other denoising methods.

2 Our Approach

The first dimension of the proposed denoising factor assigns a high weight to a pixel with a large luminance. Given an assumption that the noise distribution is independent of the measurement pixel value z , [Mitsunaga and Nayar 1999] argued that a luminance based weighting function (MN weight) of $\omega = f(z)/f'(z)$ will achieve the best signal to noise ratio, where $f(z)$ is the radiometric response function. [Akyuz and Reinhard 2007] modified the MN weight by replacing pixel value with luminance value. A broad hat function $h(z) = 1 - (2z/255 - 1)^{12}$ was used to restrict the saturated pixels which may cause color cast. Since the radiometric response function is usually monotonic increasing, we approximate the luminance based weight by a controllable hat function and a Hermite interpolation. Thus, we can significantly reduce processing time on response function recovery by defining a new weighting factor as

$$\omega(z) = \begin{cases} 1 - |\frac{z}{\beta} - 1|^\alpha, & 0 \leq z < \beta \\ 1 - 3(\frac{255-z}{255-\beta})^2 + 2(\frac{255-z}{255-\beta})^3, & \beta \leq z < 255 \end{cases}, \quad (1)$$

where two key coefficients are the denoising strength coefficient α and the saturation control coefficient β . It can be seen the smaller the value of α , the hat function will be steeper and this will result in better denoising effect. A large α gives high weights to small value (luminance) pixels, which keeps noise in the synthesized HDR image. We choose $\alpha = 2$ in the experiments. And test also shows that $\alpha = 12$ generates similar result as [Akyuz and Reinhard 2007]. The saturation control coefficient limits the near saturated pixels to avoid color cast due to gamut limitations (an empirical value $\beta = 200$ is used here).

The second dimension of the proposed denoising factor is based on exposure time. More photons reach the camera sensor with a longer exposure time (Δt), which results in a more accurate reading. Thus, the proposed 2D weighting factor is designed to multiply the geometrically normalized exposure times with the luminance based

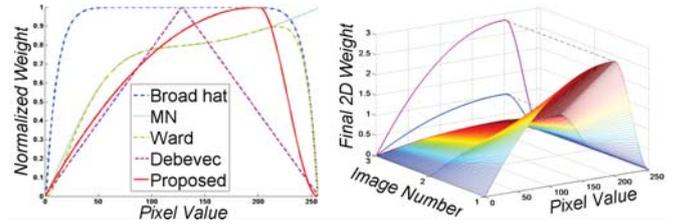


Figure 2: (Left) The proposed luminance based weighting compares to other weighting functions; (Right) An example of the proposed 2D denoising factor with 3 input images.

denoising factor as

$$W(z, \Delta t_j) = \sqrt{\Delta t_j / \sqrt{P \prod_{p=1}^P \Delta t_p}} \cdot \omega(z), \quad (2)$$

where j denotes the j th image in the total P input images. The geometrical normalization avoids overwhelming big weights caused by some very large exposure time. Then, the objective function,

$$O = \sum_{i=1}^N \sum_{j=1}^P \{W(z, \Delta t_j)[Z_{ij} - \ln E_i - \ln \Delta t_j]\}^2 + \lambda \sum_{z=Z_{max}-1}^{Z_{min}+1} [W(z, \max(\Delta t_1, \dots, \Delta t_P))g''(z)]^2, \quad (3)$$

is used to calculate camera response function (g) and synthesize the clean HDR image.

The proposed denoising method is verified by comparing it with three HDR synthesis methods. The noise is significantly reduced as compared to [Debevec and Malik 1997] with the same processing time. No blur artifact is generated as [Yao et al. 2010] and they are achieved due to spatial averaging. Comparing with [Akyuz and Reinhard 2007], the quality of the proposed is about the same. However, the proposed is achieved with only 1/5 of the processing time, since no intermediate steps for approximating response function is required.

References

- DEBEVEC, P. E. AND MALIK, J. 1997. Rendering high dynamic range radiance maps from photographs. In *Proceedings of SIGGRAPH 1997*, 369–378.
- MITSunAGA T. AND NAYAR S.K. 1999. Radiometric Self Calibration. *IEEE Conf. on Computer Vision and Pattern Recognition 1999*, 374–380.
- AKYUZ A.O. AND REINHARD E. 2007. Noise Reduction in High Dynamic Range Imaging. *Journal of Visual Communication and Image Representation 2007*, 18(5), 366–376.
- YAO W., LI Z.G., RAHARDJA S., YAO S.S., AND ZHENG J.H. 2010. Half-quadratic Regularization Based De-noising for High Dynamic Range Image Synthesis. *IEEE Int. Conf. on Acoustics Speech and Signal Processing 2010*, 18(5), 1370–1373.

*e-mail: {zhuzj, ezgli, rsusanto}@i2r.a-star.edu.sg

†e-mail: franti@cs.joensuu.fi