HIGH DYNAMIC RANGE CONTENT ACQUISITION FROM MULTIPLE EXPOSURES

Raissel Ramírez Orozco

Universitat de Girona

PhD THESIS

High Dynamic Range Content Acquisition from Multiple Exposures

Raissel Ramirez Orozco

2015

PhD THESIS

High Dynamic Range Content Acquisition from Multiple Exposures

Author:

RAISSEL RAMIREZ OROZCO

2015

DOCTORAL PROGRAMME IN TECHNOLOGY

Supervisors:

PHD. IGNACIO MARTIN
PHD. CELINE LOSCOS
PHD. ALESSANDRO ARTUSI

Presented in partial fulfilment of the requirements for a doctoral degree from the University of Girona

Dedication and acknowledgements

Flashes of my life come to my mind at this very moment, from a barefoot boy playing in the dusty streets of a suburb in the outskirts Camagüey to a man presenting his PhD in a European university; undoubtedly thanks must be offered to the many people who have made this possible.

First and foremost, I would like to express my gratitude to my supervisors. I am sure that our friendship will last much longer than this thesis, and I am proud of having you as close friends and professors.

Sincere acknowledgments to my country, Cuba. To all my teachers and professors there, from primary school to university. I must thank to the University of Girona and specially to the Group of Graphics and Geometry, for the nice environment of cordiality. I appreciate the support of the people in the CReSTIC group at the Universite de Reims Champagne-Ardenne. I am glad to have participated in the EU COST HDRi IC1005 project that gave me the option to engage with meetings and training schools, enriching my experience and broadening my circle of friends.

Thanks to my student colleagues both in Girona and Reims, for the great discussions and the nice times we spent together.

I also wish to thank to the reviewers and the members of the thesis jury, for the valuable comments that have enriched this thesis. Thanks to all the people that in a way or another have made this work possible.

The deepest thanks to my family and friends for trusting me, for their unconditional support and comprehension, for shaping the man I am. Thanks Samu, for always backing me, and for the corrections and the diagrams too ;-)

To the memory of my grandfather.
To my parents, grandmas, and cousins, for being my endless source of motivation.
To my dearest love, for complementing my life in the best of the possible ways.

Declaration

PhD. Ignacio Martin
Department of Computer Science, Applied Mathematics and Statistics (IMAE), Universitat de Girona.

PhD. Celine Loscos
CReSTIC Research Centre of the Universitè de Reims Champagne-Ardenne.

PhD. Alessandro Artusi
Department of Computer Science, Applied Mathematics and Statistics (IMAE), Universitat de Girona.

CERTIFIQUEM:

Que aquest treball, titulat "High DynamicRange Content Acquisition from Multiple Exposures", que presenta en Raissel Ramirez Orozco per a l'obtenció del títol de doctor, ha estat realitzat sota la nostra direcció i que compleix els requeriments per poder optar a Menció Internacional.

SIGNATURES:

PhD. Ignacio Martin PhD. Celine Loscos PhD. Alessandro Artusi

Girona, 10 November 2015.

List of Figures

Figure **Page**

List of Tables

Abstract

The limited dynamic range of digital images can be extended by composing images of the same scene with different exposures to produce high-dynamic-range (HDR) images. This is a standard procedure for static scenes but a challenging task for dynamic ones. Pixels of the different exposures need to be perfectly aligned before being combined into an HDR values free of artifacts. This thesis is composed of an overview of the state of the art techniques and three different methods to tackle the image alignment and deghosting problems in the HDR imaging domain.

The first method is focused on HDR image acquisition of dynamic scenes from static cameras. It detects the areas affected by motion, registers the dynamic objects over a reference image, and combines low-dynamic range (LDR) values to recover HDR values in the whole image. The motion detection method generates a ghost mask that contains pixels affected by motion in the sequence. Such pixels are selected and registered to a reference image. Once matches are found, the assembling step guarantees that all aligned pixels will contribute to the final result which enlarges the dynamic range of affected areas. Unlike previous works, our solution includes the maximal amount of information available in the sequence. The results of testing this method on several scenes are robust for cases where the dynamic objects are globally rigid.

3D HDR imaging also requires matching pixels from differently exposed images. Our second approach builds multiscopic HDR images from LDR multi-exposure images. The method is based on a patch match algorithm which was adapted and improved to take advantage of epipolar geometry constraints of stereo images. To our knowledge, it is the first time that an approach different than traditional stereo matching has been used to obtain accurate matching between the stereo images. Experimental results show accurate registration and HDR generation for each LDR view.

It is challenging to find matches inside large under/over exposed areas. We introduce the new concept of in-HDR-painting which aims to recover valid color values in such regions. We propose to replace under/over exposed pixels in the reference image by using valid HDR values from other images in the multi-exposure LDR image sequence. The algorithm is fully automatic and is based on the assumption that the scene might be dynamic and that images are not aligned. The algorithm first detects the target areas and classifies them as under-exposed or over-exposed areas. Search for matching pixels in other images starts on target area contours, imposing content-based, and geometrical constraints. The best matches are selected using a color-based criterion. Using the selected matches, an inpainting interpolation technique reconstructs the missing information in the target areas. The results show that in-HDR-painting can reconstruct HDR images even when the lowest or highest exposures are used as reference images and they contain large under/over-exposed areas.

Resumen

El limitado rango dinámico de las imágenes digitales puede ampliarse mezclando varias imágenes adquiridas con diferentes valores de exposición. Este es un procedimiento estándar para escenas estáticas pero complejo en escenas dinámicas. Los píxeles de las diferentes imágenes deben estar perfectamente alineados para combinar las diferentes exposiciones sin introducir errores. Esta tesis incluye un detallado resumen del estado del arte y tres métodos diferentes para alinear las imágenes y corregir el efecto 'ghosting' en imágenes HDR.

El primer método propone recomponer imágenes HDR de escenas dinámicas adquiridas con una cámara estática. Está centrado en detectar las áreas afectadas por el movimiento y registrar los objetos dinámicos sobre una imagen de referencia de modo que se logre recuperar información a lo largo de toda la imagen. Los métodos de detección de movimiento generan una máscara que contiene los píxeles de objetos en movimiento. Estos píxeles son seleccionados y registrados sobre la imagen de referencia. Una vez encontradas las correspondencias, nuestro método garantiza que todos los píxeles alineados contribuyan al resultado final. A diferencia de los trabajos anteriores, esta solución incluye la máxima información disponible en la secuencia de imágenes. Los resultados de probar este método en diversas escenas son prometedores en casos donde los objetos dinámicos son aproximadamente rígidos.

Las imágenes 3D HDR tambien requieren encontrar correspondencia entre píxeles de imagenes con exposición diferente. Nuestra segunda propuesta es un método para obtener imágenes HDR multiscópicas a partir de diferentes exposiciones LDR. Está basado en un algoritmo de 'patch match' que ha sido adaptado para aprovechar las ventajas de las restricciones de la geometría epipolar de imágenes estéreo. Hasta donde conocemos, es la primera vez que se utiliza un enfoque diferente a la tradicional búsqueda de correspondencias estéreo para este propósito. Los resultados experimentales muestran que el registro y la generación de las imágenes HDR correspondientes a cada vista son adecuados.

Resulta complejo encontrar correspondencias en áreas con sobre/baja exposición. Esta tesis presenta el concepto de 'in-HDR-painting' que intenta recuperar valores adecuados para estas regiones. Proponemos reemplazar dichos píxeles en la imagen de referencia usando valores correctos de otras imágenes de la secuencia. El algoritmo es completamente automático y asume las escenas son dinámicas y las imágenes no están alineadas. Primero detecta las zonas a tratar y las clasifica en saturadas o oscuras. Se buscan correspondencias para los puntos en el contorno de dichas zonas imponiendo restricciones geométricas y se seleccionan las mejores correspondencias. Un proceso de interpolación usa dichas correspondencias para reconstruir la información en las zonas afectadas. Los resultados muestran que este método puede reconstruir imágenes HDR incluso usando como referencia la imagen con menor o mayor valor de exposición en la secuencia, con amplias zonas oscuras o saturadas.

Resum

El limitat rang dinàmic de les imatges digitals pot ampliar-se barrejant diverses imatges adquirides amb diferents valors d'exposició. Aquest és un procediment estàndard per a escenes estàtiques però complex per a escenes dinàmiques. Els píxels de les diferents imatges han d'estar perfectament alineats per combinar les diferents exposicions sense introduir errors. Aquesta tesi inclou un detallat resum de l'estat de l'art i tres mètodes diferents per alinear les imatges i corregir l'efecte 'ghosting' en el domini de les imatges HDR.

El primer mètode proposa recompondre imatges HDR d'escenes dinàmiques adquirides amb una càmera estàtica. Està centrat en detectar les àrees afectades pel moviment i registrar els objectes dinàmics sobre una imatge de referència de manera que s'aconsegueixi recuperar informació al llarg de tota la imatge. Els mètodes de detecció de moviment generen una màscara que conté els píxels dels objectes en moviment. Aquests píxels són seleccionats i registrats sobre la imatge de referència. Una vegada trobades les correspondències, el nostre mètode garanteix que tots els píxels alineats contribueixin al resultat final. A diferència dels treballs anteriors, aquesta solució inclou la màxima informació disponible en la seqüència d'imatges. Els resultats de provar aquest mètode en diverses escenes són prometedors en casos on els objectes dinàmics són aproximadament rígids.

Les imatges 3D HDR tambien requereixen trobar correspondència entre píxels de imatges amb exposició diferent. La nostra segona proposta és un mètode per obtenir imatges HDR multiscópicas a partir de diferents exposicions LDR. Està basat en un algorisme de 'patch match' que ha estat adaptat per aprofitar els avantatges de les restriccions de la geometria epipolar d'imatges estèreo. Fins a on coneixem, és la primera vegada que s'utilitza un enfocament diferent a la tradicional cerca de correspondències estèreo per a aquest propòsit. Els resultats experimentals mostren que el registre i la generació de les imatges HDR corresponents a cada vista són adequats.

Resulta complex trobar correspondències en àrees amb sobre/baixa exposició. Aquesta tesi presenta el concepte de 'in-HDR-painting' que intenta recuperar valors adequats per a aquestes regions. Proposem reemplaçar aquests píxels en la imatge de referència usant valors correctes d'altres imatges de la seqüència. L'algorisme és completament automàtic i assumeix les escenes són dinàmiques i les imatges no estan alineades. Primer detecta les zones a tractar i les classifica en saturades o fosques. Es busquen correspondències per als punts en el contorn d'aquestes zones imposant restriccions geomètriques i se seleccionen les millors correspondències. Un procés d'interpolació usa aquestes correspondències per reconstruir la informació a les zones afectades. Els resultats mostren que aquest mètode pot reconstruir imatges HDR fins i tot usant com a referència la imatge amb menor o major valor d'exposició en la seqüència, amb àmplies zones fosques o saturades.

Table of Contents

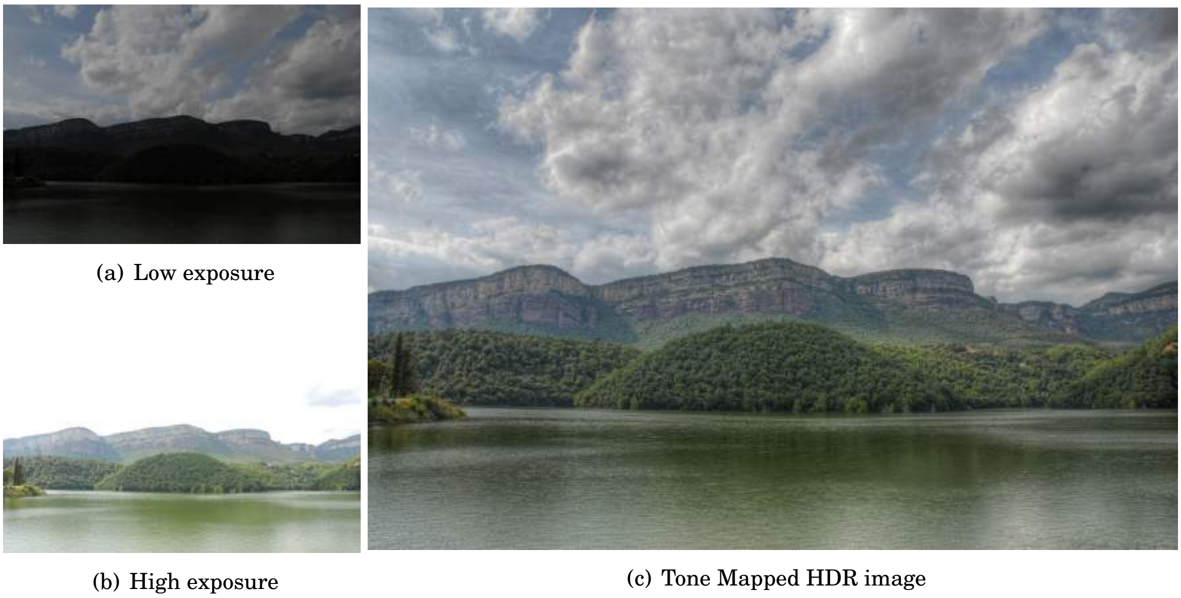
CHAPTER 1

INTRODUCTION

In digital photography, one of the main limitations for reproducing real world appearances relies on the constrained luminance and contrast ranges that are captured by most digital cameras, stored by the majority of image and video formats and reproduced in display devices [RKMS15]. There is a huge gap between the range of light that the human visual system (HVS) perceives and what common digital cameras and displays are able to capture and visualize respectively. The human eye can see objects both in a dark night and in a sunny day, despite of the luminance level of the sunlight is about $10^5 cd/m^2$ while the stars light is about $10^{-3} cd/m^2$. This means that the HVS is capable of adapting to a large variation of lighting in a range of nearly 10 orders of magnitude and about 5 orders of magnitude within the same scene [RWD⁺10]. In contrast, most user-level displays show images within a luminance range of approximately 1:300 cd/m2 [SHS⁺04] and most digital cameras produce images in a range lower than 1:1000 [JLW08].

Dynamic range in digital images can be defined as the ratio between the darkest and the brightest points captured from a scene. It can be expressed in orders of magnitude (powers of ten), in stops (powers of two) or in decibels (db). High Dynamic Range (HDR) imaging aims to increase the dynamic range recorded in a digital image from a given scene. Pixels in an HDR image are proportional to the radiance of the scene, dark and bright areas can be recorded within the same image. Visually, this means avoiding under and over exposure in such areas. The vast majority of digital images are still Low Dynamic Range (LDR), stored usually in 24 bits per pixel (8 bits per color channel in RGB) which represents approximately 2 orders of magnitude while HDR images are represented using floating point formats. Since common displays or printers represents only 8 bits per color channel, images need to be adapted (*Tonemap*) to 8 bits per color channel to display or print them in LDR devices. Figure 1.1(c) shows a tone mapped example of HDR image, notice that both bright and dark parts of the scene are properly exposed.

Common digital camera sensors are physically limited for capturing the full illumination range from nature. Finding the appropriate exposure value is challenging, especially in scenes with large dark and bright areas. For example, taking a picture on a sunny day implies in practice deciding whether to appropriately expose the bright sky or the details in the shadow, like in Figure 1.1(a) and 1.1(b). Most digital cameras provide an auto-exposure algorithm to set the ISO value, aperture and shutter speed for capturing the best exposure for a given scene. However, when the amount of energy reaching the sensor exceeds the maximum allowed value, details in bright areas are clamped to the maximum allowed value which is white(known as over-exposure or saturation). On the other hand, if not enough energy reaches the sensor under-exposure takes place (the opposite of saturation).


(a) Low exposure

(b) High exposure

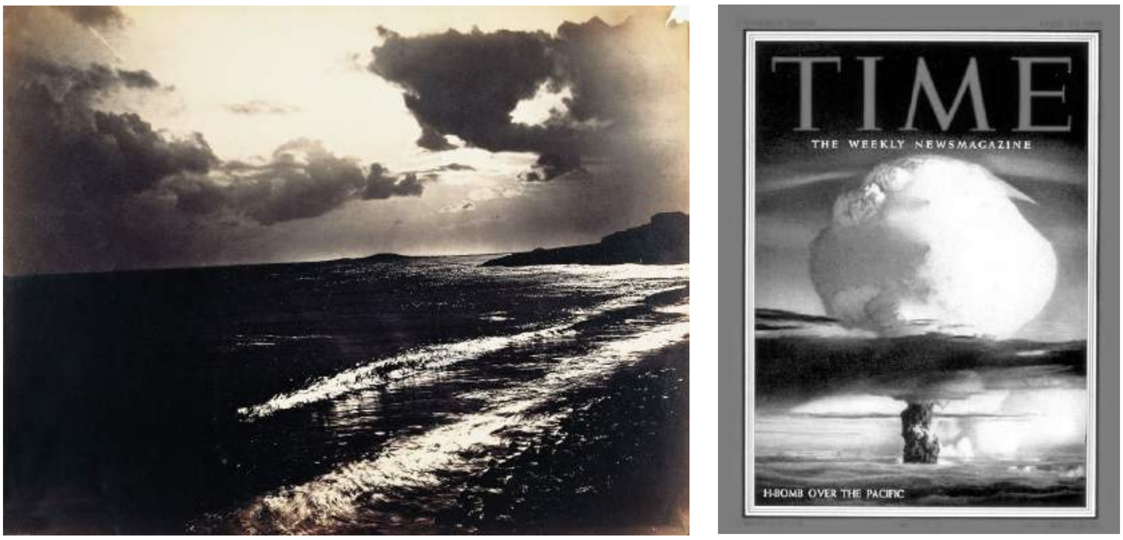
(c) Tone Mapped HDR image

Figure 1.1: Scene represented using two different LDR exposures and a tone mapped version of the HDR image.

The idea of solving these problems by enlarging the range of values represented in one image is not recent. It was pioneered by Gustave Le Gray back in 1857. In a picture of the sea (Figure 1.2(a)), he captured the extreme luminance difference between the sky and the sea by combining two negatives into a single positive print that showed details in both areas. However, the term $'HDR'$ was first cited in the 1940s by Charles Wyckoff, who implemented a local neighborhood tone mapper to combine differently exposed film layers into one single image of wider dynamic range [Cer06]. In May of 1954 Wyckoff published a picture of a nuclear explosion that was the result of combining different exposures (Figure 1.2(b)).

The combination of latest advances in digital imaging such as 4K image color resolution, 3D stereoscopic, and HDR imaging, promise an unprecedented experience for users. However, big challenges of different nature are still to be overcome before such technologies converge. In particular, there are unsolved limitations in each steps of the HDR imaging pipeline (acquisition, compression, transmission and display). Solutions are required before we can enjoy 3D HDR content on a TV at home. Among such challenges, the capture of dynamic scenes and the extension from static HDR images to HDR video and stereoscopic HDR plays a very important role.

Techniques for HDR acquisition have been a hot research topic in recent years. There are three main approaches for creating HDR content: Computer Graphics (CG) synthesized images, native HDR sensors and LDR multiple exposure combination. This work is focused on the third approach, HDR acquisition using conventional digital cameras. Multiple differently exposed LDR images can be merged to recover HDR values [MP95, DM97, MN99]. Each exposure covers a different range of light. Short exposures



(a) Le Gray, Mediterranean Sea-1857

(b) Wyckoff, HBomb-1954

Figure 1.2: Examples of dynamic range enlargement in film photography.

provide details in the brighter parts of the scene because the shutter speed is fast enough minimizing over-exposure. Meanwhile, long exposures allow to capture light coming from the darker parts of the scene.

The Automatic Exposure Bracketing function available in many digital cameras facilitates the acquisition of different exposures of the same scene consecutively. The camera automatically calculates the best aperture/ISO/shutter speed combination for the lighting conditions that minimizes under and over-exposure. Once the parameters are set for the best exposure, all values are kept constant except the shutter speed. The camera captures images consecutively varying the shutter speed to acquire different ranges of light from the scene.

1.1 Problem statement and thesis goals

HDR image acquisition of static scenes with a fixed camera is considered a standard procedure nowadays. However, dealing with dynamic scenes or cameras is still challenging. The acquisition of a multiple exposure sequence takes at least the sum of the shutter speed of each shot. If there are dynamic objects in the scene or the camera moves during the acquisition, the pixels in the sequence of images will be misaligned. Merging non-aligned exposures produce artifacts similar to the *'ghosting'* effect of large exposure times in traditional photography.

Countless image alignment or *'deghosting'* algorithms were proposed in recent years [War03, Gro06, TM07, PH08, JLW08, GSL08, SPS09, PK10], but even the best existing solutions are slow [Bog00, KAR06, GGC^{+}09, GKTT13, HGPS13], reference dependent, and might fail under highly dynamic range scenes [KUWS03, MG10, RKC09, RC11, HLL^{+}11, HGP12, SKY^{+}12, KSB^{+}13].

Although alignment and deghosting are often used as synonyms in HDR literature, in this thesis the term *alignment* is used to name methods focused on correcting global misalignment caused by camera movement and the term *deghosting* for methods dealing with local artifacts caused by dynamic objects in the scene.

This thesis presents a thorough state-of-the-art report on techniques for multiple exposures alignment and deghosting. Moreover, three different approaches to produce ghosting-free HDR content under different conditions were developed. We propose reference-independent solutions for the following multiple exposure acquisition setups:

- Dynamic scenes from a static camera.
- Multiscopic exposure sequences.
- Dynamic scenes acquired with a free camera.

1.2 HDR content acquisition from multiple exposures

The combination of multiple exposures implies dealing with images acquired in different moments, in some cases also from a different viewpoint and representing a changing scene. Therefore, some temporal and spatial aspects must be considered to tackle the problem.

1.2.1 Temporal considerations

There is a debate in the HDR community about how many exposures and which exposure values should be used to obtain a good HDR image depending on the characteristics of the scene [GN03b]. This is not a problem for static scenes, because in such case it is possible to increase or reduce the exposure time to capture a wider range of the available light. In cases where either the camera or the scene moves during the acquisition, using long exposure times might introduce large misalignment in the scene. None of the existing techniques for alignment and deghosting in HDR are robust under large misalignment [OTTE15].

Timing restrictions are a very important issue in HDR video acquisition from multiple exposures. The typical frame rate to play video is around 24 to 30 fps; it means that we need to create at least 24 HDR frames per second. Unlike for still HDR images, we cannot just capture longer exposures trying to increase the range of light captured. The sum of the exposure times for each frame must be small enough to guarantee video frame rates. The difference in exposure times between consecutive frames is limited for temporal coherence. Such limitations must be considered before extending methods designed for image alignment. Some of them are suitable to generate plausible results only when using as reference the best exposed images in each multi-exposed sequence [SKY$^+$12, HGPS13, GKTT13]. In the video context, the use of long exposure times to increase dynamic range is not always possible without compromising video frame rates.

Besides, stereo HDR requires at least two views of the same scene, while some multiscopic displays may accept more than 9 different views. To achieve stereo HDR, one HDR image per view is needed. Depending on the characteristics of the system, four different solutions can be considered:

1. Every view has the same exposure time at each shot [Ruf11]. The exposure value changes from one shot to the next. While synchronization is here simplified, it is difficult to reconstruct temporarily coherent HDR information in over- and under-exposed areas.
2. Different views have different exposure times [TKS06, LC09, SMW10, BRG$^+$14]. While finding HDR information in all areas of the images is ensured, synchronization is problematic: either each objective waits for the others before taking the next frame or frames need to be synchronized afterward.

3. Bonnard *et al.* [BLV$^+$12] propose an alternative to the first two solutions. It consists of placing neutral density filters on the camera objectives in order to simulate different exposure times. An advantage to this is that all objectives use the same exposure time even if the resulting images simulate different exposures. Synchronization is thus reduced to synchronizing the different objectives. A major drawback comes with the fact that each view takes the same exposure at each frame. Under- or over-exposed areas might remain as such through the entire video.

4. Acquiring several exposures at once for each objective is also an option. It could be done through a beam splitter [TKTS11] or a spatially varying mask [NB03]. Optical elements can split light beams onto different sensors with different exposure settings [TKTS11].

1.2.2 Spatial considerations

While taking LDR multiple exposures for HDR reconstruction, spatial variations are likely to occur. In the following table we classify those variations according to their origin and the type of misalignment they produce (sect 1.2.2.1). Types of misalignment are discussed in section 1.2.2.2. Finally, we review the different approaches specifically designed to manage misalignment of multi-exposed LDR images for HDR generation (see section 1.2.3).

1.2.2.1 Camera *vs* scene movement

Misalignment can be classified into different categories according to the kind of movement of the camera and the objects in the scene. Table 1.2.2.1 shows the types of possible misalignments present in multiple exposures for HDR generation.

Table 1.1: Different configurations of camera and scene

	Camera	Scene	Misalignment	HDR Video
1	Static	Static	-	Time-lapse
2	Static	Dynamic	Local	Camera Constrained
3	Free Path	Static	Global	Scene Constrained
4	Free Path	Dynamic	Local and Global	General case
5	Stereo / Multiscopic	Static / Dynamic	Constrained Global	Stereo / Multiscopic

A *static camera* refers to a camera fixed to a tripod or any other support that keep the objective still during the acquisition. *Free path* classification considers camera movement either because the camera is hand held or following a free path. Stereo or multiscopic acquisition includes stereo pairs of cameras or camera rigs composed by two or more positions of one or more cameras horizontally aligned, to capture respectively two or more views of the same scene. The scene is classified as *dynamic* if any object moves during the acquisition, no matter the amount of movement nor the size of the object. Otherwise the scene is considered *static*.

In every cases described above it is possible to generate HDR video. Even for static images (1st row of Table 1.2.2.1), it is possible to repeat the acquisition for a given time step and combine the resulting HDRs into a time-lapse HDR video [CA06, Est12]. Time-lapse (also known as slow motion) is a technique for capturing frames significantly slower than video frame rates. When played at video rates, time appears to be faster. This is often used to capture natural phenomena like sunrise or sunset, or in the animation industry. Camera and scene constrained sequences are in general easier to align than the general case as only one kind of misalignment takes place. The three cases (2nd, 3rd and 4th rows of table 1.2.2.1) are used in film production. The last case (5th row of Table 1.2.2.1) concerning stereo and multiscopic cameras, has become very popular since stereo and auto-stereoscopic displays appeared on the market.

1.2.2.2 Local and/or global misalignment

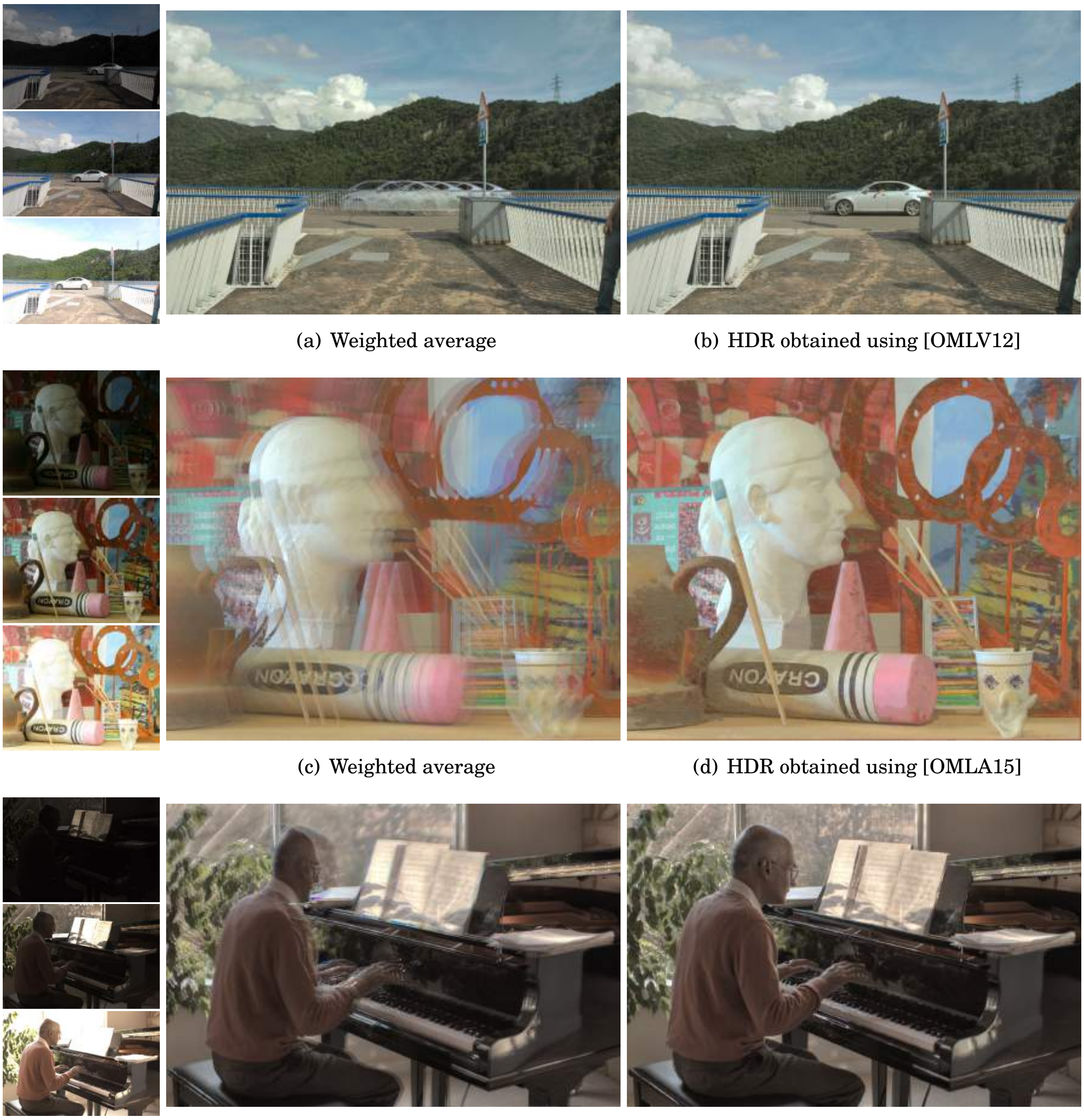
Misalignment as defined in the previous section can be categorized in three different types:

- **Global** misalignment is the consequence of camera motion (changes in position or orientation) and affects every pixel in the image (3rd and 5th rows of Table 1.2.2.1). It is common in exposure sequences acquired with hand held cameras although it is possible to find small global misalignment even for still sequences acquired using tripods (because of camera shaking with the mechanism activation or because of the wind). Between consecutive pairs of images, it is generally a small movement corresponding to translations or rotations. It may cause ghosting in the resulting HDR but some efficient techniques help to correct this misalignment. However, even for small movements, object occlusion and parallax could be difficult to solve.
- **Local** misalignment is produced by dynamic objects in the scene and affects only certain areas inside the image (2nd row of Table 1.2.2.1). Capturing a set of LDR images takes at least the sum of the shutter speed of each picture. This time is enough to introduce differences at the position of dynamic objects in the scene. In this case some areas occluded in some images may be visible in others. Depending on the speed of the dynamic object and the kind of the movement it may produce important differences between the inputs.
- **Local and Global** misalignment combines the two previous types and concerns the 4th row of Table 1.2.2.1. When a camera follows a free path to record a dynamic scene, each frame may contain both local and global misalignment. Pixels in the image could be affected by different types of transformation.

Figure 1.3 shows examples of movement in common multi-exposure sequences. The miniatures in the first column correspond to three exposures of the original sequences for each scene. The first row corresponds to local misalignment, a static camera captures consecutive images using different exposure times. Only dynamic pixels are affected by ghosting in the result, like the car in this example. Figures 1.3(a) and 1.3(b) show the results of merging the images without applying deghosting techniques and using the technique presented in the Chapter 3 [OMLV12].

In stereo sequences (Figure 1.3, second row), images were acquired at the same time by different cameras. Even if the scene is dynamic, both images correspond to the same time and no local misalignment is possible. The only misalignment possible is global, due to changes in the perspective from the two points of view. Figures 1.3(c) and 1.3(d) represent the results of merging the images without alignment and using the technique presented in the Chapter 4 [OMLA15] respectively.

In a sequence acquired from a hand held camera (Figure 1.3, third row), each exposure corresponds to different time instants but also to a slightly different viewpoint. Every pixel in the image is affected by global misalignment due to changes in the position of the camera while some pixels are also affected by local misalignment due to dynamic objects, like the pianist in this figure. Figure 1.3(e) shows the result of



(a) Weighted average

(b) HDR obtained using [OMLV12]

(c) Weighted average

(d) HDR obtained using [OMLA15]

(e) Weighted average

(f) HDR obtained using Chapter 6

Figure 1.3: Examples of local misalignment, global misalignment and a combination of both. Images courtesy of Sen *et al.*[SKY$^+$12] and the Middlebury dataset [Mid06]

merging the images directly and figure 1.3(f) is the result of using the technique presented in the Chapter 6.

1.2.3 Generating HDR content

If both the scene and the camera remain static during the acquisition, the exposures are aligned. The only possible result is an HDR image, although it can be used afterward for time lapse video production. In such case, any of the existing techniques [MP95, DM97, MN99] can be used to recover radiance values and merge them into an HDR image. Otherwise, misalignment needs to be corrected before merging the different exposures.

In cases where the camera remains static recording a dynamic scene, it is possible to detect the areas affected by dynamic objects of the scene and treat them locally. Several techniques exist for motion detection in the context of HDR image generation. Some of them focus on removing dynamic objects from the scene [KAR06, PH08, GSL08, SPS09]; this approach produces HDR images that are different from the original scene. Other works propose to replace the dynamic objects with the content of one exposure [Gro06, JLW08, LC09, GGC$^+$09, PK10, GKTT13], which in fact might introduce LDR content in the HDR. A few approaches recover HDR values by combining information of all sources in the sequence [RKC09, RC11, HLL$^+$11, OMLV12].

In the opposite case (static scene and dynamic camera), the movement between consecutive frames is very small and can be solved by finding homographies, or simply shifting one of the images. Some computationally efficient methods were proposed to solve such misalignment [Can03, War03, Cer06, TM07, Yao11].

The most difficult case is when both the camera and the scene move. In such case, dense correspondences between frames are required [ZBW11, HGP12, HGPS13, SKY$^+$12, OMLA15].

1.3 Structure of the thesis

The first chapter of this thesis provides the necessary knowledge to understand how light and color is measured. It summarizes what a digital image is and how it is represented. Chapter 2 presents a detailed introduction to HDR imaging, focused on the acquisition from multiple exposures including motion compensation and deghosting solutions.

Chapter 3 describes a first method to solve the ghosting artifacts in scenes with dynamic objects acquired from a fixed camera. Pixels in misaligned regions are detected and registered to the reference image. It includes a comparison on existing ghosting detection techniques and similarity measures for image registration.

Chapter 4 presents a method to generate multi-stereo HDR content. It is an extension for multi-stereo multiple exposure of an existing technique [SKY$^+$12]. We discuss the main drawbacks of the original method and propose improvements to use it on multi-stereo images.

Chapter 5 introduces the in-HDR-painting method to generate non reference dependent ghost free HDR images in fully dynamic scenes. This method is based on detecting the under and over-exposed areas and replace the content in such areas with correctly exposed values from other exposures in the sequence.

The results of the proposed methods are discussed at the end of each chapter and Chapter 6 summarizes as well as discusses about future work related to this thesis.

CHAPTER 2

FUNDAMENTALS AND PREVIOUS WORK

HDR imaging is inherently linked to disciplines related with light and color modeling and digital image processing. This chapter aims to provide a theoretical background on such disciplines that are fundamental for a better understanding of HDR imaging. It comprises also a detailed analysis on the state-of-the-art of multi-exposure alignment and deghosting for HDR acquisition. The last sections provide an overview of stereoscopic imaging and epipolar geometry that support a further analysis on stereoscopic HDR acquisition methods including the latest techniques for multiple exposures stereo matching.

2.1 Light measurement

The physics of light is described by two different models representing its dual nature: electromagnetic wave and particles (photons). Hence, there are two domains related with light measurement. *Radiometry*, that studies electromagnetic waves. The HVS is only sensitive to a small range of the spectrum, called visible light. Light in the visible spectrum can be described by *Photometry*, which describes the behavior of photons and light as humans perceive it. There is equivalence between radiometric quantities and the photometric ones. Table 2.1 shows the basic quantities to measure luminous energy.

Radiant energy describes the energy of light as an electromagnetic wave, it is denoted by Q_e and measured in joules (J). The flow of radiant energy in a time interval is called radiant power (P_e), measured in joules per second or watts (W). The amount of light incident on a surface is the radiant energy per time unit and per unit of area. This is the total radiant flux hitting a surface divided by its area or irradiance (E_e), measured in W/m^2. Radiance (L_e) measures the light incident on a surface from a particular direction, which is the irradiance of the surface over the solid angle of the light direction (W/m^2sr). Radiance hitting the camera sensors is transformed into digital images in a process described in section 2.3.

The radiance L_e registered in digital images can be approximated by a function known as the measurement equation. It is a function expressed in terms of the radiant energy in a given exposure time t over the surface area A of a pixel and the solid angle relative to the aperture value ω being θ the angle

Table 2.1: Photometric and Radiometric quantities.

	Quantity	Symbol	Unit
Radiometry	Radiant Energy	Q_e	J (Joules)
	Radiant Power	P_e	J/s = Watt
	Irradiance	E_e	W/m^2
	Radiance	L_e	$W/(m^2 sr)$
Photometry	Luminous Power	P_v	lm (Lumens)
	Illuminance	E_v	$lm/m^2 = lx$ (Lux)
	Luminous Intensity	I_v	$lm/sr = cd$ (Candela)
	Luminance	L_v	$lm/(m^2 sr) = cd/m^2 (Nit)$

between the surface normal and the angle of incidence.

$$L_e = \frac{d^2(dQ_e/dt)}{dA cos\theta d\omega} \tag{2.1}$$

The HVS is sensitive to a small range of wavelengths and it strongly varies with wavelength. Humans can perceive light with a wavelength in the range of approximately 400 to 700 nanometers (nm), with the highest sensitivity at 555 nm which corresponds to green light. Due to variations in the spectral sensitivity, a human may perceive a surface lit by blue light darker than one lit by green light of the same radiant power [Gut12].

The illuminance E_v corresponds to the brightness of a surface as humans perceive it. It is the result of weighting E_e with the sensitivity function $V(\lambda)$ proposed by the Commission Internationale de l'Eclairage (CIE). Each photometric quantity is the result of weighting the corresponding radiometric measure with $V(\lambda)$. They represent the same principle but adapted to our perception.

Luminous power is photometrically weighted radiant energy and is measured in lumens. If it is measured over a differential solid angle we obtain luminous intensity which is given in (lm/sr) or "candela". Illuminance is given in lumens per square meter (lm/m^2) or "lux". Luminance is the radiance as perceived by humans, it is specified in equation 2.2 and measured in (cd/m^2), also called "nits" [RWD+10].

$$L_v = \int_{380}^{830} L_{e,\lambda} V(\lambda) d\lambda \tag{2.2}$$

Radiometric and photometric quantities can be measured with lab instruments. Digital cameras are sensing devices not measuring devices. They could be used to approximately measure light but their nonlinear response to light must be characterized. This is a paramount step in HDR acquisition known as camera response function (CRF) recovery.

2.2 Color representation

Color is a perceptual phenomena although it could be defined simply as light of different wavelengths. *Colorimetry* is the field in charge of quantifying the human color perception in relation to the physics of light. Color space is a mathematical representation to describe color as a combination of primary color values (i.e. color components or color channels).

The human eye has around 130 millions of receptor cells, nearly 6 millions of them correspond to cones which are responsible to the perception of color, fine details, and fast changes. Three different types of cones determine our perception of color. Each of them is sensitive to different wavelengths: long $L(\lambda)$, middle $M(\lambda)$ and short wavelength $S(\lambda)$. The peak of wavelength sensitivities in each case is approximately 564 nm (red), 533 nm (green) and 437 nm (blue) [Boi14].

The color of a stimulus is rarely one pure wavelength but multidimensional, where each dimension is associated with particular wavelength. A visible color is a projection of this multidimensional variable to three primaries, corresponding to three types of cone cells.

In 1931, the *Commission Internationale de l'Éclairage (CIE)* standardized a set of primaries for the standard colorimetric observer. The standard was established to describe the visible color gamut (complete range or scope of a color space) in CIE XYZ color space. In 1976, the CIE presented two color spaces, the CIE LAB and the CIE LUV which are (approximately) perceptually uniform. This means that a variation in the color value will correspond to the same difference in perception. These definitions are still widely used today. However, there are multiple different color spaces grouped in five main models (CIE, RGB, YUV, HSL/HSV, and CMYK) suited for diverse specific purposes.

2.3 Digital Images

A digital image is a two-dimensional discrete representation of a continuous space stored in a digital support. Images can be generated digitally or captured using devices such as digital cameras or scanners. In case an image is acquired using a camera, it represents a projection of a real scene that passed through a set of lenses to a photosensitive surface in a digital sensor. In a digital picture, light intensity information is transformed in a color value stored at each picture element (pixel). The measured intensity values depend on the physical lighting distribution of the scene being recorded as well as on photosensitive sensor characteristics.

Sampling is the operation that converts the continuous light signal into a discrete digital representation. Digital images are formed of a finite number of points sampled from the scene. A photographic digital sensor consists of a set of sensor cells spatially distributed in the image plane that measures incoming light simultaneously. The number of recorded pixels is known as spatial resolution, given usually in rows per columns or in millions of pixels (megapixel). The resolution of digital devices has improved rapidly during the last decades, from VGA (640×480) in the '80s to today's 4K from (4096×3972) and rising.

Digital pictures are the result of exposing the camera sensor to light during certain *exposure time* that determines how long radiant flux is integrated in the sensor. Sensor cells have a fixed surface area, the lens optics with the aperture value limit the solid angle from which light reaches the sensor. The radiant energy Q_e accumulated in one shot for each cell is converted into a voltage and then to a discrete color value in a process known as *quantization*. The color depth of an image is limited by the number of bits used to store color information for each pixel. While resolution has been increasing during the last decades, a vast majority of images are still represented using 24 bits per pixel, although most sensors are capable of producing RAW images at 12-14 bits per color channel.

The cell elements of a camera sensor have a single spectral response curve instead of three different ones like our visual system. They have a single response coefficient to brightness so they can not distinguish between the wavelength distributions of different light stimuli. The most extended approach is to place a color filter over the sensor that allows only the transmission of a certain wavelength range for cell. In this

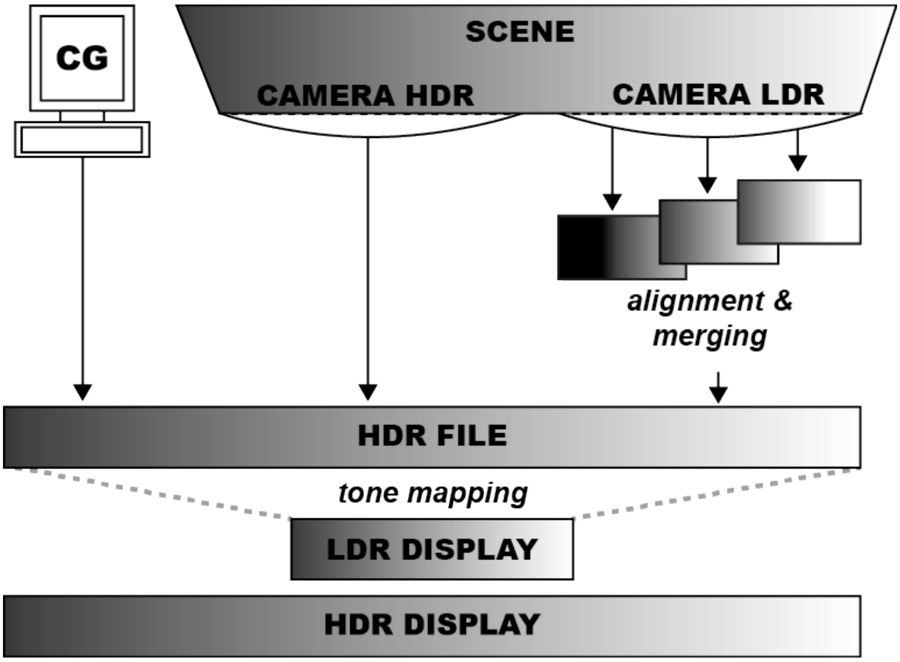


Figure 2.1: HDR image pipeline from scene to display

case, each sensor cell measures the red, green, or blue component of the incoming light in a distribution know as Bayer pattern. An interpolation process (known as demosaicing) is performed then to reconstruct a final RGB image.

Usually, the acquired image is encoded three 8-bit integer numbers (0-255). LDR images store only a small part of visible color gamut and their values correspond to a particular acquisition. LDR images have a non linear relation to the physical quantity of light reaching the camera sensor from the scene. Pixels with the same RGB color value in different images might represent very different intensities according to the acquisition setup. For instance, if a pixel has twice the intensity value of another from the same scene, it is unlikely that the sensor received twice the light. The contrast of real world scene can not be represented using LDR images, and it is not possible to reproduce accurately phenomenas like luminous surfaces or specular highlights.

2.4 HDR Images

HDR images are also known as radiance maps because they are directly related to the scene radiance. HDR pixels with certain luminance values always correspond to the same light intensity. HDR images are intended to store 'radiance' values to represent a scene accurately, rather than gamma corrected pixel color values. The range of values that can be represented is much wider than LDR images, which enables to get more details and minimizes the risk of over or under-saturated areas.

Figure 2.1 illustrates the pipeline of HDR imaging form acquisition to display. There are three ways to acquire or generate HDR contents:

- **CG Image synthesis**: CG simulates the propagation of light into a virtual 3D scene to obtain physically-based images from its projection into a virtual camera. Such simulations are performed in floating point to represent a wider dynamic range of scene radiance and minimize the quantization step.

- **Native HDR sensors**: Some HDR camera prototypes were presented lately to the research community [NB03, CBB$^+$09, TKTS11], but they are not yet available for commercial use. Commercial counterparts like the Viper camera [Tho05] or the Phantom HD camera [Res05] and the Red Epic [Red06] are only a few. Several companies claim to have HDR sensors (Kodak KAC-9628, IMS Chips HDRC sensors, Silicon Vision Products, SMaL Camera or Pixim). However, the addressed dynamic range remains limited (under 16 f-stops in most cases), and prices for such equipment are far from affordable for the average customer budget.
- **Multiple LDR exposures**: The dynamic range of a scene can be captured by a set of LDR images covering different ranges of light. This images can be merged afterwards into an HDR image. This option has become very popular and is already available in tools like Adobe Photoshop, HDRShop or Photomatix or mobile phones apps. When the LDR images are not aligned, some alignment process needs to be performed before or with the merging step to avoid artifacts. The following sections analyse existing techniques to deal with such cases.

LDR image formats like JPEG, PNG or BMP are known as *device-referred* because they were designed to cope with the capabilities of display devices. They are not directly related to the actual radiometric properties of the represented scene which makes difficult to reproduce its appearance accurately or to adapt the visualization to devices with different characteristics. On the other hand, HDR formats are *scene-referred* and they encode the actual photometric characteristics of the depicted scene. The conversion from such formats to a representation for a given device must rely on the device itself. This way devices can exploit their own capabilities to provide the best representation possible for a given image.

Pixel HDR values are stored in floating point triplets which means that the amount of possible values exceed the capabilities of the HVS in any viewing conditions [RKMS15]. They can be stored in extended formats like RGBE, HDR or OpenEXR [LJ10]. These representations increase the amount of stored data, an HDR pixel uses 96 bits while their LDR counterparts are four times less, 24 bits. This is the main challenge of HDR imaging, every step in the image pipeline (capture, storage and display) designed for LDR images needs to be adapted to deal with HDR formats.

Existing consumer devices are unable to deal with HDR values since most of the existing encoding algorithms or image processing tools are based on 8-bits images. Extending them to manipulate HDR images is not straightforward. Radiance values contained in HDR images can not be displayed on regular displays. For example, consider an HDR image with a very bright spotlight; the weak backlight of an LCD screen is unable to reproduce such amount of brightness. Even when an HDR image is generated or captured, it can only be displayed natively in HDR displays. HDR images need to be tone mapped to 8-bit images suitable for LDR displays.

2.5 HDR acquisition from multiple exposures

This section presents a detailed review of the state-of-the-art on HDR acquisition from multiple exposures, with special emphasis on image alignment. The acquisition process consists of a set of steps (setting parameters and capturing, radiance conversion, image registration, and HDR stitching) that are covered in this section. For a complete overview on HDR imaging refer to [RWD$^+$10, BADC11, RKMS15].

2.5.1 Acquisition setup

Combining multiple exposures of the same scene, each covering a different radiance range, is a solution for HDR acquisition using conventional digital cameras.

There are three parameters conditioning the amount of light reaching the sensor and how sensitive it is. The so called "exposure triangle" is formed by the aperture opening, the shutter speed (or exposure time), and the sensor sensitivity (ISO-value). The exposure value (EV) relates them in a way that each variation of 1 in EV corresponds to a change of one stop i.e., half or double of the exposure, either by halving or doubling one of them while keeping constant the rest, or a combination of changes.

Each of the variables in the exposure triangle are defined as follow:

1. The **aperture** of a camera is controlled by opening or closing a diaphragm, which is usually located in the middle of the lens assembly. The aperture is then specified as an F-number N, defined as $N = f/D$, where f is the focal length, and D is the diameter of the diaphragm opening. N is given as a sequence of square roots, e.g. f/4, f/5.6, f/8, f/11, f/16, f/22 which implies that a change in of one $f/stop$ halves or doubles the area of the aperture by 2x. Changing the aperture provokes changes in the depth of field which means changes in image focus, so it is not a good approach to control exposure for HDR merging purposes.

2. The **exposure time** is controlled by opening and closing a shutter. It is the time the sensor is exposed to light and is measured in fractions of a second. The longer the sensor is exposed, the more light enters and brighter is the image, which allows to get details in the dark areas. Longer exposure times increase the risk of having blurred images either because of hand held camera movements or moving objects. On the opposite, shorter shutter speeds reduce the amount of light and provide details in brighter areas of the scene, but dark areas will appear noisier.

3. The **ISO-value** defines the sensor sensitivity to light by varying the voltage observed at each pixel position before the A/D conversion. By convention, ISO 100 refers to no special modification, so it is the lowest ISO number available on most digital cameras. Doubling the ISO means an amplification by 2x of the voltage. Increasing the ISO makes the image brighter but also augments the noise, a random variation. Noise means random variation in the intensity of pixels that correspond to the same color in the scene.

The most common approach is to take pictures sequentially at regular exposure time intervals to ensure that each image contains useful information of different ranges of the scene. The auto-bracketing function available in many cameras is very useful to capture a set of LDR images at different EV. Using the auto exposure function, the camera automatically calculates the best aperture/shutter speed combination EV0 for the current lighting conditions to minimize over or under-exposure. Once EV0 is established, the auto bracketing function takes darker and brighter pictures respectively increasing and decreasing the shutter speed at regular intervals. Placing *neutral density filters* in front of the lens to attenuate the amount of light that enters the sensor is also an option used to control the exposure of images.

The accuracy of the radiance recovery is limited by the saturation and noise of the camera sensor. After the maximum allowed brightness level is reached the value of the sensor cells remains the same which makes impossible to measure radiance. On the contrary, there is a point where the incident light is indistinguishable from noise in the circuitry. Hence, radiance is best measured in the upper segment of the sensor cell's operating range just below saturation, where signal-to-noise ratio is the optimal [Gut12].

The optimal number of exposures depends on the dynamic range of the scene, a compromise should be found in each case. Using the higher number of images would reduce the noise in the final HDR, but also increase the acquisition time which implies bigger misalignment and ghosting artifacts in dynamic scenes.

2.5.1.1 Multiple exposures HDR video

Increasing the number of exposures or the shutter speeds is not always possible. When merging differently exposed video frames to generate HDR video this is not even an option. In this case, time constraints to keep video rates becomes first priority.

Similarly to auto exposure control of digital cameras, digital video cameras have a function called auto gain control (AGC) in charge of measuring the brightness distribution of the scene and calculating the best exposure settings for the scene conditions.

The most extended idea is to design a real-time exposure control algorithm that captures exposures at multiple steps up and down this optimal EV (for example, ± 2 stops) to obtain a high and a low exposures. In Kang's *et al.*[KUWS03] implementation, the exposure settings alternate between two different values with a ratio varying from 1 (if a single exposure is adequate to capture the scene intensities) to a maximum of 16. Mangiat [MG10] updated the exposure values every four frames trying to maximize shutter speed in the high exposure and the opposite for the low exposure as a way to increase the dynamic range covered.

Many authors use only two exposures (low and high) to generate frames of HDR video [KUWS03, ST04, MG10], while a most recent approach [KSB^{+}13] uses three (low, medium and high) exposures to generate the same number of HDR frames.

2.5.2 Image alignment and deghosting

The information of different dynamic range of the scene, captured in the multiple LDR exposures, must be merged into one HDR image. During the acquisition both the camera or the scene might move. Merging pixels in the same (x, y) position of different images that does not correspond to the same point in the scene produces artifacts in the HDR image. This is one of the main problems in HDR image acquisition using multiple exposures.

This section summarizes most of existing techniques presented in the last years to tackle this problem. They are grouped using the criteria established in Table 1.2.2.1 that classifies misalignment into global (subsection 2.5.2.1) or local (subsection 2.5.2.2). Subsection 2.5.2.3 considers approaches to avoid ghosting artifacts that does not rely on image alignment. Finally, subsection 2.5.2.4 describes the extension from image alignment and deghosting into HDR video acquisition from multiple exposures.

Previous surveys was published about this issue [SS12, HTM14, OTTE15]. Srikantha and Sidibé presented the first known classification for deghosting methods. Hadziabdic et al. [HTM14] present a comparison between state-of-the-art methods and alignment algorithms implemented on commercial software and propose a methodology to evaluate their results. More recently, Tarhan et al. [OTTE15] present a survey on deghosting methods.

2.5.2.1 Global alignment

Ward [War03] presented a technique called Median Bitmap Transform (MBT) to align differently exposed images. The algorithm transforms a set of 8 bit images (using only the green channel or a gray scale representation of the LDR) to a bitmap. The median intensity value is used as threshold because it is nearly insensitive to exposure variations. If the difference of two bitmaps is defined as a logical XOR, it shows where the images are misaligned. The alignment process is implemented iteratively to minimize differences with respect to a reference image. A pyramid of MTBs is used to speed up the process.

This method provides good results for images with a rather bimodal brightness distribution. But when a large number of pixels are near the median value, noise appears in the MTB. The noise is prevented using another threshold to exclude pixel values near to the median. Ward's proposal was designed for small translations of the camera and not for dynamic objects in the scene.

Different implementations and variations of this technique were proposed later. Grosch [Gro06] published a GPU implementation that considers also rotation of the camera not only on 8-bit images but also in radiance space. He used the CRF to predict the color of pixels in consecutive images. A threshold over the difference between the predicted and the actual image helps to identify ghost regions.

Pece and Kautz [PK10] designed a Bitmap Movement Difference (BMD) algorithm to detect and isolate clusters of moving pixels in a sequence. This method calculates a MTB for each image and mark areas that change their MTB value along the sequence. The moving areas are detected by summing all the bitmap values along the sequence and selecting pixels that are neither 0 or N, being N the number of images in the sequence. Morphological dilation and erosion help to refine the ghost mask in case of noise. If part of the scene is over or under exposed over all the sequence, or movement objects and background have similar intensity values, BMD fails.

MTB-based methods does not depend on the camera response function and their computational cost is low. On the other hand, they are only effective for small global misalignment or to detect ghosting areas.

A registration scheme based on scale invariant feature transform (SIFT) that tracks dynamic objects by matching their key points in the sequence were applied as well to this problem by Tomaszewska and Mantiuk [TM07]. A modified SIFT algorithm extracts key point descriptors that represent correspondences between key points in the reference image and the remaining LDR images. After finding SIFT features, homographies are calculated using the random sample consensus (RANSAC) algorithm.

Akyüz [Aky11] assumes that misalignment between consecutive images is translational and the correlation between pixels remains constant. For instance, if a pixel intensity is larger than its neighbor in the left and smaller than the right one, this relation will be the same in the next exposure except for over and under exposed values or pixels affected by noise. Correlation maps are constructed for areas of interest in the images based on this relation. The algorithm looks for the most similar correlation map in consecutive images using Hamming distance as a measure of similarity.

2.5.2.2 Local alignment

Bogoni [Bog00] used an optical flow based technique to perform per pixel registration applied after global affine registration. They use a Laplacian pyramid representation which decreases the sensitivity to exposure changes.

Sand *et al.* [ST04] proposed a combination of feature matching and optical flow for video matching. The idea is to identify parts of the image that can be easily matched to make the warping process. Their method is robust to changes in exposure and lighting, but dynamic objects sometimes hide parts of the scene in certain exposures and reveal them in others which leads to the optical flow parallax problem where there is not enough information to reconstruct HDR over the entire image [JLW08]. If there are objects moving at high speed in the scene, artifacts still appears [BDA+09].

Zimmer *et al.* [ZBW11] present an optical-flow, energy-based method for image alignment. A dense displacement fields between the reference image and each other image in the sequence is estimated using an energy minimization equation that combines a data term to evaluate the alignment in gradient domain and a smoothness term penalizing outliers. It is a robust method for image alignment. However, ghosting artifacts persist in areas with dynamic objects.

Sen *et al.* [SKY$^+$12] recently presented a method based on a patch-based energy minimization that integrates alignment and reconstruction in a joint optimization for HDR image synthesis. Their method relies on a patch-based nearest neighbor search proposed by [BSFG09] and a multi-source bidirectional similarity measure inspired by [SCSI08]. This method allows producing an HDR result that is aligned to one of the exposures and contains information from all the remaining exposures. The results are very accurate but dependent on the quality of the reference image. Artifacts may appear if the reference image has large under exposed or saturated areas.

Hu *et al.* [HGPS13] proposed a method to synthesize aligned images given a reference in a sequence of multiple LDR exposures. An energy minimization equation is used to calculate a color value for pixels in under/over exposed areas. The energy minimization uses two terms for radiometric and texture consistencies between the reference and the source image.

2.5.2.3 Deghosting

Some researchers focus their works on local misalignment assuming that images where acquired from a still camera or a global alignment was applied previously. Instead of finding correspondences between pixels, they try to generate HDR images without ghosting artifacts. Some different strategies have been proposed: to identify the misaligned pixels and exclude them in the HDR reconstruction, to modify the merging step to avoid misaligned pixels, detect the areas affected by misalignment, and replace them with content from the best exposure only. This section presents some of these approaches.

Background reconstruction:

Khan *et al.* [KAR06] proposed a probabilistic method for weighting pixels without any explicit movement detection. Weights are assigned not only to avoid over and under exposed values, but also according to their probability of pertaining to the background. Pixels are averaged using the weighted average function like in equation 2.5. But instead of weighting only according to the intensity values, a non-parametric estimation scheme calculates their probability of pertaining to the background. This method is computationally expensive because it requires several iterations and artifacts still may persist for scenes with deformable objects or complex transformations. The same principle is used by Pedone *et al.* [PH08] to improve Khan's work. They propagate the influence of low probabilities using energy minimization to avoid artifacts in the result. This method requires less iterations than Khan's.

Granados *et al.* [GSL08] presented a method for background estimation that can be also applied to HDR generation. It is an energy minimization method based in two assumptions: background objects are static and they represent the major part of the image. A cost function that includes constrains of intensity differences along the sequence is minimized using graph cuts.

A GPU based application was presented by Markowski [Mar09] which uses probabilities to automatically detect ghosting. A ghost map is generated for each LDR image estimating the probability that a pixel belongs to static or moving objects. This ghost map is used to exclude dynamic objects from the HDR composition.

Despite the degree of success of methods in this section, all of them share the same drawback: dynamic objects are omitted in the HDR image which makes it inconsistent with the original scene.

Ghosting detection:

Grossberg *et al.* [GN03a] proposed a method based on the Image Mapping Function (IMG). The IMF is a function that relates accurately pixel intensities of two images without recovering the CRF. The IMF is deduced from properties of the cumulative histogram. Differences between two consecutive images are calculated by applying a threshold around IMF. They can be combined with a logical XOR, and stored in a ghost mask that contains pixels affected by movement through the sequence.

Jacobs *et al.* [JLW08] presented two methods for motion detection in a sequence of LDR exposures. The first one assumes that the radiance variations across exposures are higher for pixels affected by movement. Hence, variance is considered an indicator of movement [JLW08, LJ10, RWD$^+$10]. A Variance Image (VI) is created with the weighted variance of radiance over the different exposures. Movement clusters are detected in a mask applying a threshold over the VI. The resulting binary image shows clusters of pixels that might be affected by movement. The assumption that radiance values in static pixels has low variance requires a reliable CRF estimation, if the CRF is not accurate, false positives appears in the mask. The second method calculates an Uncertainty Image (UI) using entropy as an indicator of potential movements. Local entropy at each pixel location in a neighborhood of radius five is computed, which generates the UI. The authors justify the use of entropy because it is not affected by intensity values. The pixels marked in the ghost mask are not included in the HDR merging. The VI method is implemented in the HDR reconstruction software named Photosphere but it may fail under large dynamic range scenes [PK10].

Gallo *et al.* [GGC$^+$09] presented a technique to deal with large amount of movement in the scene. The method detects region patches that do not cause artifacts when combined with a reference image. The HDR image is generated just using these patches. This method assumes that pixels measuring the same radiance have a linear relation. To select valid pixels the algorithm applies a threshold over the deviation of pixels from the predicted model. The HDR image is composed only by valid patches of each input image.

The solution proposed by Raman *et al.* [RKC09] is similar, they detect ghosting using block based comparison between exposures. They assume images are globally aligned and the first rows (5-10 upper rows in the images) are static. This region is used to calculate the IMF through a sixth order polynomial approximation. Similar to [GGC$^+$09], they compare predicted images to the actual ones and mark patches that does not follows the IMF to be ignored in the HDR composition.

Zhengguo *et al.* [LRZ$^+$10] used the IMF to detect moving objects forward and backward in the sequence using threshold operations over the sequence and predicted images. They propose to fill pixels corresponding to ghosting areas in the sequence by evaluating the IMF in a reference image bidirectionally.

Heo *et al.* [HLL$^+$11] calculate a joint probability density function (PDF) between a reference image and the rest of LDRs to estimate the global intensity transfer functions. A ghost mask is calculated by thresholding the joint PDF for each non-reference image.

Sidibe *et al.* [SPS09] presented an approach based in the fact that the inverse CRF is monotonic increasing. For two images I_1 and I_2, if their respective exposure times are related such that $\Delta t_1 < \Delta t_2$, then the radiance values satisfy $E_1 \le E_2$. This is enough to ensure that pixel intensities in both images satisfy the same order relation $I_1 \le I_2$, which can be generalized in a sequence of N exposures $\forall k \in [1...N]$ if $k < k'$, any pixel in (x, y)

$$I_{(x,y),k} \le I_{(x,y),k'}$$

Pixels that break this relation are part of dynamic regions or another unexpected variation of intensity over the sequence. This technique works in certain conditions, but it is susceptible to some color combinations between dynamic objects and background.

Granados *et al.* [GKTT13] presented an approach based on Markov Random Field. In the first step, input images are aligned with a global homography calculated from SURF key-points. The second step minimizes an energy function to find consistent and inconsistent subset of input exposures. Their function considers consistency and noise potential terms that penalize pixels prompt to introduce ghosting and noise in the resulting HDR. However, their method cannot recover the dynamic range of moving objects since moving objects are reconstructed from a single input image. The absence of semantic constraint in the HDR reconstruction may introduce artifacts such as object repetitions.

2.5.2.4 Per-frame HDR video generation

Merging stacks of differently exposed frames is the most extended approach for capturing HDR video with conventional digital cameras. After the alternating-exposure video are captured, a registration algorithm is applied to reconstruct an HDR result at every frame (see Fig. 2.2 and Fig.2.3).

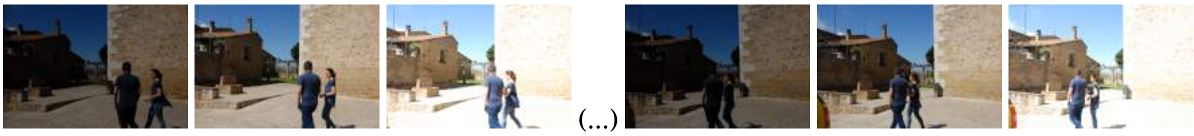


Figure 2.2: Multi-exposure video sequence alternating three different exposures.

To our knowledge, [KUWS03] proposed the first method to extend multiple exposure images methods to video sequences. Every HDR frame for a given time t_i is generated using information from adjacent frames t_{i-1} and t_{i+1}. They re-expose the short exposure frame with the long exposure times. Once images are transformed to the same exposure, motion estimation is performed for the two adjacent images. It consists of two steps:

1. Global registration by estimating an affine transform between them.
2. Gradient-based optical flow to determine dense motion field for local correction.

In the regions where the current frame is well-exposed, images are merged using a weighted function to prevent ghosting. For the over or under-exposed regions, the previous/next frames are bidirectionally interpolated using optical flow and a hierarchical homography algorithm.

Despite the novelty of this work and the promising results for some scenes, gradient-based optical flow is not accurate enough to find forward/backward flow fields. Boosting the short exposure to the long one will increase the noise, details like edges may be lost and slight variations of brightness may persist because of inaccuracies in the Camera Response Function (CRF). This may produce ghosting and errors in registration for fast non-rigid moving objects.

Sand and Teller [ST04] proposed an algorithm to register two different video sequences of the same scene. Video HDR is one of the most direct applications of their method. Differently exposed videos can be matched using their method. To initialize the process, they first compute a sparse set of features in both the reference and the target videos using a Harris corner detector [HS88], and find a preliminary set of correspondences. For each feature point identified, a matching cost is evaluated using two terms:

1. Pixel Consistency, instead of comparing equal pixels or patches in two images, they compare a single pixel in the reference image with a 3x3 patch in the source image. Correspondence is evaluated within a window around each pixel and pixel matching probabilities are assigned. The pixel consistency score is only penalized if the reference pixel is outside this a given range.

2. Motion Regression and Consistency, to determine how well a particular correspondence is consistent with its neighbors. The motion consistency score is high if the motion vector is well approximated by this regression.

Obtained matching are used to find regression predictions that are improved in a regression process. After finding high likelihood correspondences, a locally weighted linear regression method is used to interpolate and extrapolate correspondences for the rest of pixels, obtaining a dense correspondence field. This scheme is extended to all frame pairs of the video sequence. This method offers very good results for highly textured scenes but poor results otherwise. Processing each pair of frames might take up to 1.31 seconds and full video matching might take several minutes for each second of video input (on a single-processor desktop PC back in 2004).

Mangiat [MG10] proposed to improve the problems of the optical flow in Kang's [KUWS03] method by using a block-based motion estimation algorithm. They work also in a video sequence that alternate two exposure values. They use a CRF recovered using a sequence of 12 static exposures using the method presented by Devebec [DM97]. The short exposure is boosted using the CRF to match the long exposure.

They use a software [Süh08] to calculate block-based forward and backward motion estimation vectors for each frame with respect to the adjacent ones. However, such estimation is likely to fail in saturated areas. A second step of bidirectional motion estimation is performed to fill in the saturated areas with information from previous and next frames. The cost function is the Sum of Absolute Differences (SAD) adding a cost term that relates the motion vector estimated for adjacent frames. Block-based motion estimation is prone to artifacts such as discontinuities at block boundaries. Differences between the images in radiance domain are detected and assumed as artifacts. Such pixels are considered like holes and are replaced by pixels in the contour of such areas. Even though, poorly registered pixels may pass to the HDR merging step. They propose to use a cross-bilateral filter to treat the tone mapped HDR image using edge information at each frame. Despite the different attempts to avoid artifacts, fast motion (like eyes blinking for example) remains unsolved. The filtering step executed in the tone mapped images cannot be used for HDR displays.

The current state-of-the-art algorithm in HDR video reconstruction from an alternating-exposure sequence is the work of [KSB$^+$13], which extends the patch-based HDR reconstruction algorithm of Sen *et al.* [SKY$^+$12] to video. Although Sen's method produces still HDR images that are aesthetically pleasant, it does not maintain temporal coherence between subsequent frames. In their work, Kalantari *et al.* modify the HDR image synthesis equation to include an extra term that enforces temporal coherence.

To solve this new equation, their algorithm first approximates the global motion between consecutive frames using a similarity transform. It then uses optical flow to solve for the approximate local motion. To make the optical flow computation more robust, the algorithm performs a validation test between consecutive frames to ensure that the flow is consistent. Once this flow is computed, a window is set around each destination pixel with size inversely proportional to the accuracy of the flow estimation. These windows help constrain the patch-based search to ensure that the synthesized content is more coherent from frame to frame.

Once this pre-process is done, the algorithm minimizes the energy using a two-stage algorithm similar to that of Sen *et al.*, except that the bidirectional similarity term also enforces a similarity with the neighboring frames to ensure further coherence. This algorithm iterates until convergence, producing the final sequence of HDR images. Some results of this algorithm are shown in Fig. 2.3.

2.5.3 HDR merging

HDR pixels represent a radiance map $E(x,y)$ for every pixel in the image. After image alignment, every pixel (x,y) represents the same point in the scene for all LDR images. The intensity value $I(x,y)$ of LDR pixel values that were taken under different conditions need to be transformed into a common (scaled) radiance domain and merged by computing a weighted average.

When using a digital camera to acquire HDR images, values like the pixel surface area or the solid angle of the aperture are generally unknown and assumed to be constant. The result of inverting the image formation process represents radiance only up to an unknown scale. It worth specifying that in HDR the term *radiance* usually refers to a quantity that is proportional to radiometric radiance by an unknown factor [Gut12].

In many cameras there is a way to get data directly from the camera in RAW format. A RAW file is the record of data captured by sensor, it is not a single format but a general term for several proprietary file formats (.CRW, .MRW, .ORF, .NEF) that stores at least 12 bits per color channel. Most of images non-linearities are originated during conversion from RAW to 8 bits formats. This conversion includes processes like demosaicing, white balance, colorimetric interpretation, gamma correction and

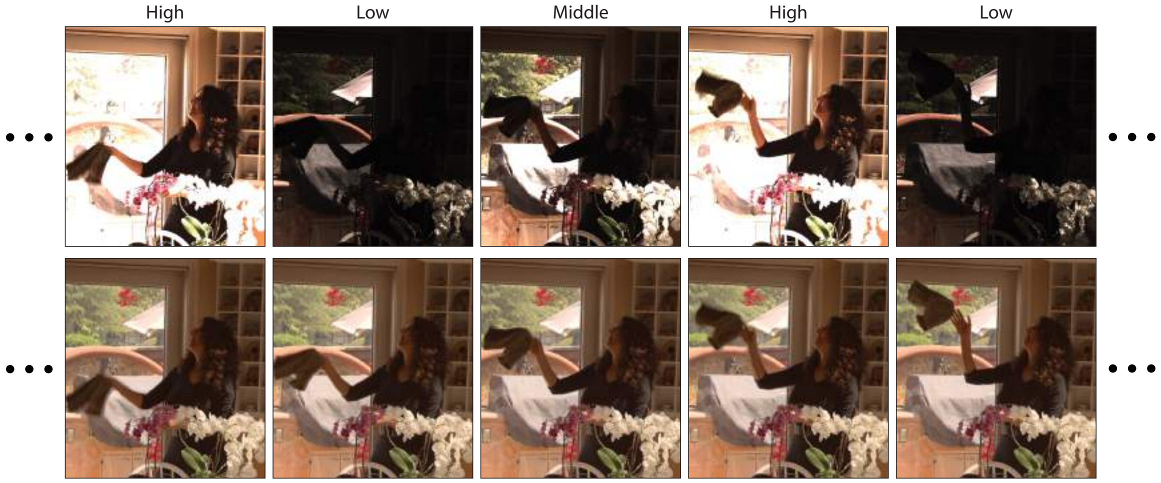


Figure 2.3: Results of the Kalantari *et al.* [KSB$^+$13] method to reconstruct an HDR video stream from a set of alternating exposures. The top row shows the input, in this case a scene that was imaged at three different exposure levels. The bottom row are the HDR frames that were reconstructed by the algorithm. Images courtesy of Kalantari *et al.*[KSB$^+$13].

noise reduction. RAW images are affected only by sensor saturation and quantization noise [Cer06]. Therefore, no CRF calculation is required while using RAW images.

A point in the scene corresponds to the same radiance $E(x,y)$ in each LDR images. Radiance is integrated inside a sensor cell for the duration of the exposure time Δt, which is different for each LDR. What a sensor cell actually measures is the exposure $E(x,y)\Delta t_i$, which is radiance integrated over time. Digital cameras introduce a function f (equation 2.3) that maps the radiance to intensity values in a range of 0 to 255. Next section 2.5.3.1 shows details on how f can be approximated.

$$I_n(x,y) = f(E(x,y)\cdot\Delta t_n) \tag{2.3}$$

Knowing the inverse of f and the shutter speed for each exposure, E can be approximated using equation 2.4.

$$\tilde{E}(x,y) = \frac{f^{-1}(I_n(x,y))}{\Delta t_n} \tag{2.4}$$

The final radiance map is calculated as a weighted average of the radiance values for each exposure[MP95, DM97, MN99]:

$$E(x,y) = \frac{\sum_{n=1}^{N} w(I_n(x,y))(\frac{f^{-1}(I_n(x,y))}{\Delta t_n})}{\sum_{n=1}^{N} w(I_n(x,y))} \tag{2.5}$$

The weighting function (w in the equation 2.5) is chosen to minimize the contribution of pixels that are under or over-exposed, hence its value should be small for pixel values close to 0 or 255. Typical weighting functions have Gaussian or Hat shape. Mann and Picard [MP95] used the derivative of the CRF for each color channel as weighting function. Devebec and Malik [DM97] and Khan *et al.*[KAR06] used simple hat functions like the one in equation 2.6 Figure 2.4.

$$w(I(x,y)) = 1-(2\cdot I(x,y)-1)^{12} \tag{2.6}$$

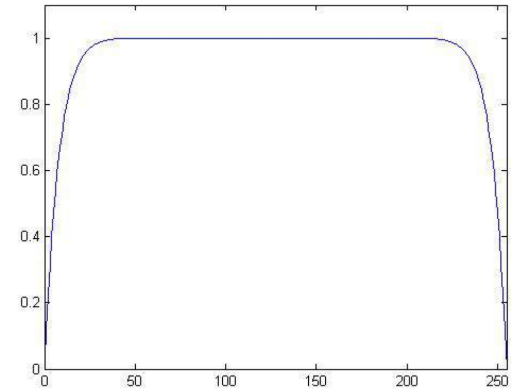


Figure 2.4: Weighting function proposed Khan *et al.*[KAR06], equation 2.6.

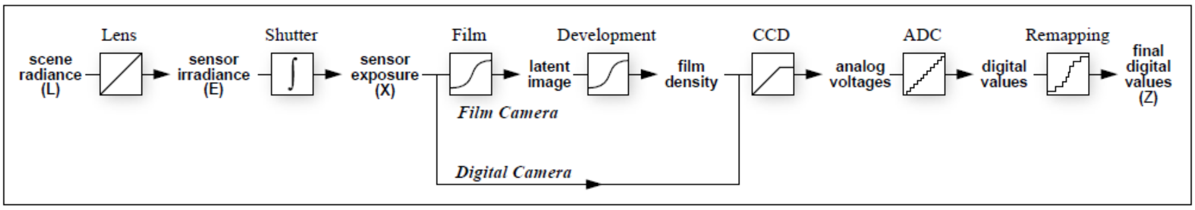


Figure 2.5: **Image Acquisition Pipeline** shows how radiance from the scene is converted to pixel intensity values for both film and digital cameras. Unknown nonlinear mappings can occur during exposure, development, scanning, digitization, and remapping. Image courtesy of Devebec *et al.*[DM97].

2.5.3.1 Camera Response Function

In order to transform pixel values to radiance, the transformation between values acquired by the sensor and pixel intensities needs to be known. This function is know as Camera Response Function (CRF), it is not linear and in most cases it is not provided by camera vendors. The most significant non-linearity occurs at the saturation point, where any pixel above this point is mapped to the same value, 255 . Figure 2.5 shows the pipeline of transformations that occurs since the light pass through the lens until digital values are assigned to pixels.

The inverse of the CRF is required to combine LDRs into an HDR image because it allows to convert from pixel intensities to radiance. Sensors produce a charge directly proportional to the amount of light that they receive. The digitization process uses analog-digital converters to transform the accumulated charge to integer values in [0, 255].

This function is assumed to be monotonically increasing and it approximate the non-linear transformations introduced at different stages of the acquisition process. The CRF depends on the camera vendor, who considers it part of their proprietary product so we need to calculate it. Under the assumption that$f$ is monotonically increasing, the existence of g in equation 2.7 is guaranteed [DM97].

$$E \cdot \Delta t = f^{-1}(I) = g(I) \tag{2.7}$$

Several approaches attempt to obtain g making assumptions on its shape and behavior. Mann and Picard [MP95] defined a Wyckoff's set like a collection of images that differ only in the exposure. Having different exposed images of the scene ensures that at least one of them would contain correct information on the different exposed areas, so all the information of the scene can be recovered and stored in a HDR image. They proposed an automatic method to combine them in a single picture of extended dynamic range and improved color fidelity.

The algorithm assumes an empirical function with gamma shape $g = \alpha + \beta I^{\gamma}$, where α is the minimal density obtained from a picture taken with lens covered, β is an arbitrary scale factor and γ is a contrast parameter estimated by regression. This method is highly restrictive so it does not lead to accurate results and does not support most of CRFs [BDA+09].

Debevec and Malik [DM97] presented instead a less restrictive method for recovering HDR images from photographs captured with conventional equipment. Their method used the different exposed images to recover the camera response function. With the response curve, pixel intensity values are converted into irradiance and combined in a HDR. It is based in the reciprocity equation 2.6, where halving the irradiance E and simultaneously doubling the exposure time Δt result in the same pixel values I. They take the natural logarithm on both sides of the equation to approximate the CRF:

$$\ln f^{-1}(I) = \ln E + \ln \Delta t \tag{2.8}$$

From this equation we know I and Δt, it is also reasonable to assume that f^{-1} is smooth and monotonic. Devebec *et al.* used least square error to calculate both f^{-1} and E minimizing the error from the set of equations resulting of equation 2.8. This approach is less restrictive than Mann's and obtains good results for images that are not too noisy [MN99].

There is another solution presented by Mitsunaga *et al.* [MN99] that does not require precise estimates of the exposure times used. They improve the previous approach using a flexible polynomial model for representing a wide range of response functions. This method determines the minimum required order N and the coefficient c_n.

$$E = f^{-1}(I) = \sum_{n=0}^{N} c_n I^n \tag{2.9}$$

2.6 Multiscopic HDR

In this section we address the generation of HDR images for two or more views. Multiscopic HDR is a special case for general HDR acquisition from multiple exposures. However, there are constrains in the geometry of stereoscopy that transform the general image alignment into a stereo matching problem. It is worth to present it in a different section for a better understanding.

Orozco *et. al.* [OMLA16] resume most relevant existing works on multiscopic and HDR video. In this section we first review the basics of stereoscopic imaging (section 2.6.1 and epipolar geometry (section 2.6.2), before we discuss the recent contribution for the generation of multiscopic HDR images (section 2.6.3).

2.6.1 Stereoscopic Imaging

Apart from a huge amount of colors and fine details, our visual system is able to perceive depth and tridimensional shape of objects. Digital images offer a representation of reality projected in two dimensional arrays. We can guess the distribution of objects in depth because of monoscopic cues like perspective, but we cannot actually perceive depth in 2D images. Our brain needs to receive two slightly different projections of the scene to actually perceive depth.

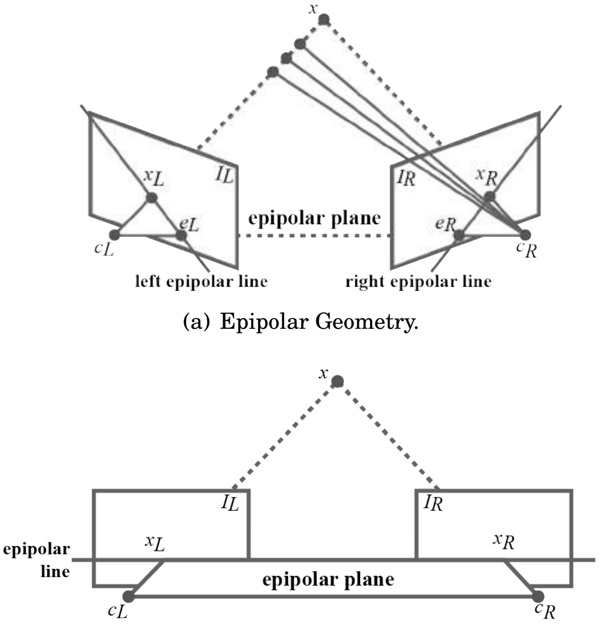
Stereoscopy is any imaging technique which enhances or enables depth perception using the binocular vision cues [SDBRC13]. Stereo images refers to a pair of images horizontally aligned and separated at a scalable distance similar to the average distance between human eyes. The different available stereo display systems project them in a way such that each eye perceives only one of the images. In recent years, technologies like stereoscopic cameras and displays have become available to consumers [UCES11, MPS12, DPPC13]. Stereo images requires to record at minimum two views of a scene, one for each eye. However, depending on the display technology it could be more. Some auto-stereoscopic displays render more than 9 different views for an optimal viewing experience [LLR13].

Some prototypes were proposed to acquire stereo HDR content from two or more differently exposed views. Most approaches [TKS06, LC09, SMW10, Ruf11, BRG$^+$14, AKCG14] are based on a rig of two cameras placed like a conventional stereo configuration that captures different exposed images. Next sections offer a background of the geometry of stereo systems (section 2.6.2) as well as a survey on the different existing approaches for multiscopic HDR acquisition (section 2.6.3).

2.6.2 Epipolar Geometry

One of the most popular topic of research in computer vision is stereo matching, which refers to the correspondence between pixels of stereo images. The geometry that relates 3D objects to their 2D projection in stereo vision is known as *epipolar geometry*. It explains how the stereo images are related and how depth can mathematically be retrieved from a pair of images.

Figure 2.6 describes the main components of the epipolar geometry. A point x in the 3D world coordinates is projected onto the left and right images I_L and I_R respectively. c_L and c_R are the two centers of projection of the cameras, the plane formed by them and the point x is known as the epipolar plane. x_L and x_R are the projections of x in I_L and I_R respectively.



(a) Epipolar Geometry.

(b) Epipolar Geometry Rectified.

Figure 2.6: Main elements of the epipolar geometry.

For any point x_L in the left image, the distance to x is unknown. According to the epipolar geometry, the corresponding point x_R is located somewhere on the right epipolar line. Epipolar geometry does not mean direct correspondence between pixels. However, it reduces the search for a matching pixel to a single epipolar line. The accurate position of a point in the space requires the correct matches between the two images, the focal length and the distance between the two cameras. Otherwise, only relative measures can be approximated.

If the image planes are aligned and their optical axes are parallel, the two epipolar lines (left and right) converge. In such case, correspondent pixels rely on the same epipolar line in both images, which simplifies the matching process. Aligning the cameras to force this configuration might be difficult but images can be aligned.

This alignment process is known as *rectification*. After the images are rectified, the search space for a pixel match is reduced to the same row which is the epipolar line. To the best of our knowledge, all methods in Stereo HDR are based on rectified images and they take advantage of the epipolar constrain during the matching process. Rectified image sets are available on the Internet for testing purposes, like [Mid06].

If corresponding pixels (matches) are on the same row on both images, it is possible to define the difference between the images by the horizontal distance between matches in the two images. The image that stores all the horizontal shifts between stereo pairs is called *disparity maps*.

Despite epipolar geometry simplifies the problem it is far from being solved. Determining pixel matches in regions of similar colour is a difficult problem. Moreover, the two views correspond to different projections of the scene, which means that occlusion takes place between objects. The HDR context adds the fact that different views might be differently exposed reducing the possibilities of finding color consistent matches.

2.6.3 Multiple exposure stereo matching

Stereo matching (or disparity estimation) is the process of finding the pixels in the different views that correspond to the same 3D point in the scene. The rectified epipolar geometry simplifies this process to find correspondences on the same epipolar line. It is not necessary to calculate the 3D point coordinates to find the correspondent pixel on the same row of the other image. The disparity is the distance d between a pixel and its horizontal match in the other image.

Akhavan *et al.* [AYG13, AKCG14] compared the different ways to obtain disparity maps from HDR, LDR and tone-mapped stereo images. A useful comparison among them is offered and illustrates that the type of input has a significant impact on the quality of the resulting disparity maps.

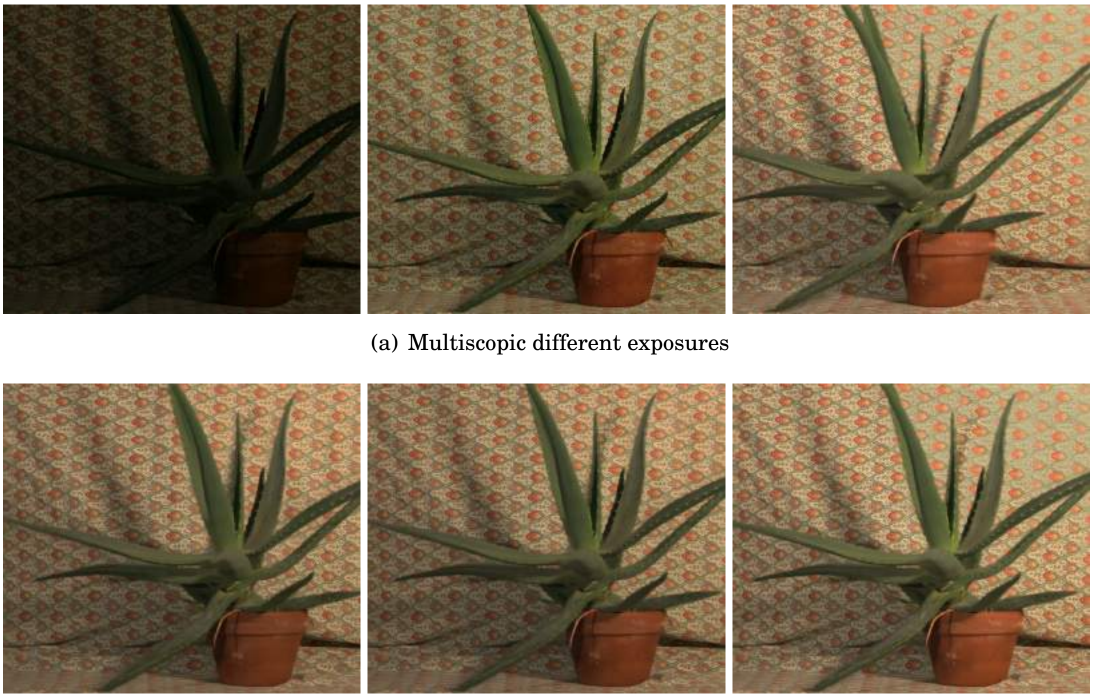
Figure 2.7 shows an example of a differently exposed multi-view set corresponding to one frame in a multiscopic system of three views. The main goal of stereo matching is to find the correspondences between pixels to generate one HDR image per view for each frame.

Correspondence methods rely on matching cost functions to compute the color similarity between images. It is important to consider that the exposure difference needs to be compensated. Even using radiance space images were pixels are supposed to have the same value for same points in the scene, there might be brightness differences. Such differences may be introduced by the camera due to image noise, slightly different settings, vignetting or caused by inaccuracies in the estimated CRF. For good analysis and comparison of the existing matching costs and their properties, refer to [SS02, HS09, BVNL14].

Exist different approaches to recover HDR from multi-view and multi-exposed sets of images. Some of them [TKS06, LC09, SMW10] share the same pipeline as in Figure 2.8. All mentioned works take as input a set of images with different exposures acquired using a camera with unknown response function. In such cases, the disparity maps need to be calculated in first instance using the LDR pixel values. Matching images under important differences of brightness is still a big challenge in computer vision.

2.6.3.1 Per frame CRF recovery methods

To our knowledge, Troccoli *et al.* [TKS06] introduced the first technique for HDR recovery from multiscopic images of different exposures. They observed that Normalized Cross Correlation (NCC) is approximately



(a) Multiscopic different exposures

(b) Multiscopic tone mapped HDR images

Figure 2.7: LDR Multiscopic sequence and the HDR counterpart. Up: 'Aloe' set of LDR multi-view images from Middlebury web page. Down: the Tone-mapped HDR result. Images courtesy of [Mid06].

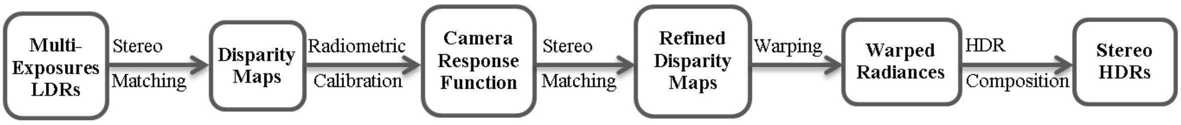


Figure 2.8: General multi-exposed stereo pipeline for Stereo HDR. Proposed by [TKS06], used by [SMW10, LC09] and modified later by [BRG$^+$14].

invariant to exposure changes when the camera has a gamma response function. Under such assumption, they use the algorithm described by Kang *et al.* [KS04] to compute the depth maps that maximizes the correspondence between one pixel and its projection in the other image. The original approach [KS04] used Sum of Squared Differences (SSD) but it was substituted by NCC in this work.

Images are warped to the same viewpoint using the depth map. Once pixels are aligned, the CRF is calculated using the method proposed by Grossberg and Nayar [GN03a] over a selected set of matches. With the CRF and the exposure values, all images are transformed to radiance space and the matching process is repeated, this time using Sum of Squared Differences (SSD). The new depth map improves the previous one and helps to correct artifacts. The warping is updated and HDR values are calculated using a weighted average function.

The same problem was addressed by Lin and Chang [LC09]. Instead of NCC, they use SIFT descriptors to find matches between LDR stereo images. SIFT is not robust under different exposure images. Only the matches that are coherent with the epipolar and exposure constraints are selected for the next step. The selected pixels are used to calculate the CRF.

The stereo matching algorithm they propose is based on a previous work [SZS03]. Belief propagation is used to calculate the disparity maps. The stereo HDR images are calculated by mean of a weighted average function. Even using the best results only, SIFT is not robust enough under significant exposure variations.

A ghost removal technique is used afterward to tackle the artifacts due to noise or stereo mismatches. The HDR image is exposed to the best exposure of the sequence. The difference between them is calculated and pixels over a threshold are rejected considering them like mismatches. This is risky because HDR values in areas under and over exposed in the best exposure may be rejected. In this case ghosting would be solved but LDR values may be introduced in the resulting HDR image.

Sun *et al.* [SMW10] (inspired by [TKS06]) also follow the pipeline described in Figure 2.8. They assume that the disparity map between two rectified stereo images can be modeled as a Markov random field. The matching problem is presented like a Bayesian labeling problem. The optimal label (disparity) values are obtained by minimizing an energy function.

The energy function they use is composed of a pixel dissimilarity term (NCC in their solution) and a disparity smoothness term. It is minimized using the graph cut algorithm to produce initial disparities. The best disparities are selected to calculate the CRF with the algorithm proposed by Mitsunaga and Nayar [MN99].

Images are converted to radiance space and then another energy minimization is executed to remove artifacts. This time the pixel dissimilarity cost is computed using the Hamming distance between candidates.

The methods presented until here have a high computational cost. Calculating the CRF from non-aligned images may introduce errors since the matching between them may not be robust. Two exposures are not enough to obtain a robust CRF with existing techniques. Some of them execute two passes of the stereo matching algorithm the first one to detect matches for the CRF recovery and a second one to refine the matching results. This might be avoided by calculating the CRF in a previous step using multiple exposures of static scenes. Any of the available techniques [MP95, DM97, MN99, GN03a] can be used to get the CRF corresponding to each camera. The curves help to transform pixel values into radiance for each image and the matching process is executed in radiance space images. This avoids one stereo matching step and prevents errors introduced by disparity estimation and image warping.

2.6.3.2 Offline CRF recovery methods

Bonnard *et al.* [BLV+12] propose a methodology to create content that combines depth and HDR video for auto-stereoscopic displays. Instead of varying the exposure times, they use neutral density filters to capture different exposures. A camera with eight synchronized objectives and three pairs of 0.3, 0.6 and 0.9 filters plus two non filtered views provide eight views with four different exposures of the scene stored in 10-bit RAW files. They use a geometry-based approach to recover depth information from epipolar geometry. Depth maps drive the pixel match procedure.

Batz *et al.* [BRG+14] present a work-flow for disparity estimation divided in the following steps:

- *Cost initialization* consists in evaluating the cost function, Zero Normalized Cross Correlation (ZNCC) in this case, for all values within a disparity search range.

The matching is performed on the luminance channel of radiance space image using patches of 9x9 pixels. The result of searching for disparities is the Disparity Space Image (DSI), a matrix of $m \times n \times d + 1$ for an images of $m \times n$ pixels with *d+1* being the disparity search range.

- *Cost aggregation* smooth the DSI and find the final disparity of each pixel in the image. They use an improved version of the cross-based aggregation method described by Mei *et al.* [MSZ+11]. This step is performed not in the luminance channel like in the previous step but in the actual RGB images.
- *Image warping* is in charge of actually shifting all pixels according to their disparities. Dealing with occluded areas between the images is the main challenge in this step. The authors propose to do the warping in the original LDR images which adds a new challenge: dealing with under and over exposed areas. A backward image warping is chosen to implicitly ignore the saturation problems. The algorithm produces a new warped image with the appearance of the reference one by using the target image and the corresponding disparity map. Bilinear interpolation is used to retrieve values at subpixel precision.

Selmanovic *et al.* [SDBRC14] propose to generate Stereo HDR video from a pair of HDR and LDR videos, using an HDR camera [CBB+09] and a traditional digital camera (Canon 1Ds Mark II) in stereo configuration. This paper is an extension to video of a previous one [SDBRC13] focused only on stereo HDR images. In this case, one HDR view needs to be reconstructed from two different sources.

Their method proposes three different approaches to generate the HDR:

1. *Stereo correspondence* is computed to recover the disparity map between the HDR and the LDR images. The disparity map allows to transfer the HDR values to the LDR image. The sum of absolute differences (SAD) is used as a matching cost function. Both images are transformed to Lab color space which is perceptually more accurate than RGB.

   The selection of the best disparity value for each pixel is based on *winner takes all* (WTA) technique. The lower SAD value is selected in each case. An image warping step based on Fehn's work [Feh04] is used to generate a new HDR image corresponding to the LDR view. The SAD stereo matcher can be implemented to run in real time but the resulting disparity maps could be noisy and not accurate. The over and under exposed pixels may end up in a wrong position. In large areas of the same color and hence same SAD cost, the disparity will be constant. Occlusions, reflective or specular objects may cause some artifacts.

2. *Expansion operator* could be used to produce an HDR image from the LDR view. Detailed state-of-the-art reports on LDR expansion were previously published [BDA+09, HS11]. However, in this case, we need the expanded HDR to remain coherent with the original LDR. Inverse tone mappers are not suitable because the resulting HDR image may be different from the acquired one, producing results not possible to fuse through a common binocular vision.

   They propose an expansion operator based on a mapping between the HDR and the LDR image using the first one as reference. A reconstruction function maps LDR to HDR values (equation 2.10) based on an HDR histogram with 256 bins putting the same number of HDR values in each bin as there are in the LDR histogram.

$$RF = \frac{1}{Card(\Omega_c)} \sum_{i=M(c)}^{M(c)+Card(\Omega_c)} C_{hdr}(i) \tag{2.10}$$

In equation 2.10, $\Omega_c = \{j = i..N : c_{ldr}(j) = c\}, c = 0..255$ is the index if a bin Ω_c, $Card(\cdot)$ returns the number of elements in the bin, N is the number of pixels in the image, $c_{ldr}(j)$ are the intensity values for the pixel j, $M(c) = \sum_0^c Card(\Omega_c)$ is the number of pixels in the previous bin and c_{hdr} are the intensities of all HDR pixels sorted ascending. RF is used to calculate the look-up table (LUT) and afterward expansion can be performed directly assigning the corresponding HDR value to each LDR pixel.

The expansion runs in real time, is not view dependent, and avoids stereo matching. The main limitation is again on saturated regions.

3. *Hybrid method* combines the two previous ones. Two HDR images are generated using the previous approaches (Stereo Matching and Expansion Operator). Pixels in well exposed regions are expanded using the first method (expansion operator) while matches for pixels in under- or over-exposed regions are found using SAD stereo matching adding a correction step. A mask of under and over saturated regions is created using a threshold for pixels over 250 or below 5. The areas out of the mask are filled in with the expansion operator while the under or over exposed regions are filled in with an adapted version of the SAD stereo matching to recover more accurate values in over or under exposed regions.

Instead of having the same disparity over the whole under or over exposed region, this variant interpolates disparities from well exposed edges. Edges are detected using a fast morphological edge detection technique described by Lee [LHS87]. Even though, some small artifacts may still be produced by the SAD stereo matching in such areas.

Orozco *et al.* [OMLA15] presented a method to generate multiscopic HDR images from LDR multi-exposure images. They adapted a patch match approach [SKY$^+$12] to find matches between stereo images using epipolar geometry constrains. This method reduces the search space in the matching process and includes an improvement of the incoherence problem described for the patch-match algorithm. Each image in the set of multi-exposed images is used as a reference, looking for matches in all the remaining images. These accurate matches allow to synthesize images corresponding to each view which are merged into one HDR per view that can be used in auto-stereoscopic displays.

2.7 Summary

This chapter introduced the main concepts related to high dynamic range imaging. It provided definitions of many underlying concepts including light, color, digital and HDR imaging. Methods of capturing HDR content (e.g. the multiple exposure technique) and details in each step of the HDR acquisition pipeline. Finally, an analysis of existing stereo HDR acquisition techniques was covered.

CHAPTER 3

RECOVERING HDR FOR MOVING OBJECTS

3.1 Introduction

This chapter presents a technique to deal with dynamic objects in HDR reconstruction by gathering HDR information of dynamic objects from a set of differently exposed LDR images [OMLV12]. The input images are assumed to be globally aligned. Our goal is to reach the maximum HDR coverage in the image as permitted by the input LDR images. No objects are removed from the image and pixels in the areas affected by movement are registered and merged in HDR values.

Once regions in movement are identified, each dynamic object is registered to a reference image. We increase the dynamic range by combining registered pixels, allowing HDR values even in areas affected by movement. The best results are achieved in scenes where the dynamic objects as well as their movement in the image sequence are roughly rigid. It means objects that don't change their shape considerably during the sequence (cars, motorbikes, planes) and that their motion trajectory can be approximated by translations.

3.2 Recovering HDR for Moving Objects

One common approach to solve local misalignment is either to exclude dynamic objects from the HDR image or to replace such areas with content from one exposure only. Excluding content from the image makes it incoherent with the original scene and replacing the content with only one exposure might reduce the dynamic range in such areas comparing to the rest of the scene.

The aim of this work is to provide HDR values even for areas in movement. The input sequence are LDR images acquired at different exposure times, with the same aperture value and from the same viewpoint or globally registered. We assume misalignment to be local, well-defined areas of the image rather than the full image. In other words, we assume that images are perfectly aligned except for moving objects.

We propose a novel approach based on a framework of three steps to detect the affected areas, register them and combine the content into a coherent HDR image, as shown in Figure 3.1. In each step, an analysis and comparison of previous works helps to chose the appropriate solution.

1. **Ghost detection**: Four of the most used methods for ghost detection were implemented and compared achieving different degrees of success, as discussed in section 3.2.1. The result of this step is a mask where clusters of pixels affected by movement are identified (red box of Figure 3.1).
2. **Registration**: The second step is registering regions affected by movement to a reference image (green square of Figure 3.1). We implemented and compared four similarity measures for image registration. An image pyramid is implemented to speed up the registration process. This step is described in section 3.2.2.
3. **HDR composition**: Finally, both the registered areas and the rest of the image are merged into an HDR image (blue square in Figure 3.1). Pixels from dynamic areas are carefully combined after registration excluding possible outliers. Ghosting areas are replaced with the obtained HDR values (section 3.2.3).

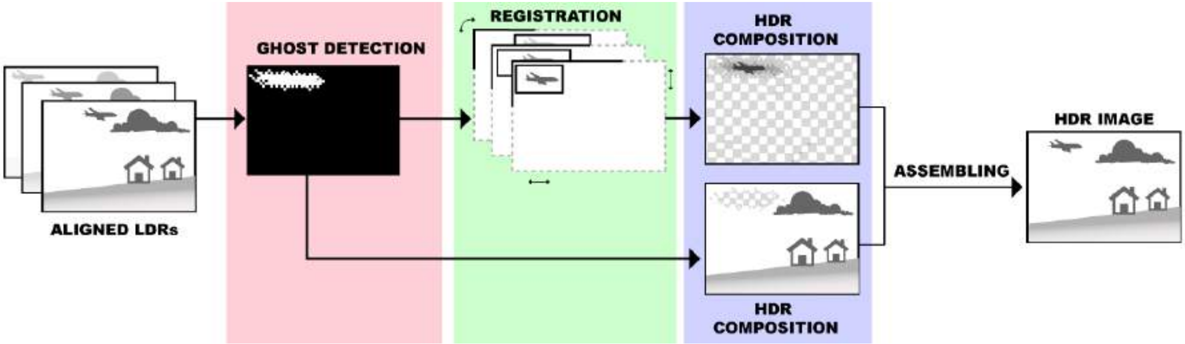


Figure 3.1: Framework to achieve HDR image in dynamic scenes acquired from static cameras.

3.2.1 Ghost mask generation

This section is focused on detecting areas affected by movement. Like shown in Figure 3.2, moving objects may appear in different parts of the image for each image in the sequence. It is important not only to register the moving objects but to tackle occluded areas behind them.

We implemented and compared four well-known methods already presented in the previous chapter. Since they attempt to identify regions that if merged produce ghosting artifacts, they are often known as ghost detection methods. Depending on the method used, ghost detection can be performed either in the LDR images or once they are transformed to radiance space. The objective is to identify pixels that change unpredictably along the sequence (Figure 3.2).

3.2.1.1 Median Threshold Bitmap (MTB)

Ward [War03] found that the median intensity value is nearly insensitive to exposure variations, specially for image with a rather bimodal brightness distribution. He used the median intensity value of a gray representation of images to build MTB from each image in the sequence.

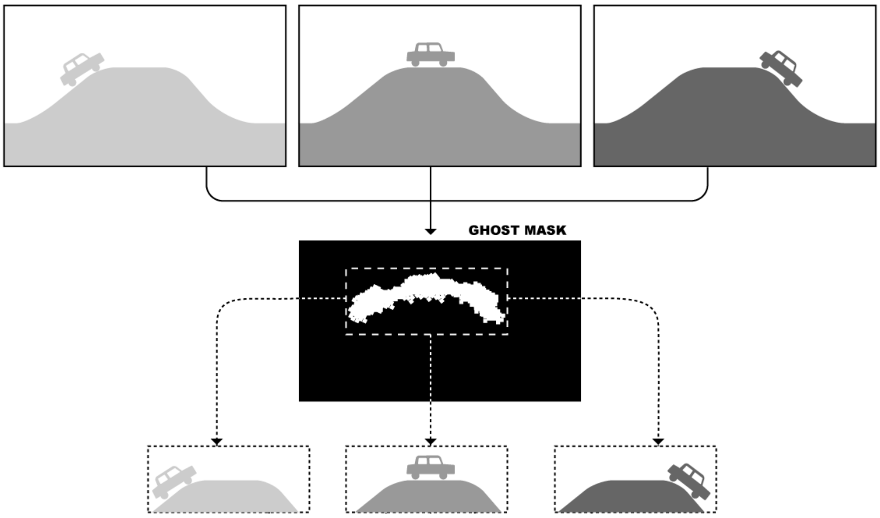


Figure 3.2: Ghost mask generation.

Pece and Kautz [PK10] proposed to use the MTB to detect clusters of pixels affected by movement in a sequence using simple binary operations. If there is no movement in the scene, pixels in the same position are expected to have the same value in all binary bitmaps. The difference between MTB images indicates pixels changes along the sequence. The difference between all bitmaps can be calculated using logical XOR or the following expression:

$$\sum_{i=1}^{N} M_i(x,y) \notin 0,N \tag{3.1}$$

Where M_i corresponds to the MTB of each image in a sequence of N exposures. Summing all the bitmap values along the sequence and selecting pixels that are neither 0 or N.

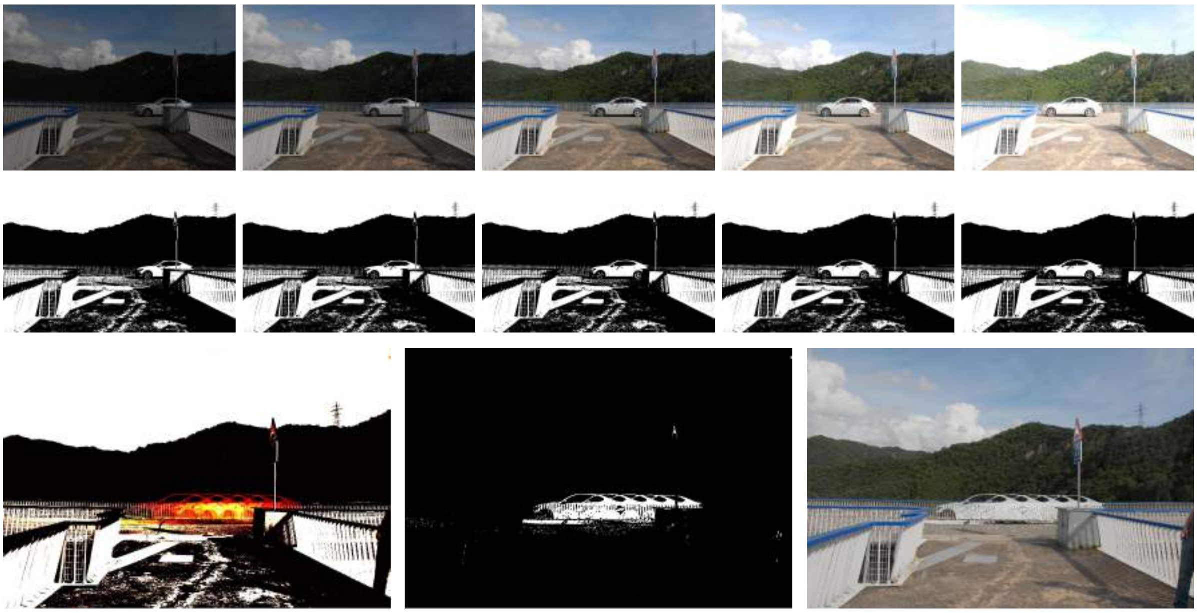
The result is a bitmap containing pixels that are not constant through the sequence either because movement or false positives. False positives are pixels that change their value along the sequence but does not represent any object movement, they are mainly consequence of noise. Morphological operations of erosion and dilation help to eliminate such noise.

Figure 3.3 shows in the first row a sequence example of five LDR images. The images were acquired from a tripod and show a car moving with a static background. The corresponding bitmaps are shown in the second row. Figure 3.3(a) shows the sum of all bitmaps according to equation 3.1, in color appears pixels that are neither 0 or 5. Such pixels are stored in a binary mask which represent the movement in the sequence (Figure 3.3(b)).

There are cases where this method fails to detect movement. When colors from the dynamic object and the background are both at the same side of the median threshold, this method fails. The example in Figure 3.3 shows clearly that the car glasses are not detected as movement due to their similarity with the background, neither parts of the car that overlaps the white bars are detected.

3.2.1.2 Variance-based methods

Variance of pixels along the sequence of LDR images may be an indicator to detect areas affected by movement [JLW08]. The variance of pixel intensities $VI(i,j)$ over the input LDR images is calculated



(a) Sum of MTBs

(b) Bitmap Movement Difference (BMD)

(c) Pixels detected as movement

Figure 3.3: Bitmap Movement Difference for ghost detection.

using Equation 3.2.

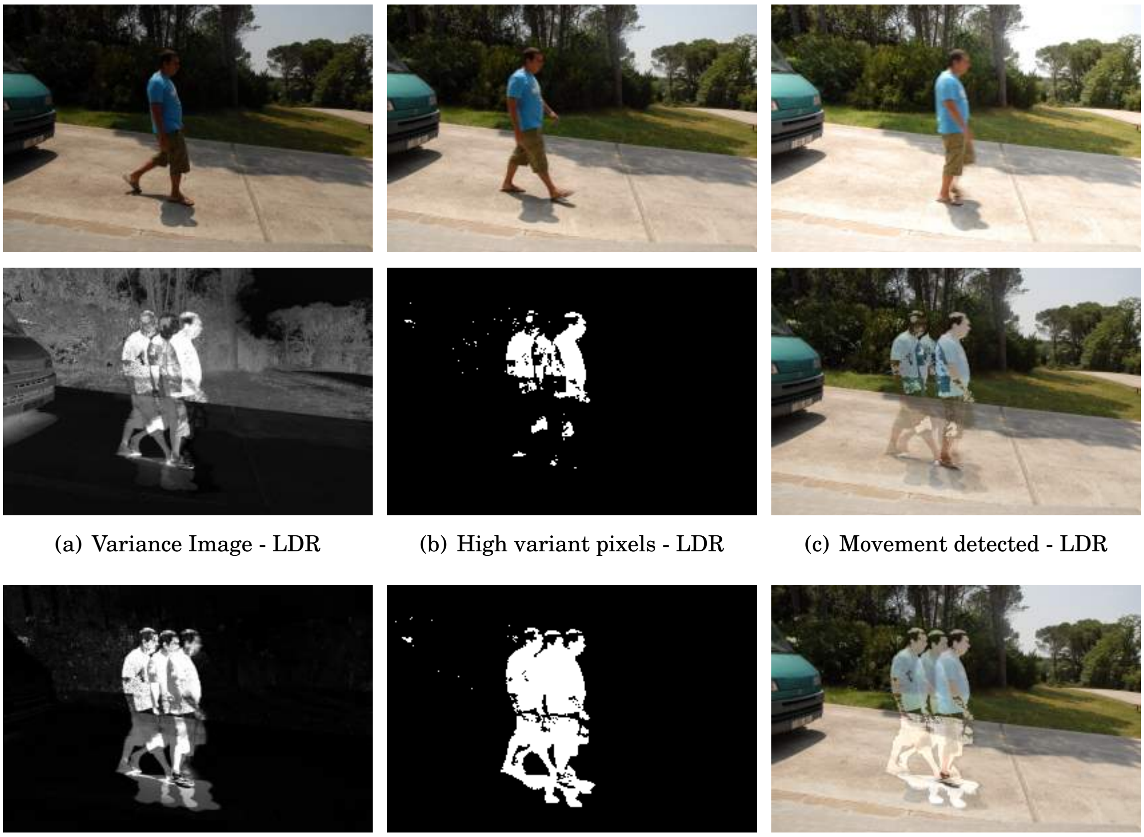
$$VI(i,j) = \frac{\sum_{n=0}^{N} w(I_n(i,j))I_n(i,j)^2 / \sum_{n=0}^{N} w(I_n(i,j))}{(\sum_{n=0}^{N} w(I_n(i,j))I_n(i,j))^2/(\sum_{n=0}^{N} w(I_n(i,j)))^2} \tag{3.2}$$

The variance could be influenced by saturation and noise. A weighting function w is used to minimize the influence of under- and over-exposed values. We use the hat function proposed by Khan *et al.* [KAR06] which is the same used in the final HDR composition.

High variance values are selected from the VI applying a threshold to obtain a binary image of dynamic pixels. Some high variant pixels remain in the binary image that corresponds to noise or very small movements. Morphological erosion and dilation are used to refine the final mask from noisy areas.

Pece [PK10] found that variance can fail in brightness peaks regions such as highlights, shining objects or direct sun light. The success of using variance in ghost detection also depends on the relation between colors of the dynamic object and the background. If the color of the dynamic object is similar to the background the method fails.

The results can be improved if images are transformed to radiance space instead of using LDR images. It requires a pre-calibration of the CRF using a static sequence of images. Notice the difference in the results in the second and third rows of Figure 3.4. Second row corresponds to the results using the original LDR sequence while the third row shows the results using the radiance instead of the original RGB values.



(a) Variance Image - LDR (b) High variant pixels - LDR (c) Movement detected - LDR

(d) Variance Image - Rad (e) High variant pixels - Rad (f) Movement detected - Rad

Figure 3.4: Variance of a sequence in ghost detection. The first row shows the LDR images, the second row the result of using the variance over the original LDR sequence and the third row shows the variance applied over the images in radiance space.

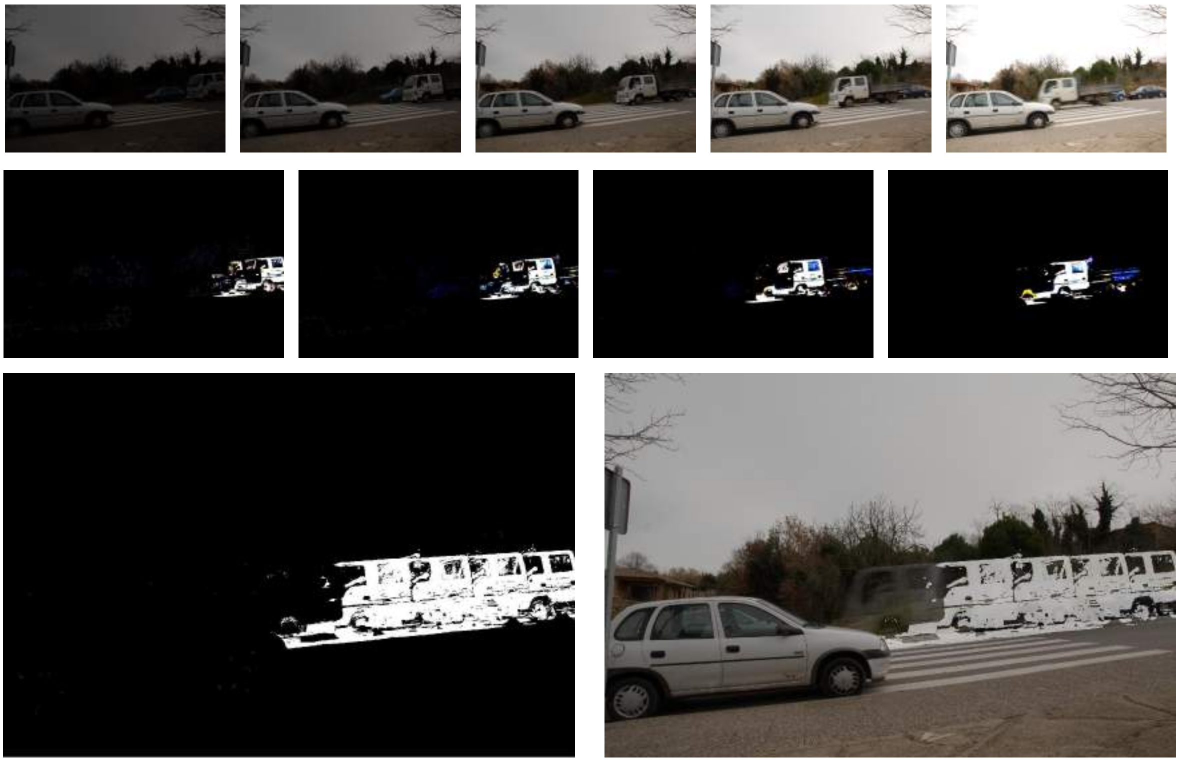
3.2.1.3 CRF-based pixels prediction

Usually, multiple exposure sequences are ordered increasingly by their exposure times, from darker to brighter. Sidibe *et al.* [SPS09] proposed a simple approach based on the fact that the CRF is monotonically increasing. Pixels from consecutive images must be related by the same relation as their exposure times. In a sequence of N exposures $\forall n \in [1...N]$, for the exposure times $\Delta t_n < \Delta t_n + 1$, we can assume that the following relation is valid for any pixel in (i,j):

$$I_{(i,j),n} \leq I_{(i,j),n'} \tag{3.3}$$

Pixels breaking this relation might be considered as movement. This method not only detects movements but also the unexpected variation of a pixel color. Gallo *et al.* [GGC$^+$09] improve this result assuming a linear relation $y = x + ln(EV)$ between images and using a threshold to select pixels far from this line.

Figure 3.5 shows this technique applied on a sequence ordered increasingly by the exposure time. The second row shows pixels that does not follow an increasing order between consecutive images. Figure 3.5(a) shows the interception of the partial differences and Figure 3.5(b) shows the pixels marked as movement in the sequence.



(a) Interception of partial differences

(b) Movement detected

Figure 3.5: Pixels breaking exposure order [SPS09].

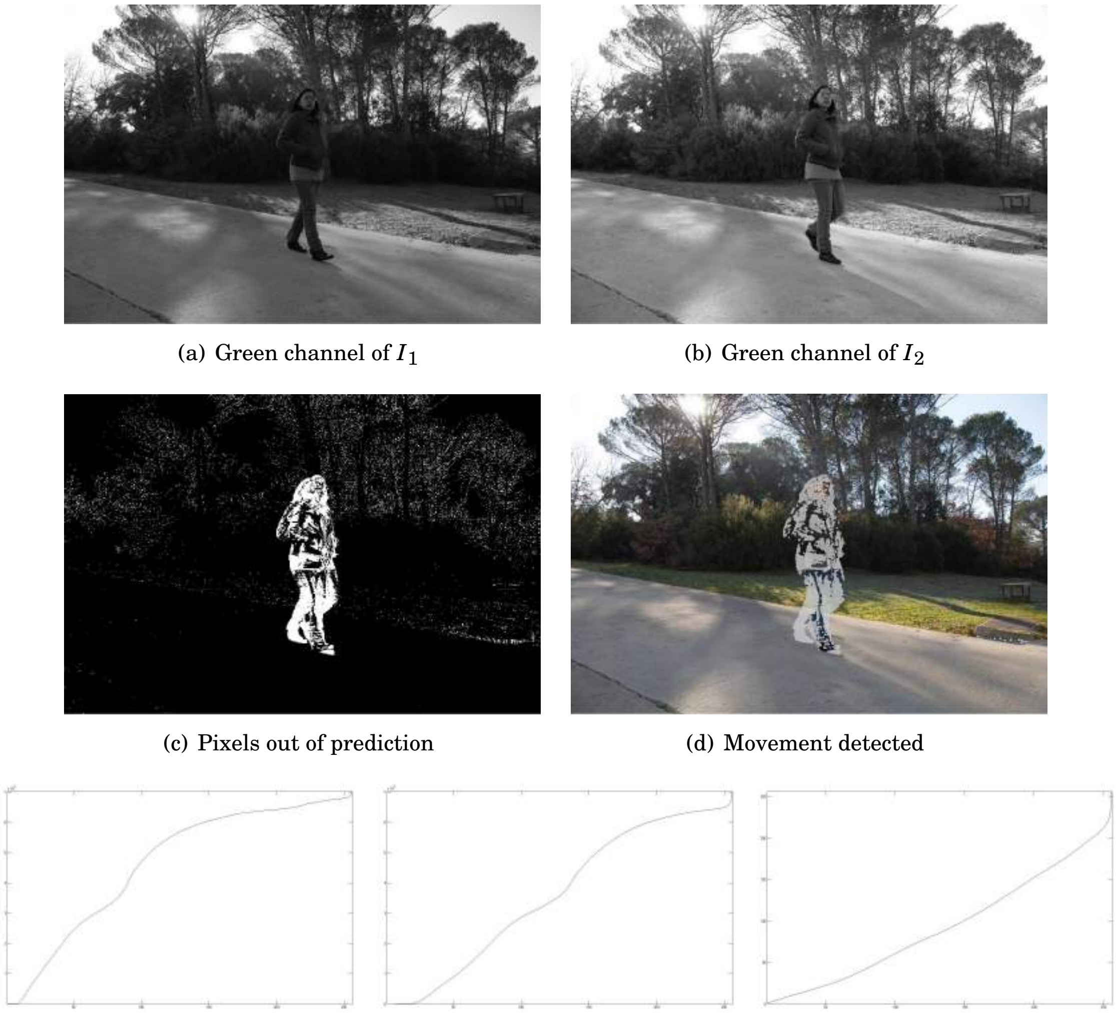
The relation in equation 3.3 is necessary but not enough to guarantee motion detection. The example in the sequence from Figure 3.5 shows a clear example when this relation is not enough. The last displacement of the car is not detected because it is white, which means that pixels are brighter than in the previous image so the equation 3.3 is satisfied. This method is not effective to detect complete movement of an object with more intensity than the background.

Gallo *et al.* [GGC^{+}09] assumed a linear relation between the images and used a threshold to select pixels in movement but problems persist because this assumption is rarely true. Grossberg and Nayar [GN03a] proposed a method to get the intensity mapping function (IMF) τ. The IMF is derived from the relation between the images based on comparing the cumulative histograms of consecutive exposures instead of comparing the pixel values, as described in equation 3.4.

$$I_2 = \tau(I_1) = g^{-1}(kg(I_1)) \tag{3.4}$$

where k expresses the relation between radiance in both images. Given the histogram of one image, the histogram of the second is necessary and sufficient to determine the intensity mapping function. The area of the image with intensities in the range $[0, I]$ is given by a function $H(I)$. Being h the continuous histogram, this can be expressed like:

$$H(I) = \int_0^I h(u)du$$



(a) Green channel of I_1

(b) Green channel of I_2

(c) Pixels out of prediction

(d) Movement detected

(e) Cumulative histogram H_1

(f) Cumulative histogram H_2

(g) Image Mapping Function

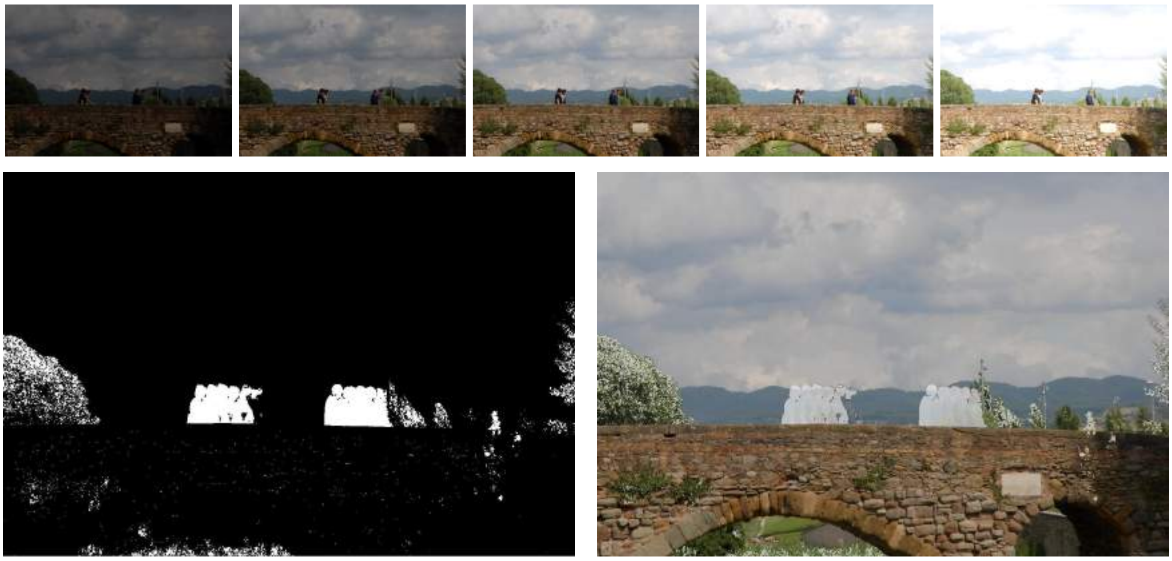
Figure 3.6: Cumulative histograms and IMF computation

This represents the cumulative histogram of the image (Figure 3.6(e) and 3.6(f)). Assuming ideal conditions, each intensity in I_2 maps to an intensity in I_1 defined by $I_1 = \tau(I_2)$ then the set of pixels in one image with intensity less than I_1 must be the same that pixels in the other with intensity less than I_2, these sets must also have equal area $H_1(\tau(I_2)) = H_2(I_2)$. Finding τ is possible using only the cumulative histograms from the two images. Histograms must be normalized and linear interpolation is used to invert the cumulative histogram.

Equation 3.2.1.3 replaces $I_1 = u$ and calculates τ from the cumulative histograms H_1 and H_2.

$$\tau(u) = {H_2}^{-1}(H_1(u))$$

The main contribution of this method is to show that the IMF can be determined without alignment because small scene movement does not change the histogram significantly. This makes it valid also for scene with large moving objects as long as histograms remain approximately constant in the sequence.



(a) IMF ghost masks combined

(b) Detected movement

Figure 3.7: Ghost map detected using IMF.

The masks calculated for consecutive image pairs are combined with a logical XOR in a final ghosting mask for the sequence that contains pixels affected by movement. Figure 3.7 shows five LDR images of a scene of people walking over a bridge, trees in the background are also moving because of the wind.

3.2.2 Image Registration

Image registration is an intense research field in computer graphics, vision and image processing. Registration is the process of matching two or more images of the same scene taken under different conditions (sensor, time, viewpoint or optical settings). During registration, one image is taken as reference and the rest (target images) are transformed until a match is found [ZFS05]. Surveys classifying and analyzing several techniques were presented previously [Bro92, MV98, ZF03, WPA09]. Registration techniques are traditionally used in remote sensing, medical imaging, cartography or computer vision. Most registration methods can be classified in two main groups:

- Feature-based methods: use salient structures (borders, lines or points) spread all over the image, recognizable in both images and invariable in time. Some of these methods were used for alignment in HDR reconstruction. However, dynamic objects in a sequence may not be well defined due to large exposure times and movement. Details in one exposure may not be visible in the rest because of over or under exposure.
- Intensity-based methods: attempt to match images without any explicit features detection, matching directly pixels intensities. A similarity measure (cost function) is defined between the source and the target and transformations are applied until the similarity measure reaches a maximum or minimum according to the function selected.

This work compares the results of intensity based registration using four similarity measures. The area of movement in the sequence is detected in the ghost detection step described in section 3.2.1. Figure 3.8

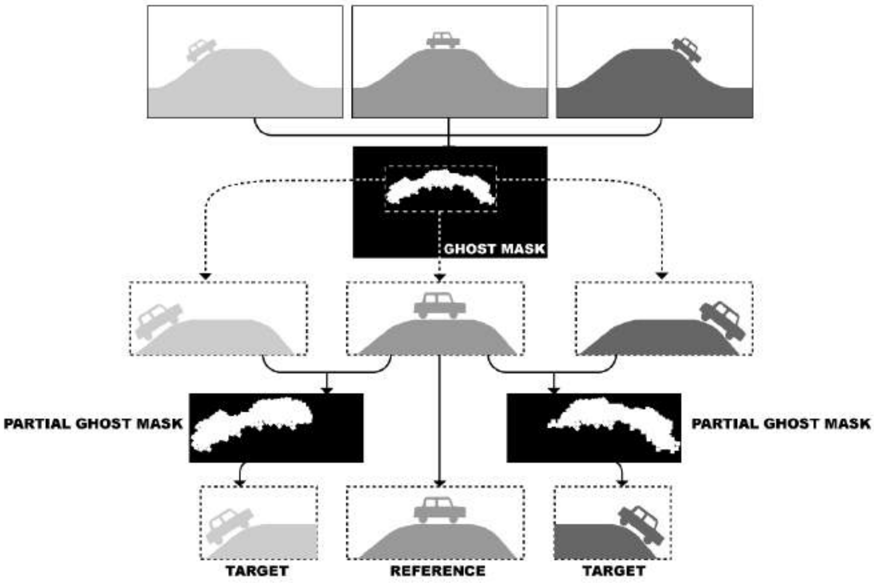


Figure 3.8: Defining target images from the ghost mask.

shows that the ghost mask can be cropped to focus only in regions of interest. The sub-images containing the movement detected can still be reduced. The partial ghost mask between pairs of them are smaller than the whole affected area. Differences between consecutive images are calculated generating partial ghost masks and sub-images are cropped according to the partial masks as shown in the last row of Figure 3.8.

All target images will be registered over the reference image. Registration is a highly time consuming task, the use of a pyramid of images helps to speed it up. In each level images are sampled down by a factor of two, as shown in Figure 3.9. Using the images from the last level, the target images are translated, rotated and scaled iteratively over the whole reference image evaluating the similarity measure in each iteration. Once the best match for the lowest level of the pyramid is found, we proceed with the higher level. Then we check only for transformations in an offset around the match point obtained in the previous step. The size of the offset depends on the size of target images but usually a vicinity of 10 percent of the dimensions of the image around the previous matching position gives good result images.

The following sections introduce the different similarity measures we implemented. The results of each of using them are presented in the section 3.3.2.

3.2.2.1 Sum of Squared Difference (SSD)

SSD is the simplest and the most intuitive way of measuring similarity [Anu70, SMA78, RS84]. The minimum value of SSD corresponds to the transformation T that better match the images I_n and $I_{n'}$.

$$SSD = \sum_{n=1}^{N} (I_n - T(I_{n'}))^2 \tag{3.5}$$

However, there are some problems using SSD for HDR registration because it is not invariant to changes in lighting conditions across the image sequence [Lew95].

3.2.2.2 Normalized Cross Correlation (NCC)

NCC assumes that corresponding intensities in the images have a linear relationship [CHH04]. This metric is used in images taken with the same device at different times [RR08]. We implemented it using an approach presented by Lewis [Lew95]. The best matching corresponds to the transformation that maximizes the NCC value.

$$NCC = \frac{\sum_{n=1}^{N} I_n \cdot T(I_{n'})}{\sqrt{\sum_{n=1}^{N} I_n^2 \cdot \sum_{n=1}^{N} T(I_{n'})^2}} \tag{3.6}$$

3.2.2.3 Mutual Information (MI)

MI is a measure of statistical dependence between two random variables or the amount of information that one variable contains about the other [RR08]. The Mutual Information can be defined like:

$$MI(I_n, I_{n'}) = H(I_n) - H(I_{n'}|I_n) = H(I_n) + H(I_{n'}) - H(I_n, I_{n'}) \tag{3.7}$$

Being H the Shannon[Sha48] entropy. For an image the entropy is calculated from the intensity histogram where N is the number of bins and p_i the value of each bin:

$$H = -\sum_{i=1}^{N} p_i log(p_i) \tag{3.8}$$

Since Viola *et al.* [VW95], several papers were presented mainly for multimodal image registration either minimizing joint entropy or maximizing mutual information [MFS14]. Most of them use a pyramidal approach to speed up the registration process [ZFS05]. We use Normalized Mutual Information (NMI) as similarity measure:

$$NMI(I_n, I_{n'}) = \frac{H(I_n) + H(I_{n'})}{H(I_n, I_{n'})} \tag{3.9}$$

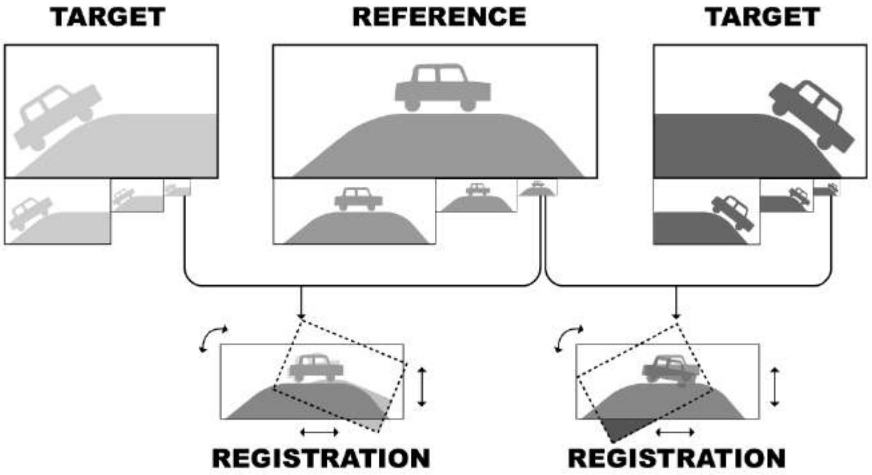


Figure 3.9: Image pyramid registration.

3.2.2.4 Median Threshold Bitmap (MTB)

It is the same MTB principle used for ghost detection in the previous section. It was adapted by Grosch [Gro06] to allow finding the correct transformation for registration once the transformation that minimizes bitmap differences is found. Despite this is not a popular measure for image registration, it has been used repeatedly in HDR.

3.2.3 HDR composition

This section shows how to proceed once the areas affected by movement are detected and registered over the reference image. The ghost mask classifies pixels in static or dynamic. The static pixels can be composed in an HDR image E using a weighted average [MP95, DM97, MN99] defined in Equation 3.10. The equation represents a weighted average of the radiance of the N images I_n in the sequence. The radiance is calculated using the inverse of the CRF (f^{-1}) and the exposure time Δt and we use the weighting function of Equation 3.11 proposed by Khan *et al.* [KAR06].

$$E(i,j) = \frac{\sum_{n=1}^{N} w(I_n(i,j))(\frac{f^{-1}(I_n(i,j))}{\Delta t_n})}{\sum_{n=1}^{N} w(I_n(i,j))} \tag{3.10}$$

$$w(I_n) = 1 - (2 \cdot \frac{I_n}{255} - 1)^{12} \tag{3.11}$$

Pece and Kautz [PK10] suggested to replace all pixels in the ghost mask by the best exposed LDR. The ghost mask contains pixels from dynamic objects all over the sequence, but usually movement does not affect more than two or three consecutive images. Gallo *et al.* [GGC$^+$09] calculated partial ghost mask for each pair of consecutive images. This ensures that only pixels from dynamic objects are excluded in each LDR image. However, under or over exposed regions might remain untreated in the result.

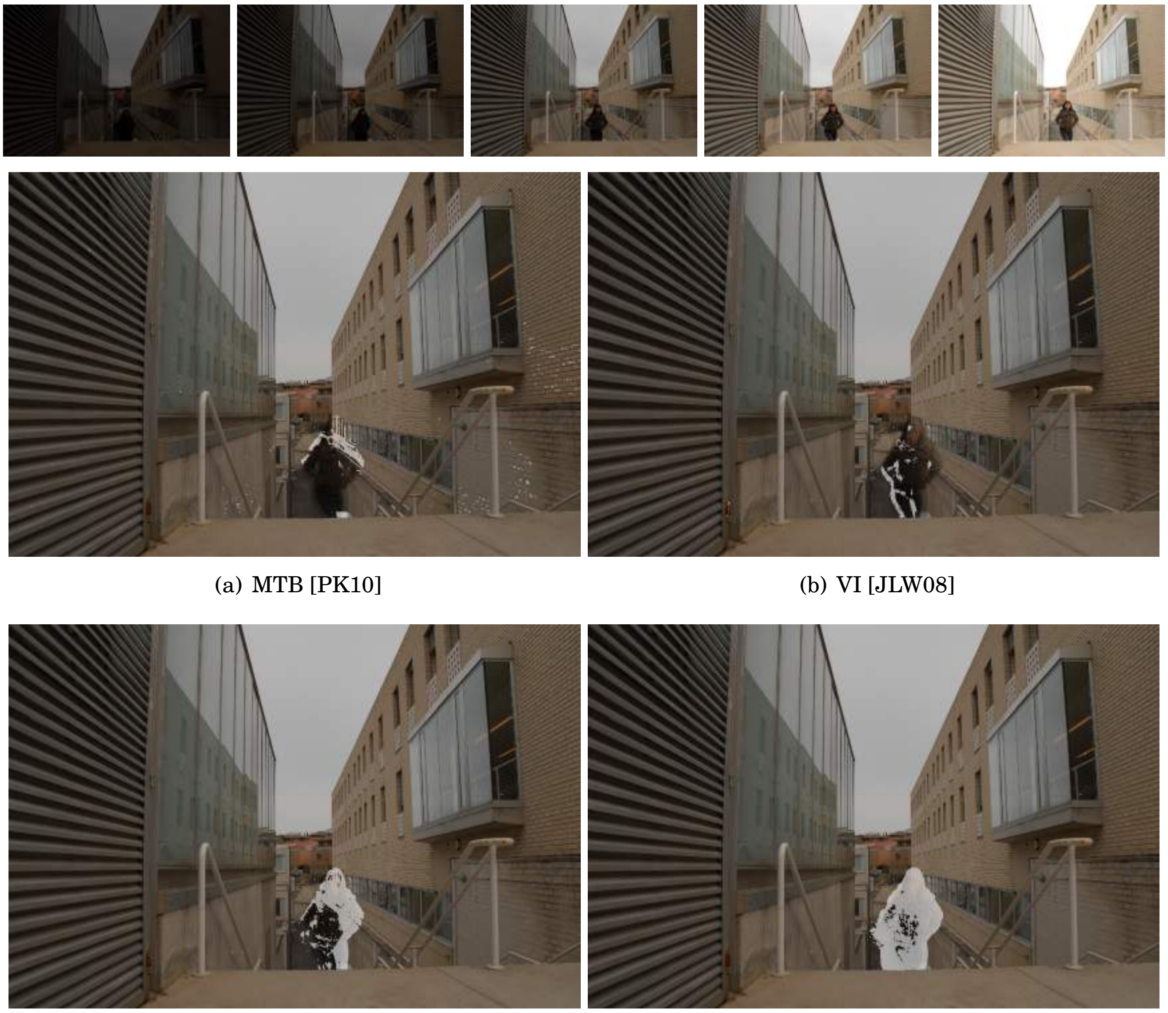
We can fill these regions with HDR values recovered for dynamic regions after registration. Even after registration it is important to prevent artifacts produced by small misalignment. In the HDR composition we only consider pixels that are correctly aligned. We calculate the difference between aligned images and the reference and discard the pixels that doesn't match.

3.3 Results and discussion

This section analyses and compares the results of the ghost detection and registration steps and the results of the proposed framework for full HDR recovery. All images for the test described were captured using a tripod and the auto bracketing function of a NIKON D200 camera. All parameters except the shutter speed were kept constant during the capture. Results vary depending on the ghost detection technique used since it determines the accuracy of the mask that represent dynamic content, the similarity measure used for registration and the thresholds used in each case.

3.3.1 Ghost Detection

Figures 3.11 and 3.10 show five LDR images from two different scenes and the ghost detection results obtained using the four implemented methods (see section 3.3.1). The set of images were carefully selected to show the weakness of each method.



(a) MTB [PK10]

(b) VI [JLW08]

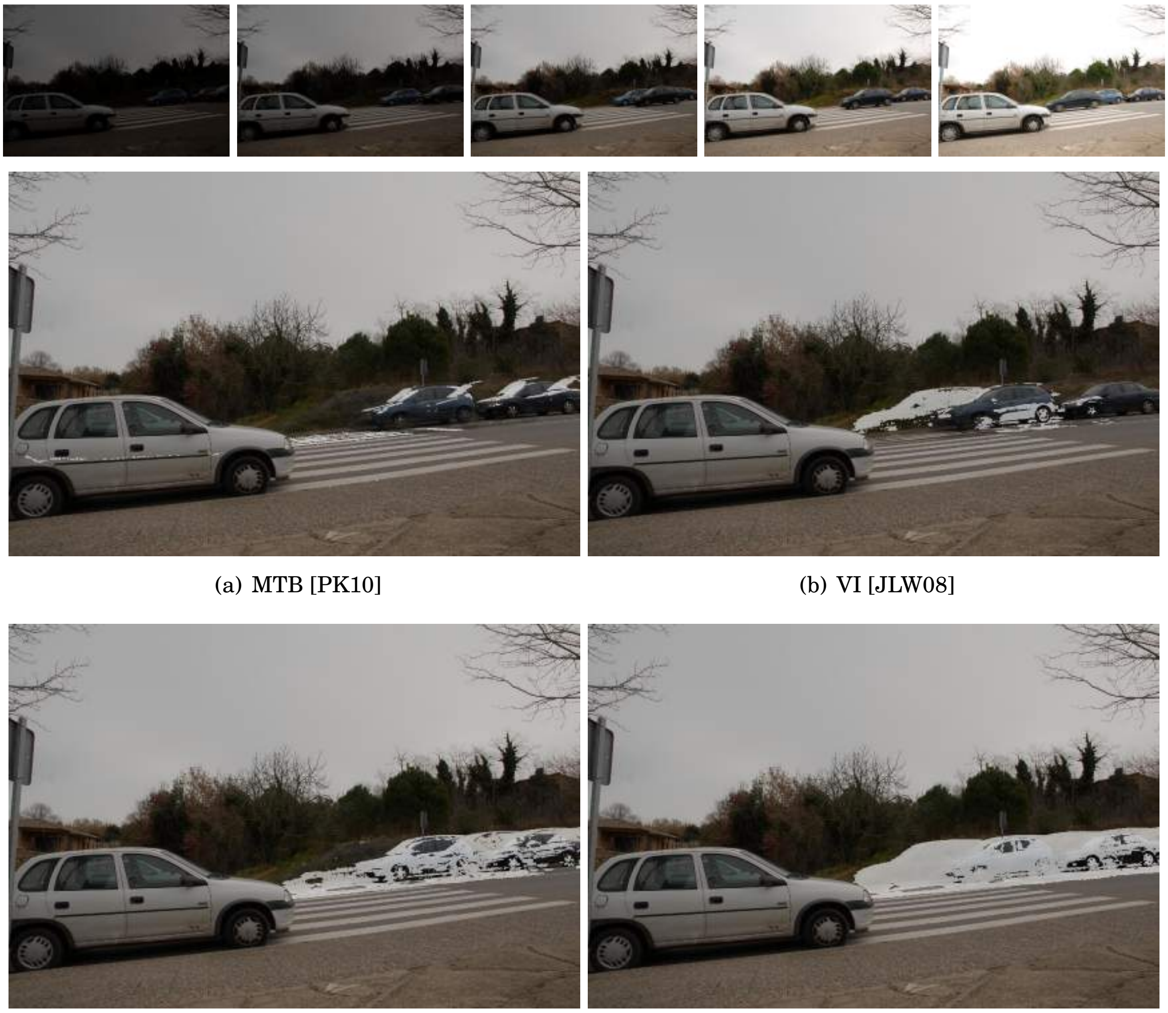
(c) CRF [SPS09]

(d) IMF Threshold

Figure 3.10: (Top) Set of input LDR images. (a-d) Results of ghost detection using the different methods presented in section 3.3.1. Note that IMF provides the best results.

The MTB method is fast and easy to implement. It is accurate for scenes with a rather bimodal brightness distribution. However, it fails if the dynamic object and the background are both smaller or bigger than the median value of intensities. In Figure 3.10(a) only pixels from non overlapping regions of clothes are detected, the same in Figure 3.11(a), where only the front glass is detected because the rest is very similar to the background.

The success of using variance in ghost detection also depends on the relation between colors of the dynamic object and the background. Variance method may fail in areas where the brightness is too high (highlights, shining objects, direct sun light) or when the movement is too slow that produces overlapping [PK10]. If the color of the dynamic object is similar to the background, the method fails [JLW08]. Figures 3.10(b) and 3.11(b) show an example where the variance method fails because dynamic pixels are very similar to the background. The threshold value is also an issue to take into account since the results directly depend on it.



(a) MTB [PK10]

(b) VI [JLW08]

(c) CRF [SPS09]

(d) IMF Threshold

Figure 3.11: Another example of ghost detection. (Top) Set of input LDR images. (a-d) Results of ghost detection using the different methods presented in section 3.3.1. In this case IMF is also the method with best results.

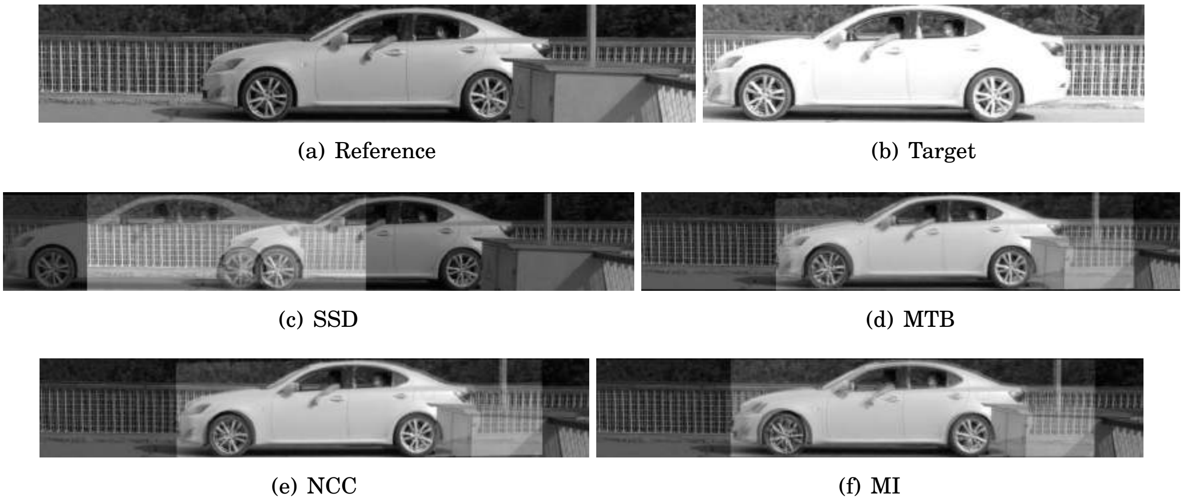
The CRF-based method proposed by Sidibe [SPS09] is very simple and does not depend on threshold values. It is based on the assumption that pixel intensities increase through the sequence, which is not always true in dynamic scenes. Any situation that breaks such assumption is detected as movement, like when the black car is moving in front of a dark background in Figure 3.11(c).

The most robust method in our tests was the one based on the IMF. Getting the IMF allows to predict values between pairs of images and to compare predicted values with the actual ones. The results of this method are the best in all tested cases. Nevertheless, the appropriate selection of the threshold value is crucial for the success of this method.

3.3.2 Image Registration

The ghost mask obtained in the previous step defines the area that we need to register. The results of image registration highly depend on the ghost detection step. The target images are selected from the bounding box of pixels affected by movement, like shown in Figure 3.3. If the ghost detection is not correct, bounding boxes do not correspond to dynamic objects and the registration step may fail too. The goal of the registration algorithm is to find the transformation that match the target image into the reference one. To find the matching we iterate over the search space in order to minimize (or maximize, depending in the similarity measure) the error between the target and the reference images.

Figures 3.12 and 3.13 shows an example of registration of two images from a sequence and the results obtained with the four implemented measures presented in section 3.2.2.



(a) Reference

(b) Target

(c) SSD

(d) MTB

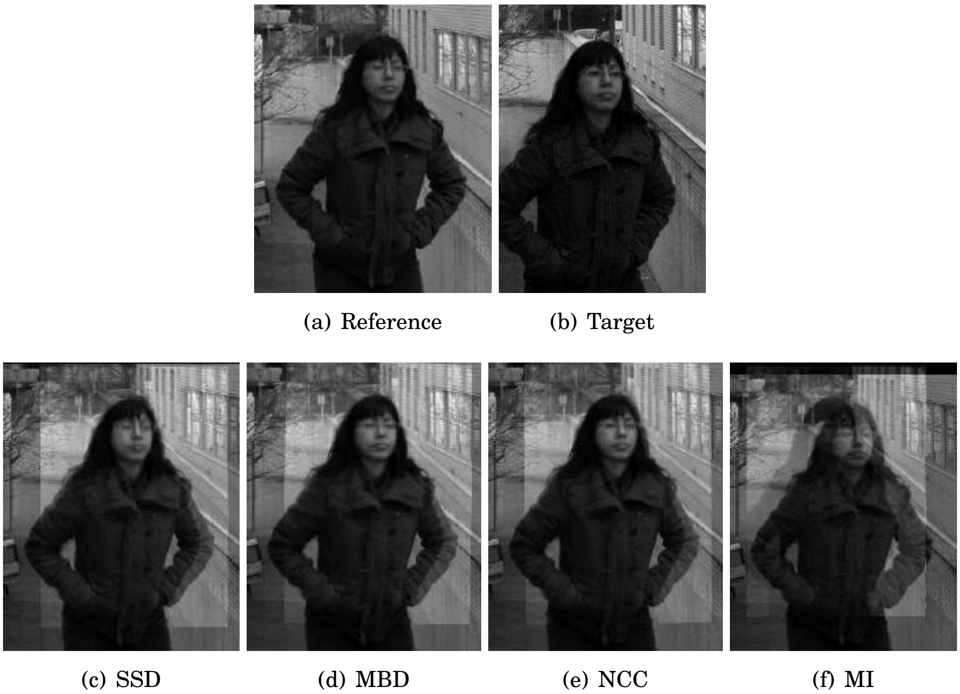
(e) NCC

(f) MI

Figure 3.12: Results of the registration of the reference sub-image (a) to the target sub-image (b). NCC is the method that achieves best results.

The original implementation was updated including translations and scaling. Evaluating the similarity measure for each possible translation is the most time consuming step of our method. We implemented CUDA kernels within our application in Matlab to perform this step.

Even when SSD finds a correct matching in some cases, it requires very specific conditions to work properly. It is not appropriate for registering images with very different exposures. The distance between images is affected by over- and under-exposed values.



(a) Reference (b) Target

(c) SSD (d) MBD (e) NCC (f) MI

Figure 3.13: Results of the registration of the reference sub-image (a) to the target sub-image (b). Note how NCC gets the best registration.

Our results showed that the MTB method is suitable for most cases. Its implementation is very simple and it is the fastest of the tested methods. However, similarly to variance for ghost detection, it fails if the background and the object have similar colors.

MI is a statistical measure and it is not affected by intensity changes. Most of results obtained using it are satisfactory but directly depend on the entropy of images which can be perturbed by over- or under-saturated pixels and the size of the moving objects. It is also the slowest method among the implemented ones.

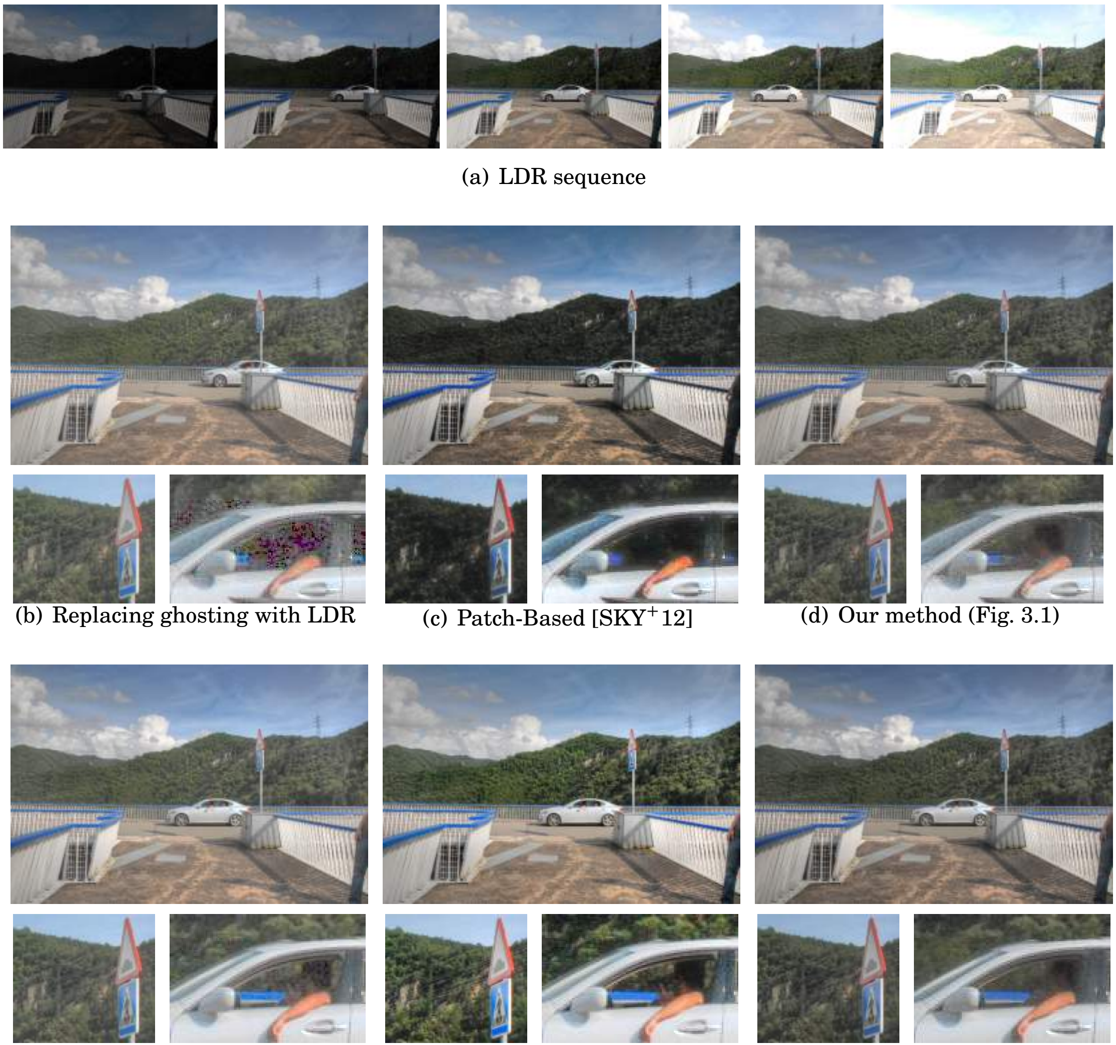
We obtained the best relation cost/performance when using NCC. It is insensitive to intensity changes which makes it a strong measure for images with different exposures.

3.3.3 HDR composition

Most previous work deal with dynamic areas in two different ways: (1) dynamic areas are removed from the final HDR image, reconstructing only the background [KAR06, PH08, GSL08, Mar09, SPS09, SL12] or (2) replaced by LDR content from the best exposure [Gro06, JLW08, GGC$^+$09, SPS09, PK10, HLL$^+$11, ADGM13]. The first approach is not coherent with the original scene because some content is missing in the HDR image. Replacing areas with LDR content may introduce over or under saturated areas in the dynamic areas of the resulting HDR image.

Algorithms using optical flow [KUWS03, ST04, MG10, ZBW11] aim to register the movement by tracking some feature points in the sequence. They are robust for slow movement (that introduces small displacements of pixels throw the sequence) but it can fails otherwise, the selection of features that could be tracked through a multiexposed sequence is also a problem to concern about.

Other approaches propose to match patches centered in every pixel of the image [SKY$^+$12, HGP12]. Such cases are robust when the reference image does not contains big under or over exposed areas but it can introduce artifacts in such areas otherwise.



(a) LDR sequence

(b) Replacing ghosting with LDR

(c) Patch-Based [SKY$^+$12]

(d) Our method (Fig. 3.1)

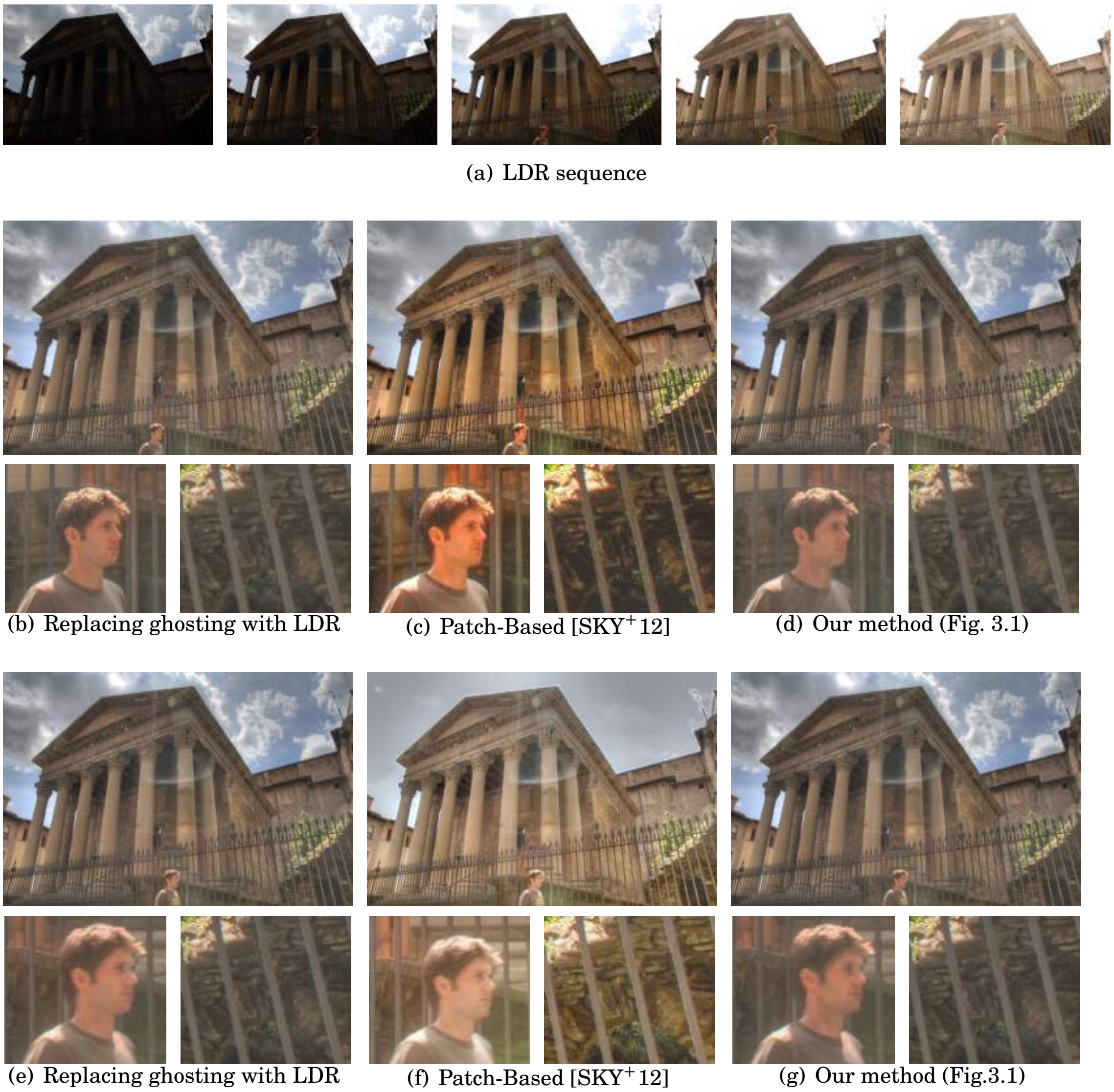
(e) Replacing ghosting with LDR

(f) Patch-Based [SKY$^+$12]

(g) Our method (Fig. 3.1)

Figure 3.14: HDR reconstruction of the sequence (a). (b)-(d) are the result of HDR reconstruction using the first image as reference.(e)-(g) are the result of HDR reconstruction using the third image as reference.

Our results are consistent in terms of the dynamic range recovered in the final HDR image and it preserve all the content from the original scene. Some artifact may persist after the registration because we only check for translation, rotation and scaling in the image plane. Other transformations like perspective or rotations in the objects space could be frequent and our method can not deal with them. To avoid such artifacts we check which of the registered pixels actually match to the reference image. We calculate the error between the registered image and the reference and exclude pixels with error bigger than a given threshold (5 percent for most of our tests).



(a) LDR sequence

(b) Replacing ghosting with LDR

(c) Patch-Based [SKY^{+}12]

(d) Our method (Fig. 3.1)

(e) Replacing ghosting with LDR

(f) Patch-Based [SKY^{+}12]

(g) Our method (Fig.3.1)

Figure 3.15: HDR reconstruction of the sequence (a). (b)-(d) are the result of HDR reconstruction using the third image as reference.(e)-(g) are the result of HDR reconstruction using the last image as reference.

Figure 3.14 helps to compare our solution to those that replace dynamic areas with content from the LDR reference only and a patch based approach. We conveniently selected as a reference the firs image (results (b) to (d)) in the sequence because it contains very under saturated values both in the mountains of the background and inside the car and the image in the middle of the sequence (results (e) to (g)). All images where tone mapped using the default tonemapper of Photomatix 4.2.1[3]. Replacing ghosting with the reference image introduces under saturated values in image (b) because pixels inside the car are under

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

saturated in the reference image.

The patch based method highly depends on the reference, since the matching takes place at patch level and includes a random component it may get wrong matching if the reference is under or over saturated like both inside the car or the background mountains. On the other hand when selecting the best exposure as reference, this method shows very accurate results. For some applications it is very important to get the same HDR values through the sequence independently of the reference image and this is not possible using most of the previous approaches.

In Figure 3.15 we compare our results obtained between our method and other two state of the art methods selecting the best exposure and the most saturated exposure of the sequence. Similarly to the previous sequence we can notice that our method is less sensitive to the reference choice. This is an important advantage if we want to apply this method to HDR video using different exposure sequences.

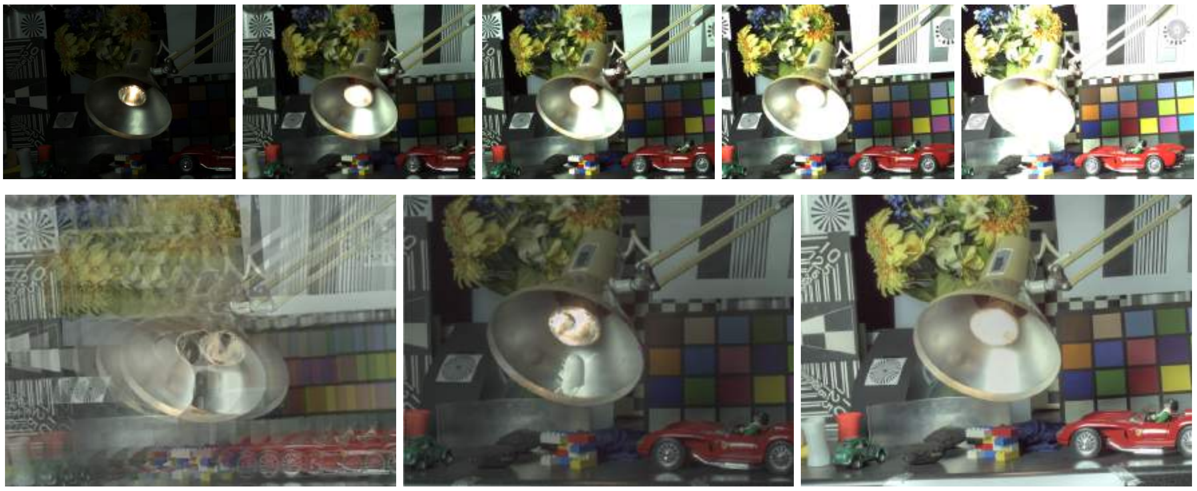
3.4 Summary

This chapter presented a method for HDR images generation of dynamic scenes. Our method detects areas affected by movement, matches them in a reference image and recovers HDR values from such areas in a sequence of LDR images. Promising results were obtained for scenes where dynamic objects were roughly rigid.

We implemented some state-of-the-art algorithms according to descriptions given by their authors. The selected algorithms were developed in Matlab using CUDA kernels. Regarding the ghost detection, we implemented four approaches obtaining different degrees of success. Even when all implemented techniques produce good results for some kind of scenes, the best results are most of the time obtained using a threshold over differences of pixels predicted with the IMF and the actual values. It is a very important step since all the process relies on its results. Any improvement of this step will have a positive impact on the reconstructed HDR scene.

CHAPTER 4

MULTISCOPIC HDR IMAGE SEQUENCE GENERATION



(a) Non aligned (b) Bätz *et al.* [BRG$^+$14] (c) Our Result

Figure 4.1: Set of LDR multiview images from the IIS Jumble data-set, courtesy of Bätz [BRG$^+$14]. The top row shows five views with different exposure values. The bottom row shows HDR images obtained without alignment (a), using Bätz's method (b), and using our proposed patch-match method (c).

4.1 Introduction

High Dynamic Range content generation has been recently moving from the 2D to 3D imaging domain introducing a series of open problems that need to be solved. 3D images are displayed in two main ways: either from two views for monoscopic displays with glasses or from multiple views for auto-stereoscopic displays. Most of current auto-stereoscopic displays accept from five to nine different views [LLR13]. To our knowledge, HDR auto-stereoscopic displays do not exist yet. However, HDR images are device independent and that means that they store values from the scene they represent independently of the device that will project them. Actually, HDR images existed long before the first HDR prototype appeared. Similarly to displaying tone-mapped HDR images on LDR displays, it is possible to feed LDR auto-stereoscopic displays with tone-mapped HDRs, one per each view required by the display.

Some of the techniques used to acquire HDR images from multiple LDR exposures have been recently extended for multiscopic images [TKS06, LC09, SMW10, BRR11, BLV$^+$12, OMLA13, OMLA14, BRG$^+$14, SDBRC14]. However, most of these solutions suffer from a common limitation: they rely on accurate dense stereo matching between images which is not robust in case of brightness difference between exposures [BVNL14].

This chapter presents a solution to combine sets of multiscopic LDR images into HDR content using image correspondences based on the Patch Match algorithm [BSFG09]. This algorithm was recently used by Sen *et al.* [SKY$^+$12] to build HDR images preventing from significant ghosting effects. Furukawa and Ponce [FP10] noticed the importance of improving the coherence of neighboring patches, an issue tackled in this chapter. Their results were promising for multi-exposure sequences where the reference image is moderately under exposed or saturated but it fails when the reference image has large under exposed or saturated areas.

The method described in this chapter improves Barnes *et al.* [BSFG09] approach for multiscopic image sequences (Figure 4.2). It also reduces the search space in the matching process and improves the incoherence of the matches in the original algorithm. Each image in the set of multi-exposed images is used as a reference; we look for matches in all the remaining images. Accurate matches allow to synthesize a set of HDR images, one for each view used for HDR merging.

The main contributions of this chapter can be summarized as follows:

- We provide an efficient solution to multiscopic HDR image generation.
- Traditional stereo matching produces several artifacts when directly applied on images with different exposures. We introduce the use of an improved version of patch-match to solve these drawbacks.
- Patch-match algorithm was adapted to take advantage of the epipolar geometry reducing its computational costs while improving its matching coherence drawbacks.

4.2 Patch-based multiscopic HDR Generation

The input for multiscopic HDR is a sequence of LDR images (formed of RAW or 8-bit RGB data) as shown in the first row of Figure 4.1. Each image is acquired from a different viewpoint, usually from a rig of cameras in a stereo distribution or multi-view cameras (see Figure 4.3). If the input images are in a 8-bit format, an inverse CRF needs to be recovered for each camera involved in the acquisition. This calibration step is performed only once, using a static set of images for each camera. The inverse of the CRFs is used to transform the input into radiance space. The remaining steps are performed using radiance space values instead of RGB pixels.

An overview of our framework is shown in Figure 4.2. The first step is to recover the correspondences between the **n** images of the set. We propose to use a nearest neighbor search algorithm (see section 4.2.1) instead of a traditional stereo matching approach. Each image acts like a reference for the matching

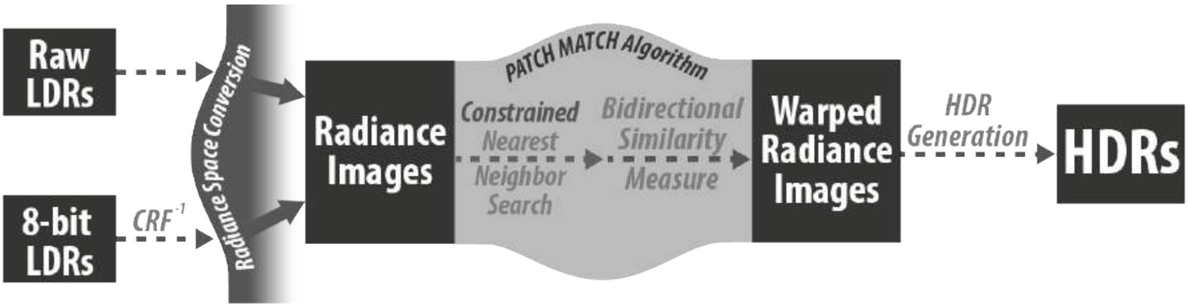


Figure 4.2: Proposed framework for multiscopic HDR Generation. It is composed by three main steps: (1) radiance space conversion, (2) patch match correspondences search and (3) HDR generation.

process. The output of this step is **n-1** warped images for each exposure. Afterward, the warped images are combined into an output HDR image for each view (see section 4.2.2).

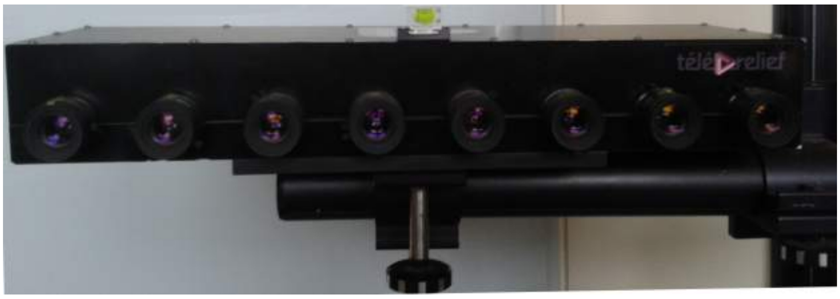


Figure 4.3: The Octo-cam, multi-view camera prototype.

4.2.1 Nearest Neighbor Search

For a pair of images I_r and I_s, we compute a Nearest Neighbor Field (NNF) from I_r to I_s using an improved version of the method presented by Barnes *et al.* [BSFG09]. NNF is defined over patches around every pixel coordinate in image I_r for a cost function **D** between two patches of images I_r and I_s. Given a patch coordinate $\mathbf{r} \in I_r$ and its corresponding nearest neighbor $\mathbf{s} \in I_s$, $NNF(\mathbf{r}) = \mathbf{s}$. The values of NNF for all coordinates are stored in an array with the same dimensions as I_r.

We start initializing the NNFs using random transformation values within a maximal disparity range on the same epipolar line. Consequently the NNF is improved by minimizing **D** until convergence or a maximum number of iterations is reached. Two candidate sets are used in the search phase as suggested by [BSFG09]:

1. *Propagation* uses the known adjacent nearest neighbor patches to improve NNF. It quickly converges but it may fall in a local minimum.

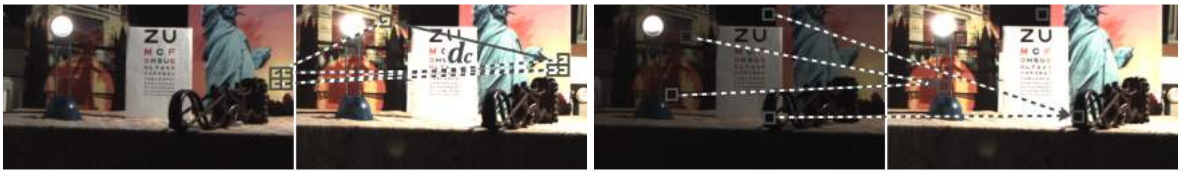
2. *Random search* introduces a second set of random candidates that are used to avoid local minimums. For each patch centered in pixel v_0, the candidates u_i are sampled at an exponentially decreasing

distance v_i previously defined by Barnes *et al.* :

$$u_i = v_0 + w\alpha^i R_i \tag{4.1}$$

where R_i is a uniform random value in the interval [-1,1], w is the maximum value for disparity search and α is a fixed ratio (1/2 is suggested).

Taking advantage of the epipolar geometry both search accuracy and computational performance are improved. Geometrically calibrated images allow to reduce the search space from 2D to 1D domain, consequently reducing the search domain. The random search of matches only operates in the range of maximum disparity in the same epipolar line (1D domain), avoiding to search in 2D space. This reduces significantly the number of samples to find a valid match.



(a) Coherency

(b) Completeness

Figure 4.4: Patches from the reference image (Left) look for their NN in the source image (Right). Even when destination patches are similar in terms of color, matches may be wrong because of geometric coherency problems. Images from the 'Octocam' dataset courtesy of [BLV+12]

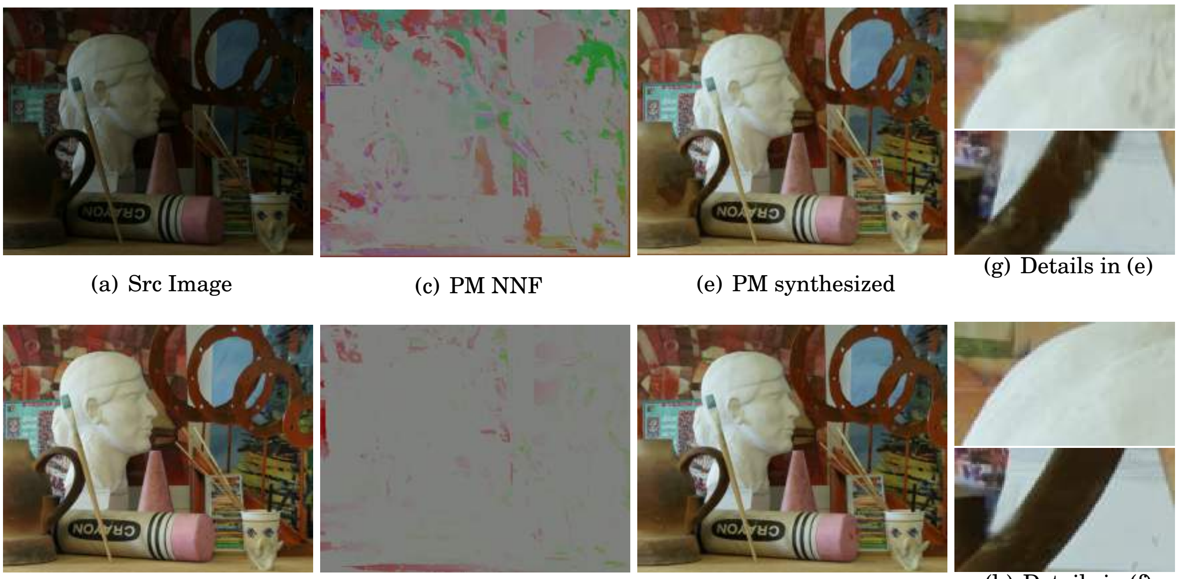
However, the the original NNFs approach [BSFG09] used in the patch match algorithm has two main disadvantages, the lack of completeness and coherency. This problems are illustrated in Figure 4.4 and the produced artifact in Figure 4.5. The lack of coherency refers to the fact that two neighbor pixels in the reference image, may match two separated pixels in the source image like in Figure 4.4(a). Completeness issues refer to more than one pixel in the reference image matching the same correspondence in the source image, like shown in Figure 4.4(b).

To overcome this drawback we propose a new distance cost function D by incorporating a coherence term to penalize matches that are not coherent with the transformation of their neighbors. Both Barnes *et al.* [BSFG09] and Sen *et al.* [SKY+12] use the Sum of Squared Differences (SSD) described in equation 4.3 where **T** represents the transformation between patches of **N** pixels in images I_r and I_s. We propose to penalize matches with transformations that differ significantly form it neighbors by adding the coherence term **C** defined in equation 4.4. The variable d_c represents the Euclidean distance to the closest neighbor's match and Max_{disp} is the maximum disparity value. This new cost function forces pixels to preserve coherent transformations with their neighbors.

$$D = SSD(r,s)/C(r,s) \tag{4.2}$$

$$SSD = \sum_{n=1}^{N}(I_r - T(I_s))^2 \tag{4.3}$$

$$C(r,s) = 1 - d_c(r,s)/Max_{disp} \tag{4.4}$$



(a) Src Image (c) PM NNF (e) PM synthesized (g) Details in (e)

(b) Ref Image (d) Ours NNF (f) Ours synthesized (h) Details in (f)

Figure 4.5: Matching results using original Patch Match [BSFG09] (Up) and our version (Down) for two iterations using 7x7 patches. Images in the 'Art' dataset courtesy of [Mid06].

Figures 4.5(d) and 4.5(f) correspond to the results including the improvements presented in this section. Figures 4.5(c) and 4.5(d) show a color representation of the NNFs using HSV color space. The magnitude of the transformation vector is visualized in the saturation channel and the angle in the hue channel. Areas represented with the same color in the NNF color representation mean similar transformation. Objects in the same depth may have similar transformation. Notice that the original Patch Match [BSFG09] finds very different transformations for neighboring pixels of the same objects and produces artifacts in the synthesized image.

4.2.2 Image alignment and HDR Generation

The nearest neighbor search step finds correspondences among all the different views. The matches are stored in a set of $n^2 - n$ NNFs. This information allows to generate $n-1$ images with different exposures realigned on each view. The set of aligned multiple exposures per view feeds the HDR generation algorithm to produce a HDR image for every view (see Figure 4.6).

Despite the improvements in the cost function presented in the previous section, NNF may not be coherent in occluded or saturated areas. However, even in such cases a match to a similar color is found between each pair of images $I_r; I_s$. This makes possible to synthesize images for each exposure corresponding to each view.

Direct warping from the NNFs is an option, but it may generate visible artifacts as shown in Figure 4.7. We use Bidirectional Similarity Measure (BDSM) (Equation 4.5), proposed by Simakov *et al.* [SCSI08] and used by Barnes *et al.* [BSFG09], which measures similarity between pairs of images. The warped images are generated as an average of the patches that contribute to a certain pixel. It is defined in equation 4.5 for every patch $\mathbf{Q} \subset I_r$ and $\mathbf{P} \subset I_s$, and a number $\mathbf{N}$ of patches in each image respectively. It consists of two terms: *coherence* that ensures that the output is geometrically coherent with the reference and

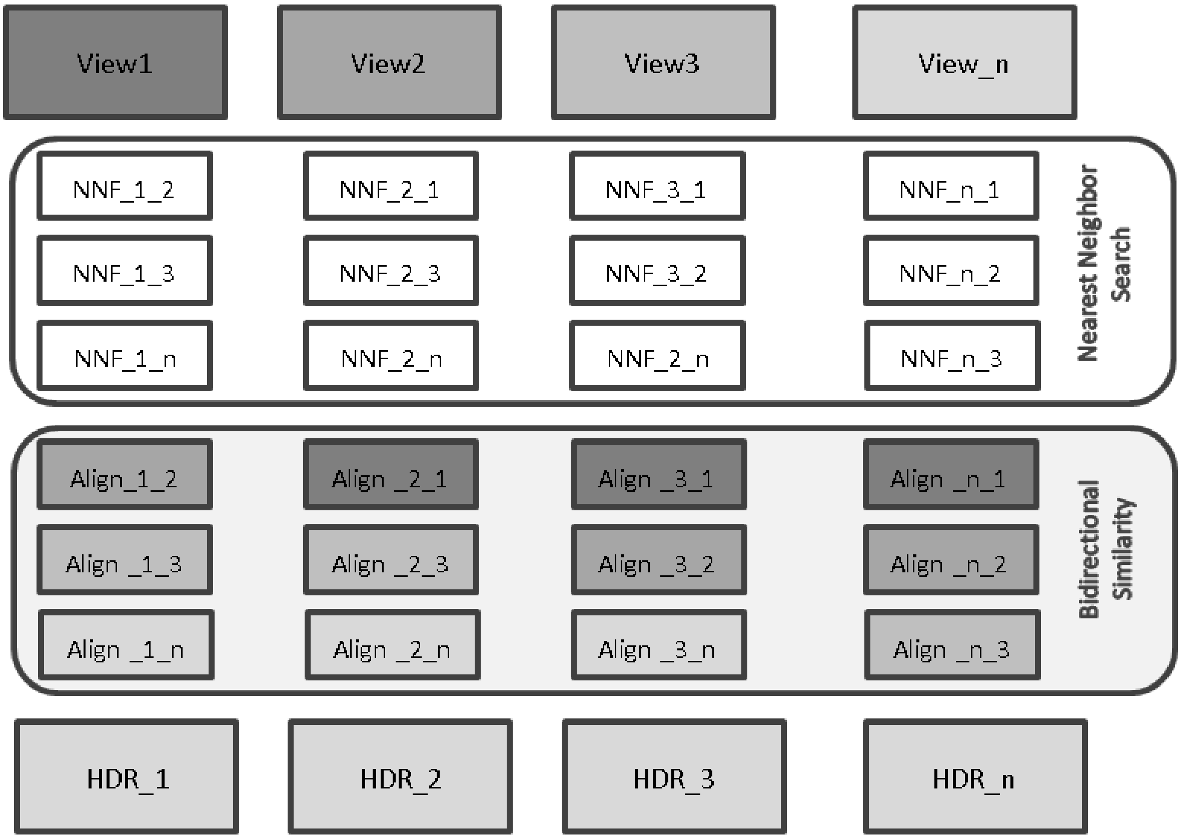


Figure 4.6: The nearest neighbor search step generates $n^2 - n$ NNFs. This is used to generate $n-1$ aligned images per view, using Bidirectional Similarity. HDR images are generated using input from each view and the corresponding aligned images.

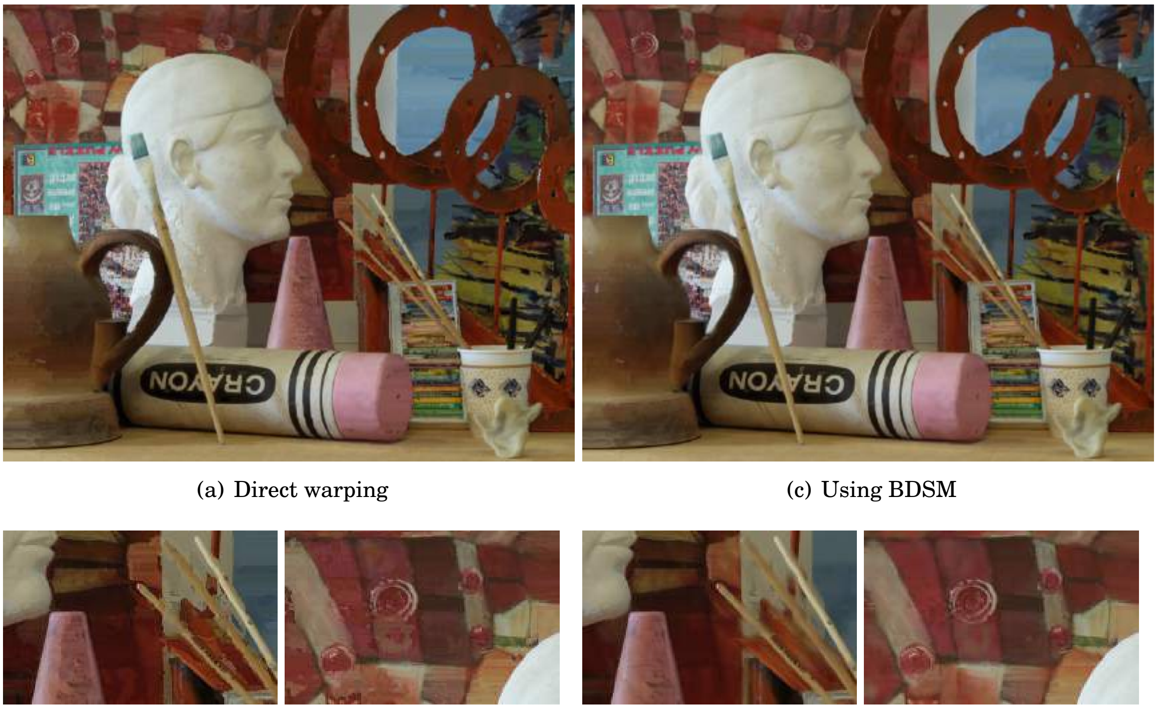
completeness that ensures that the output image maximizes the amount of information from the source image:

$$d(I_r, I_s) = \overbrace{\frac{1}{N_{I_r}} \sum_{Q \subset I_r} \min_{P \subset I_s} D(Q,P)}^{d_{completeness}} + \overbrace{\frac{1}{N_{I_s}} \sum_{P \subset I_s} \min_{Q \subset I_r} D(P,Q)}^{d_{coherence}} \tag{4.5}$$

This improves the results by using bidirectional NNFs ($I_r \to I_s$ and backward, $I_r \leftarrow I_s$). It is more accurate to generate images using only two iterations of nearest neighbour search and bidirectional similarity than four iterations of neighbour search and direct warping. Table 4.1 shows some values of Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) of images warped like the ones in Figure 4.7 comparing to the reference LDR image. The values in the table corresponds to the average MSE and PSNR calculated per each channel of the images in L*a*b* color space, using equations 4.6 and 4.7 respectively.

$$MSE(I, I') = \frac{1}{N} \sum_{1=0}^{N} (I(i) - I'(i))^2 \tag{4.6}$$

$$PSNR(I, I') = 10 log_{10}(\frac{MAX^2}{MSE(I, I')}) \tag{4.7}$$



(a) Direct warping

(c) Using BDSM

(b) Details in (a)

(d) Details in (c)

Figure 4.7: Images 4.7(a) and 4.7(c) are both synthesized from the pair in Figure 4.5. Image 4.7(a) was directly warped using values only from the NNF of Figure 4.5(c), which corresponds to matching 4.5(a) to 4.5(b). Image 4.7(c) was warped using the BDSM of Equation 4.5 which implies both NNFs of Figures 4.5(c) and 4.5(d). Notice the artifacts on the edges and the sharp changes on (a) and (b).

Table 4.1:

Iterations	Direct Warp		Bidirectional	
	MSE	PSNR	MSE	PSNR
1	2.42199	41.8613	2.17003	42.3734
2	2.40966	41.8764	2.17195	42.3762
4	2.4137	41.8728	2.16708	42.3846

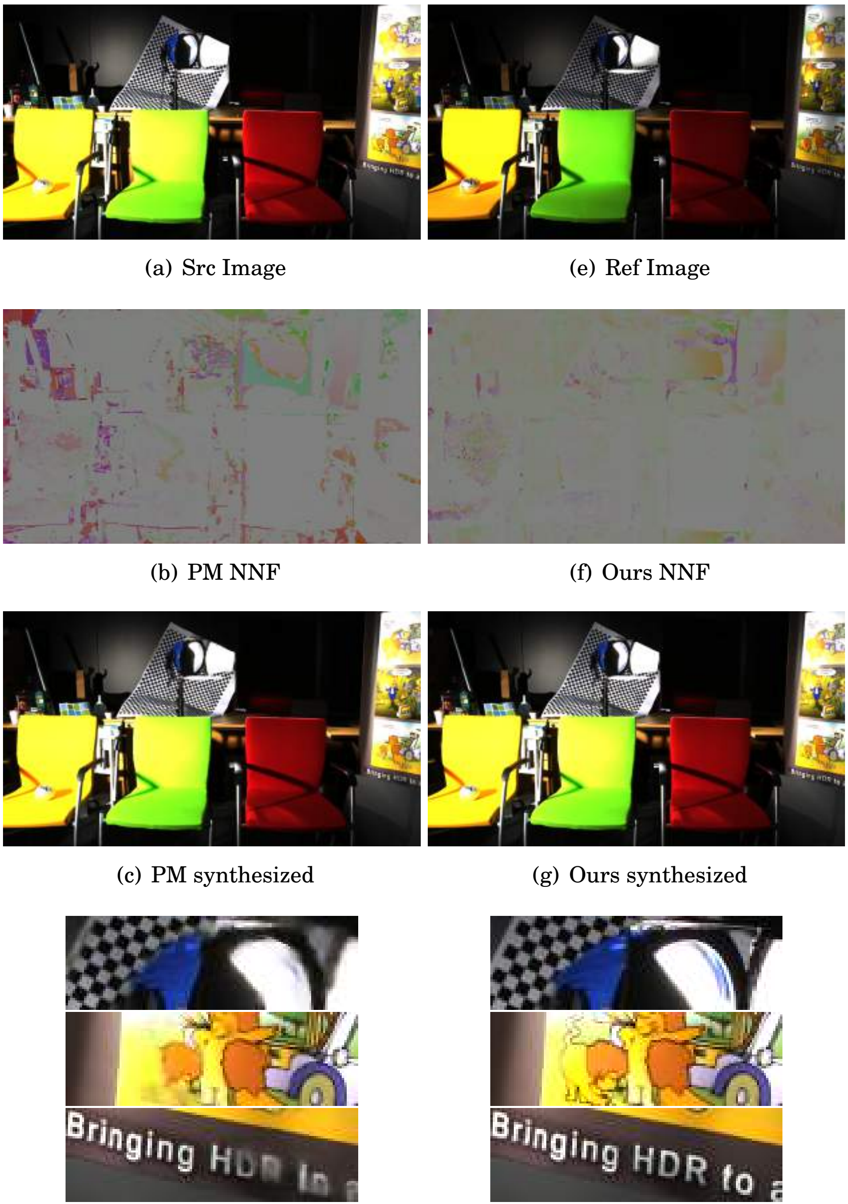
Since the matching is totally independent for pairs of images, it was implemented in parallel. Each image matches the remaining other views. This produces **n-1** NNFs for each view. The NNFs are in fact the two components of the BDSM of equation 4.5. The new image is the result of accumulating pixel colors of each overlapping neighbor patch and averaging them.

$$E(i,j) = \frac{\sum_{n=1}^{N} w(I_n(i,j))(\frac{f^{-1}(I_n(i,j))}{\Delta t_n})}{\sum_{n=1}^{N} w(I_n(i,j))} \tag{4.8}$$

$$w(I_n) = 1 - (2\frac{I_n}{255} - 1)^{12} \tag{4.9}$$

The HDR images (one HDR per view) are generated using a standard weighted average [MP95, DM97, MN99] as defined in Equation 4.8 and the weighting function of Equation 4.9 proposed by Khan *et al.* [KAR06] where I_n represents each image in the sequence, w corresponds to the weight, f is the CRF, Δt_n is the exposure time for the I^{th} image of the sequence.

4.3 Results and discussion



(a) Src Image

(e) Ref Image

(b) PM NNF

(f) Ours NNF

(c) PM synthesized

(g) Ours synthesized

(d) Details in (e)

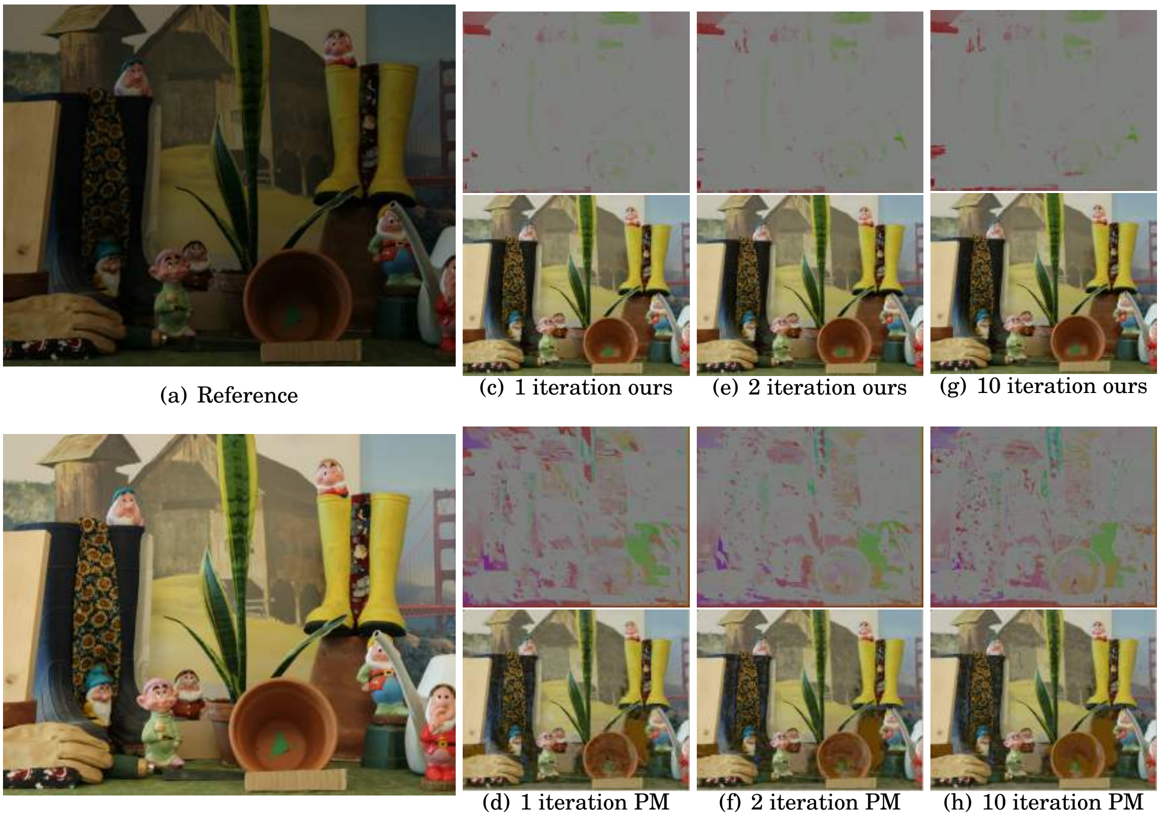
(h) Details in (f)

Figure 4.8: Comparison between original Patch Match and our method (2 iterations, 7x7 patch size). Images 4.8(b) and 4.8(f) show the improvement on the coherence of the NNF using our method. Images courtesy of [SDBRC14].

Five data-sets were selected in order to demonstrate the robustness of our results. For the set 'Octo-cam' all the objectives capture the scene at the same time and synchronized shutter speed. For the rest of data-sets the scenes are static. This avoids the ghosting problem due to dynamic objects in the scene. In all figures of this section we use the different LDR exposures for display purposes only, the actual matching is performed in radiance space.

The 'Octo-cam' data-set are eight RAW images with 10-bit of color depth per channel. They were acquired simultaneously using the Octo-cam [PcPD$^+$10] with a resolution of 748x422 pixels. The Octo-cam is a multi-view camera prototype composed by eight objectives horizontally disposed. All images are taken at the same shutter speed (40 ms) but we use three pairs of neutral density filters that reduce the exposure dividing by 2, 4 and 8 respectively. The exposure times for the input sequence are equivalent to 5, 10, 20 and 40 ms respectively [BLV$^+$12]. The objectives are synchronized so all images corresponds to the same time instant.

The sets 'Aloe', 'Art' and 'Dwarves' are from the Middlebury web site [Mid06]. We selected images that were acquired under fixed illumination conditions with shutter speed values of 125, 500 and 2000 ms for 'Aloe'and 'Art' and values of 250, 1000 and 4000 ms for 'Dwarves'. They have a resolution of 1390 x 1110 pixels and were taken from three different views. Even if we have only 3 different exposures we can use the seven available views by alternating the exposures like shown in Figure 4.11.



(a) Reference

(c) 1 iteration ours

(e) 2 iteration ours

(g) 10 iteration ours

(b) Source

(d) 1 iteration PM

(f) 2 iteration PM

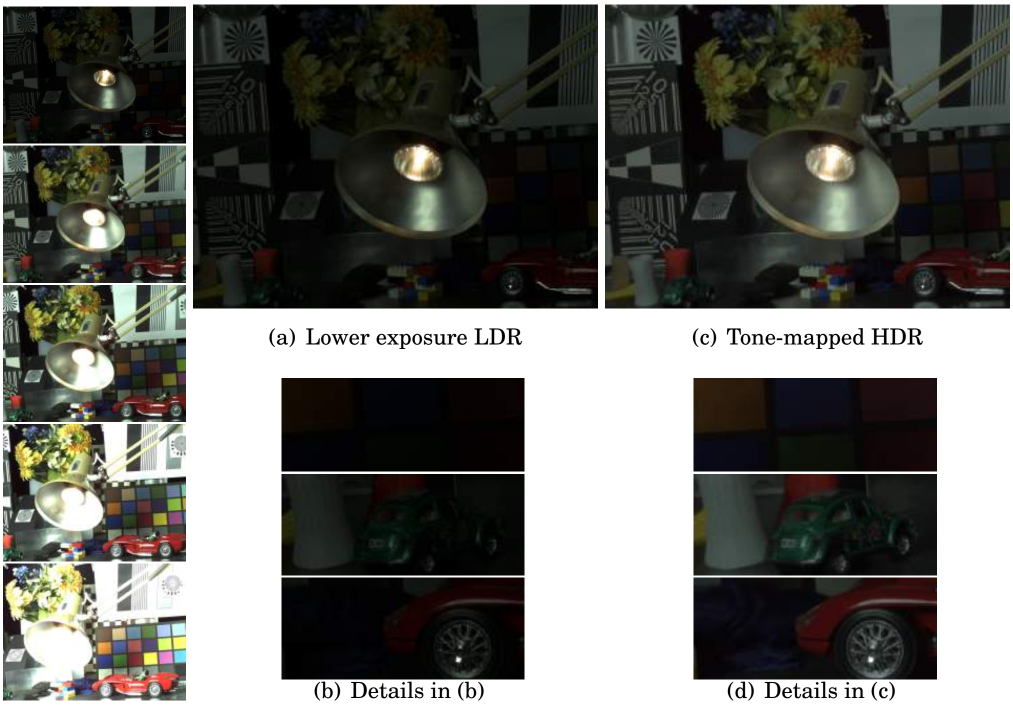
(h) 10 iteration PM

Figure 4.9: Two images from the 'Dwarves' set of LDR multi-view images from Middlebury [Mid06]. Our method with only two iterations achieve very accurate matches. Notice that the original patch match requires more iterations to achieve good results in fine details of the image.

The last two data-sets were acquired from two of the state-of-the-art papers. Bätz *et al.* [BRG$^+$14] shared their image data set (IIS Jumble) at a resolution of 2560x1920 pixels. We selected five different views from their images. They where acquired at shutter speeds of 5, 30, 61, 122 and 280 ms respectively. Pairs of HDR images like the one in Figure 4.8, both acquired from a scene and synthetic examples come from Selmanovic *et al.* [SDBRC14]. For 8-bit LDR data sets, the CRF is recovered using a set of multiple exposure of a static scene. All LDR images are also transformed to radiance space for fair comparison with other algorithms.

Figure 4.8 shows a pair of images linearized from HDR images courtesy of Selmanovic *et al.* [SDBRC14] and the comparison between the original PM from Barnes *et al.* [BSFG09] and our method including the coherence term and epipolar constrains. The images in Figures 4.8(b) and 4.8(f) represent the NNF. They are encoded into an image in HSV color space. Magnitude of the transformation vector is visualized in the saturation channel and the angle in the hue channel. Notice that our results represent more homogeneous transformations, represented in gray color. Images in Figure 4.8(c) and 4.8(g) are synthesized result images for the **Ref** image obtained using pixels only from the **Src** image. The results correspond to the same number of iterations (2 in this case). Our implementation converges faster producing accurate results in less iterations than the original method.

All the matching and synthesizing processes are performed in radiance space. They were converted to LDR using the corresponding exposure times and the CRF for display purposes only. The use of an image synthesis method like the BDSM instead of traditional stereo matching allows us to synthesize values also for occluded areas.


(a) Lower exposure LDR

(c) Tone-mapped HDR

(b) Details in (b)

(d) Details in (c)

Figure 4.10: Details of the generated HDR image corresponding to a dark exposure. Notice that under-exposed areas, traditionally difficult to recover, are successfully generated without visible noise or misaligned artifacts. IIS Jumble data-set courtesy of [BRG$^+$14].

Figure 4.9 shows the NNFs and the images synthesized for different iterations of both our method and the original patch match. Our method converges faster and produce more coherent results than [BSFG09]. In occluded areas the matches may not be accurate in terms of geometry due to the lack of information. Even in such cases, the result is accurate in terms of color. After several tests, only two iterations of our method were enough to get good results while five iterations were recommended for previous approaches.

Figure 4.10 shows one example of the generated HDR corresponding to the lowest exposure LDR view in the IIS Jumble data-set. It is the result of merging all synthesized images obtained with the first view as reference. The darker image is also the one that contains more noisy and under-exposed areas. HDR values were recovered even for such areas and no visible artifacts appears. On the contrary, the problem of recovering HDR values for saturated areas in the reference image remains unsolved. When the dynamic range differences are extreme the algorithm does not provide accurate results. Future work must provide new techniques because the lack of information inside saturated areas does not allow patches to find good matches.

The inverse CRFs for the LDR images were calculated from a set of aligned multi-exposed images using the software RASCAL, provided by Mitsunaga and Nayar [MN99]. Figure 4.11 shows the result of our method for a whole set of LDR multi-view and differently exposed images. All obtained images are accurate in terms of contours, no visible artifacts comparing to the LDR were obtained.

Figure 4.12 shows the result of the proposed method in a scene with important lighting variations. The presence of the light spot introduces extreme lighting differences between the different exposures. For bigger exposures the light glows from the spot and saturate pixels not only inside the spot but also around it. There is not information in saturated areas and the matching algorithm does not find good correspondences. The dynamic range is then compromised in such areas and they remain saturated.

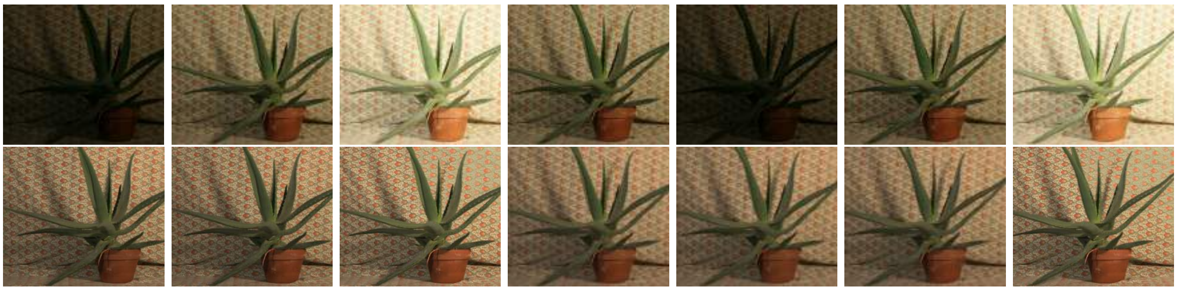


Figure 4.11: Up: 'Aloe' set of LDR multi-view images from Middlebury web page [Mid06]. Down: the resulting tone mapped HDR taking each LDR as reference respectively. Notice the coherence between the tone mapped HDR images.

Two of the dataset used in the tests provide aligned multiple exposures for each view, which allows to generate ground truth HDR images per view. Figure 4.13 shows the results of comparing some of our results to ground truth images using the HDR-VDP-2 metric proposed by Mantiuk *et al.* [MKRH11]. This metric provides some values to describe how similar two HDR images are.

The quality correlate Q is 100 for the best quality and gets lower for lower quality. Q can be negative in case of very large differences. The images at the right of each pairs in figure 4.13 are the probability of detection map. It shows where and how likely a difference will be noticed. However, It does not show what this difference is. Images at the left on each pair show the contrast-normalized per-pixel difference

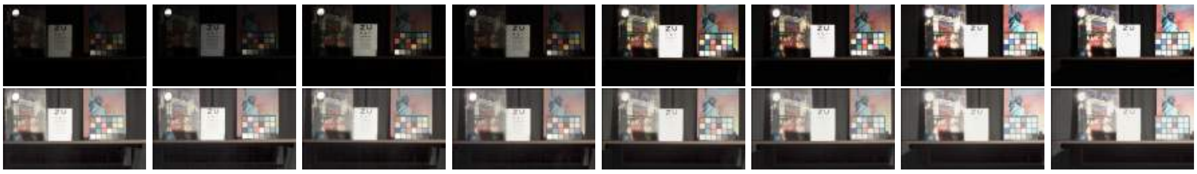
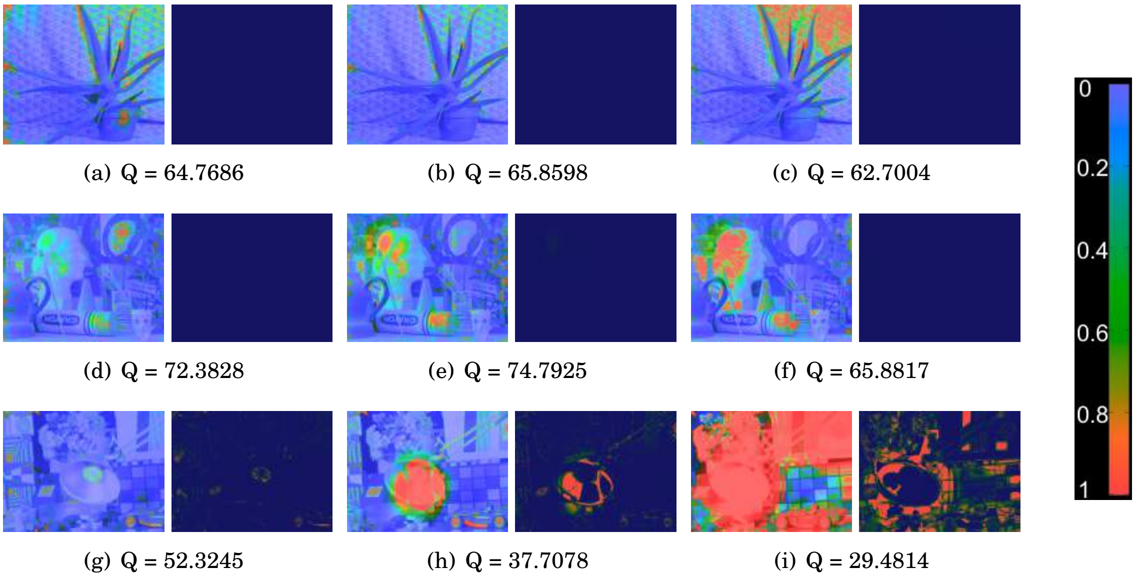


Figure 4.12: Up: Set of LDR multi-view images acquired using the Octo-cam [PcPD+10]. Down: the resulting tone mapped HDR taking each LDR as reference respectively. Despite the important exposure differences of the LDR sequence, coherent HDR results are obtained. However, highly saturated areas might remain saturated in the resulting HDR. Images courtesy of [BVNL14].

weighted by the probability of detection. The resulting images do not show probabilities. However, they better correspond to the perceived differences.



(a) Q = 64.7686 (b) Q = 65.8598 (c) Q = 62.7004

(d) Q = 72.3828 (e) Q = 74.7925 (f) Q = 65.8817

(g) Q = 52.3245 (h) Q = 37.7078 (i) Q = 29.4814

Figure 4.13: . HDR-VDP-2 Comparison between ground truth HDR images and our results. Each pair corresponds to the probability of detection (left) and the contrast-normalized per-pixel difference (right) of low, medium and high exposures corresponding to different views. The first row corresponds to the three first views of the 'Aloe' data set (Figure 4.11). The second row to the 'Art' data set (Figure 4.5) and the third row to the IIS Jumble dataset (Figure 4.10) HDR-VDP-2 Comparison. Images courtesy of Middlebury [Mid06] and [BRG+14] respectively.

The results illustrate that in general, no difference are perceived. Except in areas that appear totally saturated in the reference image, like the head of the sculpture in the 'Art' data set or the lamp in the IIS Jumble. In such cases visible artifact appears because the matching step fails to find valid correspondences.

Our method is faster than some previous solutions. [SKY+12] mention that their method takes less than 3 minutes for a sequence of 7 images of 1350x900 pixels. The combination of a reduced search space and the coherence term effectively implies a reduction of the processing time. On a Intel Core i7-2620M 2,70 GHz with 8 GB of memory, our method takes less than 2 minutes (103 ± 10 seconds) for the Aloe data set with a resolution of 1282x1110 pixels.

4.4 Summary

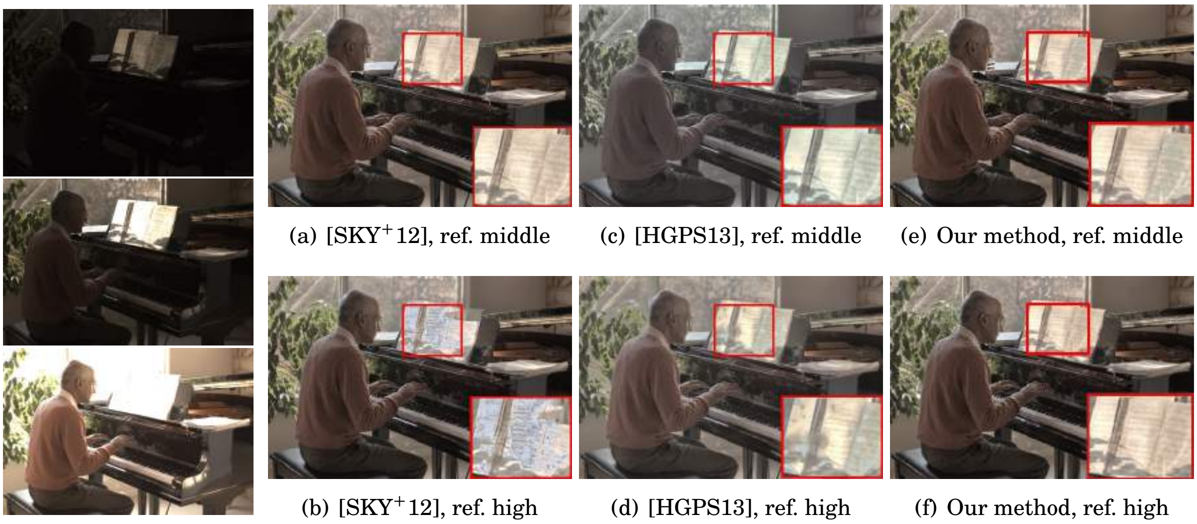
This chapter presented a framework for auto-stereoscopic 3D HDR content creation that combines sets of multiscopic LDR images into HDR content using image dense correspondences.

Our novel approach extends the well known Patch Match algorithm, introducing an improved random search function that takes advantage of the epipolar geometry. Also a coherence term is used for improving the matching process.

These modifications allow to extend the original approach to work for HDR stereo matching, while improving its computational performances. We have presented a series of experimental results showing the robustness of our approach, in the matching process, when compared with the original approach and its qualitative results.

CHAPTER 5

IN-HDR-PAINTING



(a) [SKY$^+$12], ref. middle
(c) [HGPS13], ref. middle
(e) Our method, ref. middle

(b) [SKY$^+$12], ref. high
(d) [HGPS13], ref. high
(f) Our method, ref. high

Figure 5.1: This figure presents results of different state-of-the-art methods using two reference images, the middle exposure (top row) and the highest exposure (bottom row). Images in (a) and (b) correspond to results of Sen *et al.* [SKY$^+$12], notice the artifacts of (b) in areas that are saturated in the reference image. Images in (c) and (d) are results of using [Hu et al. 2013], HDR information is not recovered areas saturated in the reference. (e) and (f) show results of the in-HDR-painting method presented in this chapter. Plausible HDR images are produced independently of the selected reference, including highly saturated areas. Source images courtesy of Sen *et al.* [SKY$^+$12]

5.1 Introduction

Finding per pixel dense correspondences in differently exposed sequences is slow and is prone to produce artifacts due to mismatches. This chapter presents a technique called In-HDR-Painting that reconstructs HDR images by replacing only the under/over exposed areas in a reference image with correctly exposed pixels from the appropriate source images in the LDR sequence. After detecting under/over exposed areas, this method finds correspondence matches for correctly exposed pixels in the contour of such areas. The correspondence for pixels in under/over exposed regions is determined by interpolating the correspondences of pixels in the contour. Bright pixels in over exposed areas are replaced with properly exposed pixels from a source image with lower exposure than the reference, and the other way around for under exposed areas. Unlike previous approaches, this algorithm can handle large under/over exposed regions.

The key contributions of this chapter can be summarized as follows:

- The problem of HDR reconstruction is tackled for the first time, using the concept of inpainting, where an adaptive interpolation process is used to fill-in the under/over exposed areas.
- This is a reference-independent solution that automatically recovers HDR values for under or over exposed areas in dynamic scenes independently of the selected reference image.
- Our algorithm can generate HDR images from a sequence of multiple exposure LDR images with both camera and scene motion without producing ghosting effects.

5.2 Image inpainting previous work

The concept of digital inpainting has been largely studied [IP97, BSCB00, CS01, Har01, OBMsC01, CPT04, KSD$^+$14]. Inpainting refers to algorithms designed to reconstruct images by filling regions where the data is lost or corrupted. It is based on the similar concept of artist inpainting used to modify an existing image in an undetectable way. Inpainting is useful in various applications such as restoring damaged areas of the input image, as well as removing unwanted objects. Despite the fact that researchers have adopted different names to describe the same problem [KCS02], we can identify it as a typical interpolation problem where the information is inferred into the region of interest based on the information available in the neighboring regions. A detailed survey on inpainting approaches was presented by Bertalmío *et al.* [BSCB00].

Partial Differential Equations (PDE's) are used by Chan and Shen [CS01] because it automatizes the interpolation process, has the advantages to be free from object segmentation or edge detection, and do not impose any topological constraints [KCS02]. However, this technique fails often when the region to be inpainted is large. This is due to the fact that the information available in the input image is not enough to recover the original content. To solve this problem, the missing information can be extracted from other images of the same scene taken from different points of view [KCS02].

Harrison [Har01] introduced a procedure for synthesizing an image with the same texture as a given input image by successively adding pixels selected from the input image. Pixels are chosen by searching the input image for patches that closely match pixels already present in the output image. Oliveira *et al.*[OBMsC01] presented a simple inpainting algorithm that was two to three orders of magnitude faster than the state-of-the-art method, but only valid to reconstruct small missing and damaged portions of images.

Criminisi *et al.*[CPT04] presented an algorithm for removing large objects from digital images. Their method combines texture synthesis and inpainting in a best-first algorithm in which the confidence in the synthesized pixel values is propagated. The algorithm consists of three steps that are repeated until the whole target region is filled. First, a priority value is calculated for pixels on the contour of the target region. The area around the pixel with the highest priority is the first to be filled. In the second step, the target area is filled with patches that looks alike to the one centered in the pixel with higher priority. The third step recalculates all priorities of pixel on the contour.

Kalantari *et al.*[KSD$^+$14] used patch-based synthesis to replace given areas in the image. Instead of fixed patch sizes, they use content-adaptive masks. The method propose two alternatives, user manual annotation of the boundary edges inside the hole or training of a learning model to preserve details inside the areas to be filled.

5.3 In-HDR-painting Algorithm

The input of in-HDR-painting is a sequence of LDR images taken at different exposure times, it includes dynamic scenes acquired from a moving camera. The proposed algorithm recovers an HDR image corresponding to a reference image of the LDR sequence and replaces the under/over-exposed pixels with valid information from the other LDR images of the sequence. This solution is independent on the choice of the reference image and potentially recovers the HDR content even in large over/under-exposed areas.

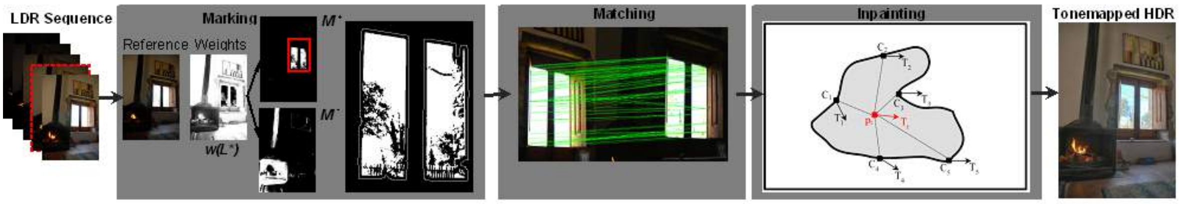


Figure 5.2: Diagram explaining the in-HDR-painting framework step-by-step. Marking step is in charge of identifying the under/over-exposed (target) areas in the reference LDR image. Contours of the target areas contains properly exposed pixel for the search of match correspondences between the contours and a source image in the LDR sequence (Matching step). Finally the pixels information found in the matching step is used to replace under/over exposed pixels in the reference through an inpainting interpolation process.

The framework is depicted in Figure 5.2 and has three main steps:

1. The LDR sequence is converted to the $CIE\ L^*a^*b^*$ color space. The L channel is used to mark the under-over-exposed areas (target areas) (section 5.3.1).

2. The images are transformed to radiance space before finding matches between pixels in the contour of target areas and a source image in the LDR sequence (section 5.3.2).

3. The information to replace the target areas with valid information is inferred through an inpainting approach (section 5.3.3).

5.3.1 Marking Step

An HDR image corresponds to a radiance map that approximately represents the light values registered by the camera sensor. Recovering radiance values from only one exposure is possible except for under/over exposed pixels. In such cases, information from other exposures is required. Combining a sequence of different exposures with a weighted average is the predominant approach. This method proposes to recover radiance values for the pixels that are properly exposed and replace the under/over exposed with valid information from other exposures. The first step is to identify and mark these *target* regions containing under/over exposed pixels.

The previous work to merge multiple exposures included weighting functions to determine the contribution of pixels to the HDR image and exclude the under/over saturated ones from the HDR merging. Figure 5.3 shows some weight functions and the result of weighting two differently exposed images. The right-most column shows the graphic representation of the weight function (in black) and the inverse CRF (in red). The two first columns show weight in gray scale images. The weight is a value in the range [0,1] where 0 (black) means that a pixel is totally under or over exposed and hence is not good to recover HDR values and 1 (white) shows pixels that are well exposed.

Figure 5.3(c) shows the triangular hat function suggested by Debevec and Malik [DM97]. Using this function only pixels with the medium value get the maximum weight. Mitsunaga and Nayar [MN99] (Figure 5.3(d)) proposed a function that favors pixes where the signal-to-noise ratio (SNR) and response to radiance are high. In this case pixels saturated get the highest weight. Reinhard *et al.*[RWD$^+$10] proposed to multiply this function with a broad hat function like shown in Figure 5.3(e) to avoid maximum weights to saturated pixels.

We use the hat function presented by Khan *et al.*[KAR06] defined with the equation 5.1 (Figure 5.3(e)). We weight the luminance L^* channel normalized to the range $[0,1]$ of the image in $CIE\ L^*a^*b^*$ color space. The two masks identifying the over-exposed target areasM^+(equation 5.2), and the under-exposed target areas M^-(equation 5.3) respectively are obtained using a threshold value over the *w(L*)*.

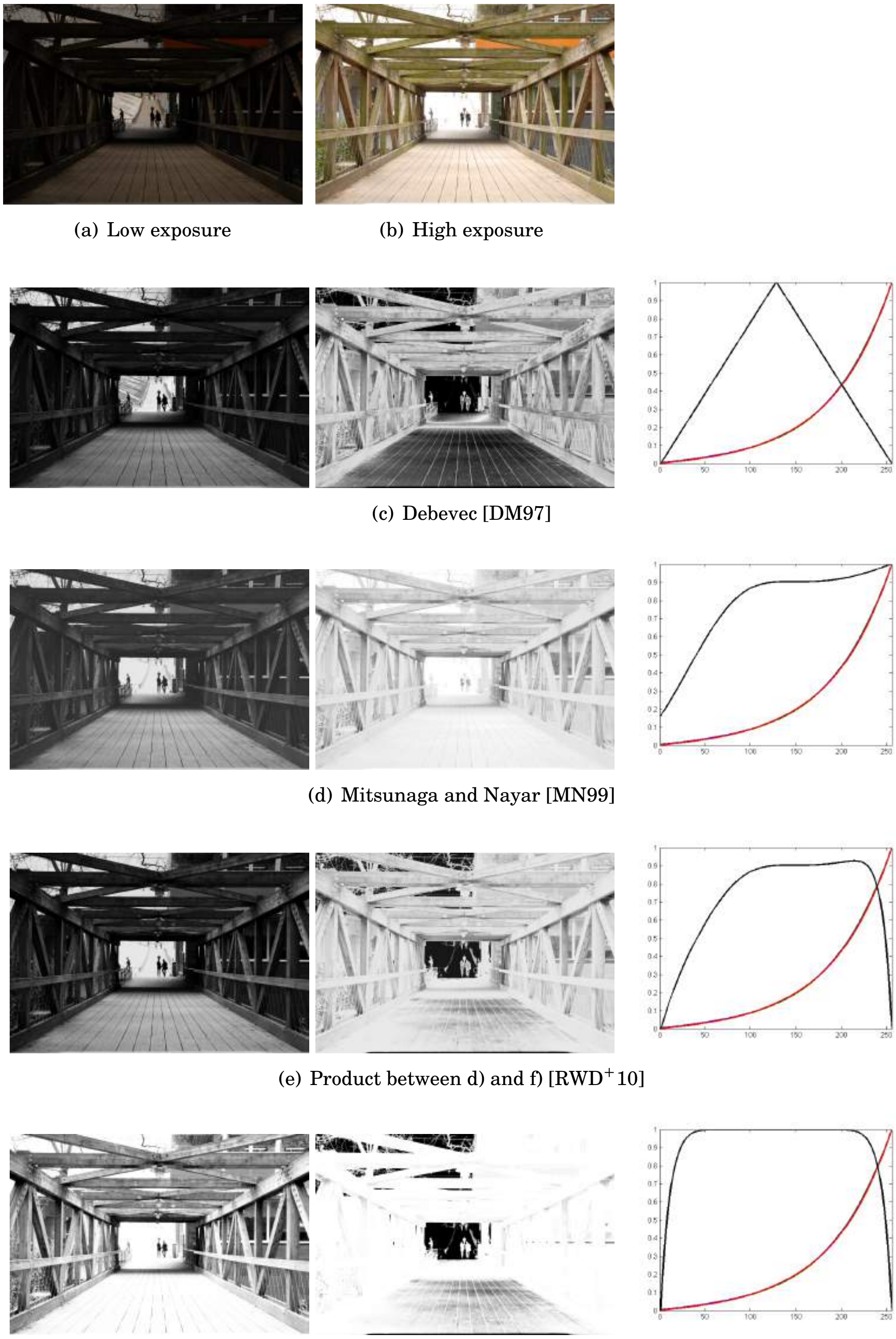
$$w(L^*) = 1 - (2 \cdot L^* - 1)^{12} \tag{5.1}$$

$$M^+ = w(L^*) < T \cap L^* < 0.5 \tag{5.2}$$

$$M^- = w(L^*) < T \cap L^* > 0.5 \tag{5.3}$$

Figure 5.4 shows an example of the results of the target identification step. The weight map *w(L*)* shown in Figure 5.4(b) represents the identified under and over-exposed areas. In this case, the fireplace and the window respectively, that are classified identified in two different masks M^- Figure 5.4(c) and M^+ Figure 5.4(e) using equations 5.2 and 5.3.

However, target areas do not represent defined clusters of pixels. Camera sensors often generate noise mainly in the underexposed areas, making the marking step not fully reliable in detecting these target areas. To prevent this issue we impose a constraint for the target areas to be solid regions representing cluster of pixels. We apply a morphological closing and opening operations to remove small holes and isolated pixels consequence of noise. It consists in combining erosion and dilation of the target areas with a structural element, i.e., a 5 × 5 squared kernel. Results are shown in Figures 5.4(d) and 5.4(f).



(a) Low exposure

(b) High exposure

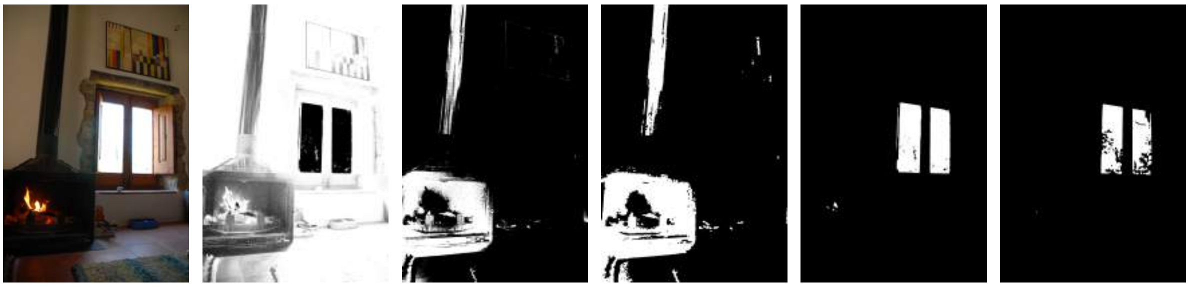
(c) Debevec [DM97]

(d) Mitsunaga and Nayar [MN99]

(e) Product between d) and f) [RWD$^+$10]

(f) Khan *et al.* [KAR06]

Figure 5.3: Results of using different weighting functions on two different exposures a) and b). The column at the right shows the weighting function (black curve) and the inverse CRF (red curve) for a Nikon D200 camera used to acquired the images.



(a) Reference (b) Weight w(L*) (c) Mask M^- (d) M^- filtered (e) Mask M^+ (f) M^+ filtered

Figure 5.4: Reference image (a) its associated weights (b) and the corresponding masks M^- and M^+ before (c), (e) and after the morphological operations (d), (f)

5.3.2 Matching Step

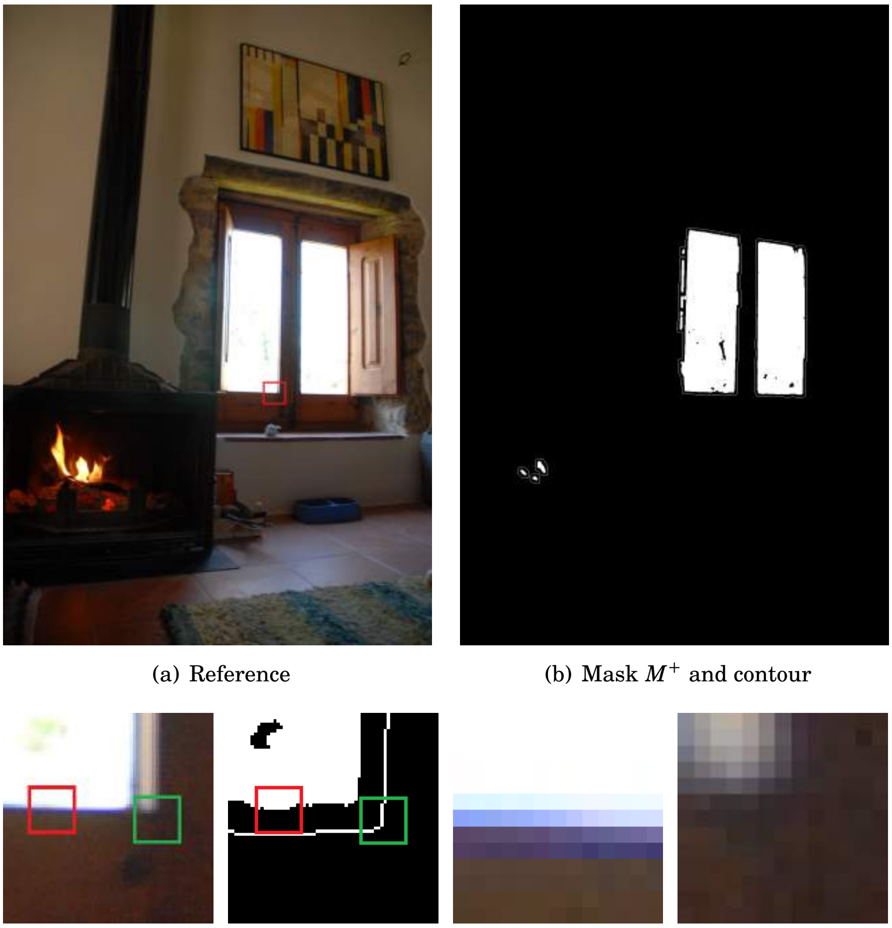
Once the under/over-exposed areas are identified, the next step is to find the content to replace under/over exposed pixels in the target areas with valid information, while preserving intact edges and details. Finding matches for pixels inside the target areas is nearly impossible, those areas are close to either black or white values and corresponds to properly exposed areas in other exposures. This is why previous works fail in large under and over exposed areas. Our approach instead, aims to interpolate correspondences for them by matching the contours of the target areas and interpolating the results for pixels inside.

Patches centered in the border of the target area are partially under/over exposed (Figure 5.5(e)) which might difficult the matching process. We find an external contour (Figure 5.5(d)) to guarantee that all pixels in the patch are properly exposed (Figure 5.5(f)). The external contour is calculated with a morphological dilation over the masks, using a squared kernel of size equal to half of the patch size +1. Detecting the contour line is easy on binary images, the contour is isolated using the gradient of the dilated mask. This step is essential because having under/over exposed pixels in a contour patch would reduce the amount of valid information for the Nearest Neighbor Search (NNS) and the possibilities of finding reliable matches.

For each squared patch of pixels centered on this contour, we use a patch-based NNS to find matches on the appropriated LDR source [BSGF10]. NNS technique finds matches by minimizing a cost function (the Sum of Squared Distance in this case) between patches of two images. It has been found a reliable strategy for matching in previous methods to match different exposures [SKY$^+$12, HGPS13, OMLA15]. In order to be more accurate we impose the constraint that a contour patch does not cover any under/over exposed pixel.

Problems with the geometrical coherence of the NNS search were discussed already [OMLA15]. The matches are found based on the color information of patches, but no geometrical constraints are considered in the search. In low textured areas where homogeneous patches looks all the same, mismatches may appear and adjacent pixels in static regions of the reference image may find different transformations (see Figure 5.6(b)). However, the probability that several matches find the same transformation is much lower than that of any individual match to be wrong [HSGL11]. We define the consistency of a match as the variance of the transformations obtained for each pixel inside the patch (Figure 5.6(c)).

We use only the best matches of the contour for calculating the correspondences of pixels inside the target areas. Two criteria define which matches are more reliable:



(a) Reference

(b) Mask M^+ and contour

(c) Zoom in a)

(d) Zoom in b)

(e) Red patch

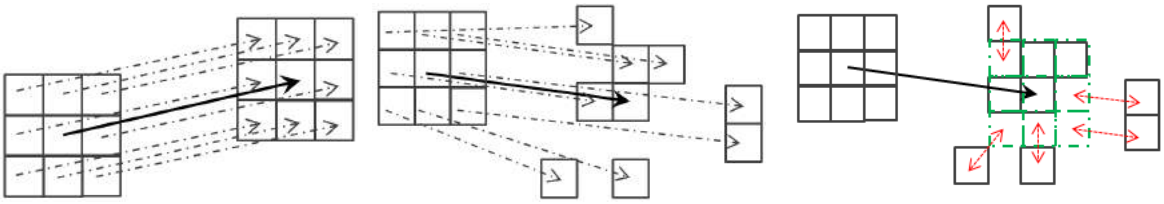
(f) Green patch

Figure 5.5: Reference image (a) and the corresponding M^+ mask with the dilated contour in white (b). (d) and (e) are zooms over (a) and (b) respectively. Patches centered in the contour of target areas (red square), like the one shown in (e), contain many pixels with saturated values. Dilating the contour increases the number of valid pixels inside the patch, like the example shown in (f). Having well defined patches augment the possibilities of finding reliable matches.

1. the SSD error between a patch and its nearest neighbor obtained during the NNS step.
2. the consistency of the transformations corresponding to every pixel inside the patch.

The reliability R (equation 5.6) of a patch P_r is defined as the product between the $SSD(P_r,P_s)$ being P_r a patch in the reference image and its nearest neighbor in the source image P_s and the consistency (variance of the distances between the predicted coherent matches and the actual ones). The euclidean distance d between a predicted match m_p for a pixel p_i with transformation T_i inside a patch which center has a transformation T_c and the actual match m_a, is calculated using equation 5.4. The variance (σ^2) of such distances (Figure 5.6(c)) is calculated using equation 5.5.

$$d(m_p,m_a) = \sqrt{(p_i+T_i)-(p_i+T_c)^2}) \tag{5.4}$$



(a) Consistent match (b) Inconsistent match (c) Consistency of matches

Figure 5.6: NNS match overlapping patches based on color information. If patches are part of an static object pixels, all pixels inside it should have similar transformations (a). But it is possible that pixels in patches from a static regions find different transformations which are not geometrically coherent (b). Figure in (c) shows in green the predicted position if all pixels have the same transformation as the center of the patch. The consistency of matches inside a patch is measured as the variance over the distances between the predicted matches for coherent matches and the actual ones (red arrows) is used to measure a patch coherency. In coherent matches this value must be equal to zero.

$$\sigma_d^2 = \frac{1}{N}\sum_{n=1}^{N}(d_i - \bar{d})^2 \tag{5.5}$$

$$R(P_r) = SSD(P_r, P_s) * (1 + \sigma_d^2) \tag{5.6}$$

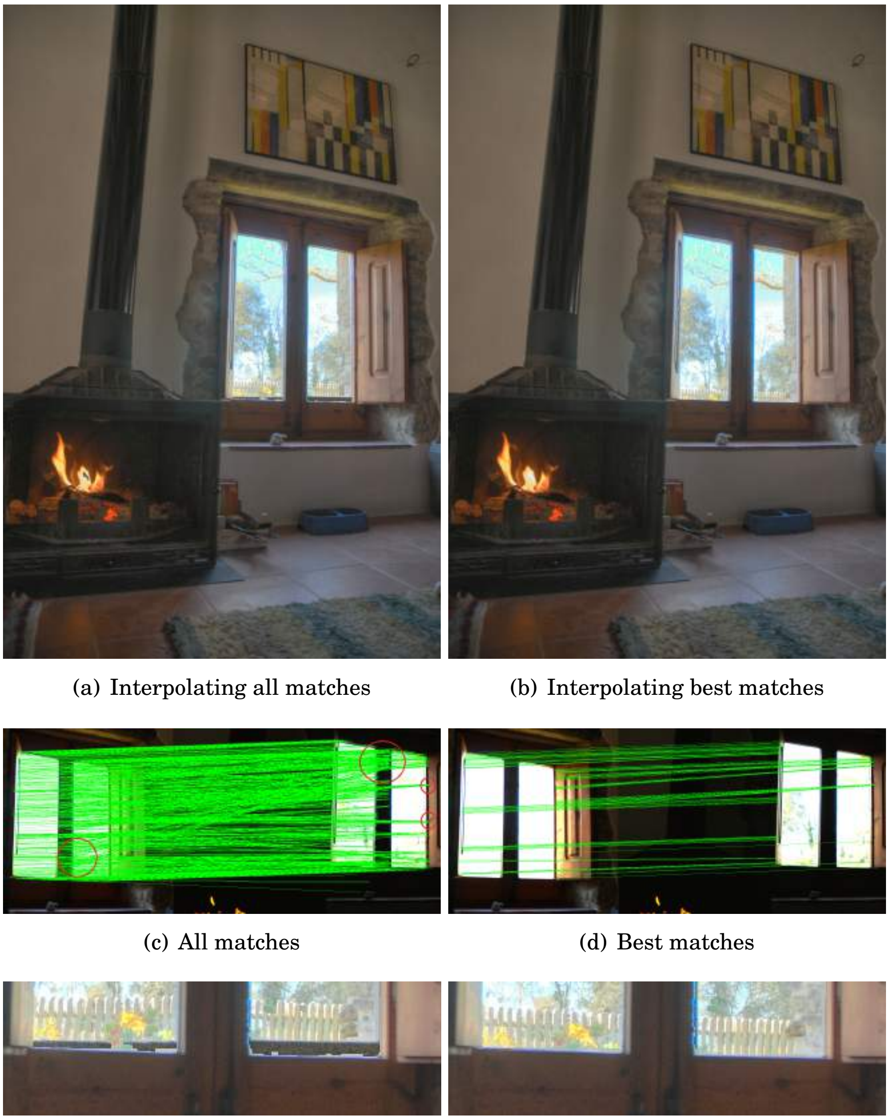
5.3.3 Inpanting from Multiple LDR Sources

Most inpainting methods sample the surroundings of the target area and use it as source of information to complete the missing pixels [BSCB00]. The information is propagated generally by growing from the contours to the interior using constraints to preserve edges and details. Previous methods fail though in the case of large under/over-exposed areas because there is not enough reliable information in the reference image to complete the target area.

In multiple exposure sequences, the information to replace under/over exposed areas could be available in the other LDR images. This method uses pixels properly exposed from low exposures to replace saturated regions in the reference image and the opposite for under exposed areas. The challenge is to 'paste' this information in the target areas, while respecting details like edges in the reference image.

Matching pixels is subject to local transformations that may approximate complex transformations in the scene. We propose to calculate the values inside the target areas as the linear interpolation of the matches in the contour. We use linear interpolation and weight the contribution of each match in the contour based on their reliability value as computed in equation 5.6 and the distance to the contour. Only the most reliable patches act like control points C_i in our interpolation scheme. We select the best points (the best 10% worked properly in our experiments) and normalize their reliability value (normalized reliabilities are denoted by $\hat{R}_i$), these points act like control points for the interpolation process.

$$c_i = (1 - \hat{R}_i) / \sqrt{p_t^2 - C_i^2} \tag{5.7}$$



(a) Interpolating all matches

(b) Interpolating best matches

(c) All matches

(d) Best matches

(e) Zoom in a)

(f) Zoom in b)

Figure 5.7: Nearest Neighbor Search minimizes the SSD between patches, which consider only color information and any geometric information. More than one patch may match the same destination in the source image and mismatches may appear in low textured areas. The inpainting interpolation using wrong matches produces artifacts in the results (a). We select the best matches according to our consistency criteria to produce accurate inpainting results (b).

$$\hat{c_i} = \frac{c_i}{\sum c_i} \tag{5.8}$$

$$T_t = \hat{c_1}T_1 + \hat{c_2}T_2 + ... + \hat{c_i}T_i \tag{5.9}$$

For every pixel p_t in the target area we find a transformation T_t (equation 5.9) that matches a pixel in the source image by linearly interpolating the transformations T_i of the control points C_n. The coefficients that determine the contribution of the transformations of each C_n are calculated taking into account the normalized reliability value and the distance from p_t to each control point. This helps to reinforce the contribution of good matches reducing the influence of possible outliers (see Figure 5.8).

5.4 Results and Discussion

The algorithm was implemented in C++ using OpenCV [Ope15], and it was tested on a large set of images including images from state-of-the-art papers. These images were acquired under different lighting conditions such as indoor and outdoor environments, and all of them corresponds to dynamic scenes i.e. either due to camera movement or movement of objects.

We have tested our algorithm using both 8-bits and RAW formats, without using any pre-alignment technique. In all cases, we have selected the reference with large under/over exposed regions to prove that our method is independent of the reference. The HDR results shown in this section are tonemapped using Photomatix version 4.2.1 [htt15].

Results were generated using the following parameters. For the marking step, the threshold T was chosen as 0.9. As discussed in section 5.3.1 this value, as verified in the experimental algorithm design, produces reliable results for all the cases shown in this section. The patch size used for the matching step has an important role on the results. Small patches may be very difficult to match, while bigger patches are time consuming. We have used an automatic trade-off, in all these experiments, where the patch size is assigned proportionally to the image size. In these experiments, the results are generated using a patch size equal to 1.5% of the image's smallest dimension, but never smaller than 7×7 pixels. Once the contour matches are calculated, we select only the best 10% of matches for the inpainting step.

Figure 5.9 shows an example on how the reconstruction result fails when the size of the patch is too small (3×3 pixels in this example). This is due to two main reasons: (1) small patches have limited information which increases the probability of finding false positive matches and (2) noise and/or limited accuracy in the CRF estimation may lead to wrong matches.

To evaluate the quality of our approach, we compared our results with the following state-of-the-art methods Sen *et al* [SKY+12], Hu *et al* [HGPS13], Granados *et al* [GKTT13] and Zimmer *et al*. [ZBW11].

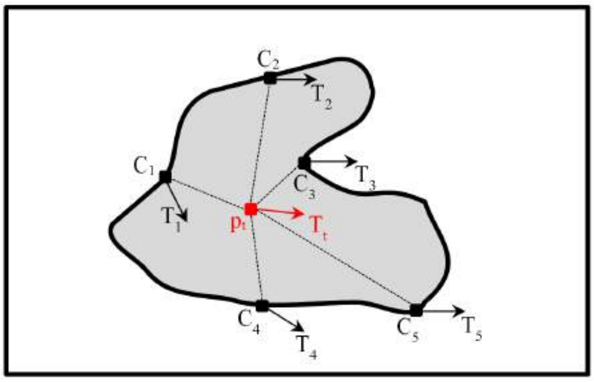
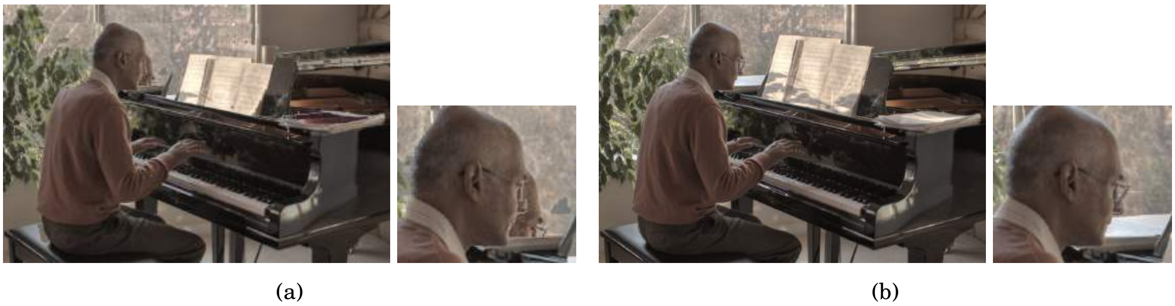


Figure 5.8: Diagram of the interpolation scheme. A selection of the best matches of pixels in the contour of the target areas (C_n) is used to interpolate the transformation T_t for every pixel p_t in the target area by interpolating the transformations T_n of the best matches in the contour.


(a)

(b)

Figure 5.9: Example of the reconstructed HDR image using different sizes of patches: (a) 3 × 3 pixels and (b) 9 × 9 pixels.

5.4.1 Quality Evaluation

We started evaluating our method showing a result where the reference image was chosen as the highest exposed image (Figure5.10 right side). In this case, the reference image presents large over-exposed areas (bumper), and the LDR sequence presents moderate movements of the roster. The reconstructed HDR image, shows well reconstructed HDR values even in the areas with limited information (over-exposed) and with dynamic movements.

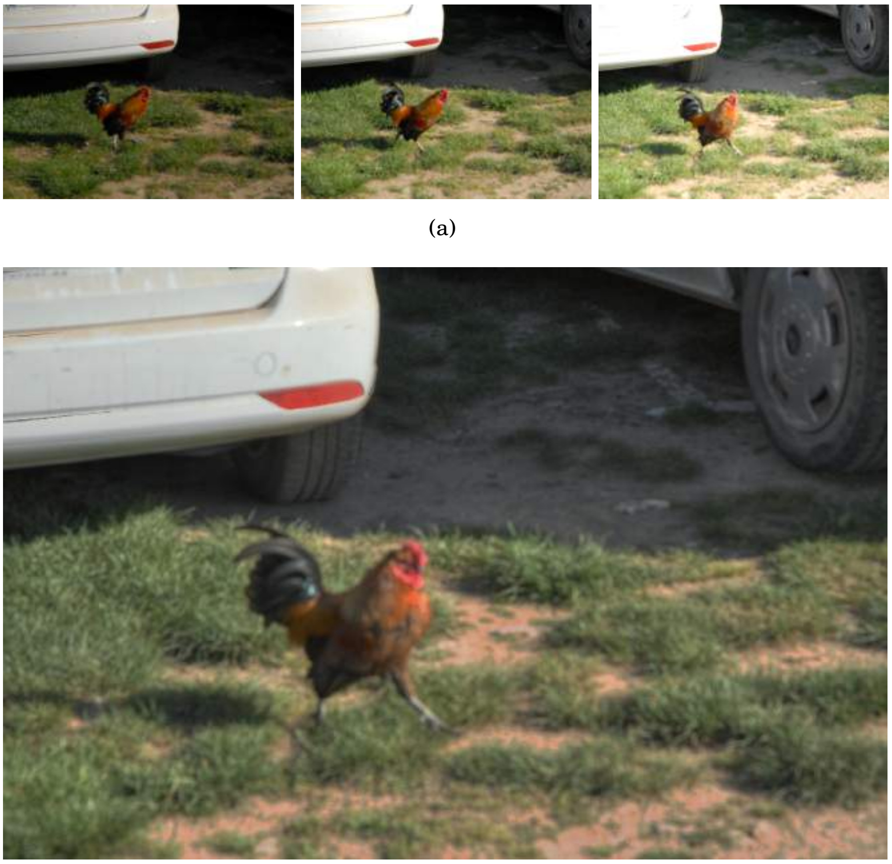
Figures 5.11 and 5.13 show the reconstructed HDR image from a LDR sequence of 5 images taken at different exposure times. Here the HDR reconstructed image is compared with two of the state-of-the-art methods, Sen *et al* [SKY$^+$12] and Hu *et al* [HGPS13], used in this evaluation.

In Figure 5.11, the over-exposed image, in the LDR sequence, is taken as reference. In the lower row of Figure 5.11 is shown a zoomed area where artifacts and/or lost of details are visible for both state-of-the-art methods. In particular Hu *et al*'s [HGPS13] clearly looses details in the window area, while Sen *et al*'s [SKY$^+$12] introduces visible artifacts.

The second row of Figure 5.13 shows for each method the reconstructed HDR images taking as reference the two extreme exposed images in the LDR sequence. In two zoomed areas, artifacts and/or lost of details are visible for both state-of-the-art methods. In particular Hu *et al*'s [HGPS13] clearly looses details in the sky for low exposures, while introducing artifacts for higher exposures. Sen *et al*'s [SKY$^+$12] looses details in the sky area for higher exposures, and details in the lower exposure in the area of the houses as highlighted in the zoom area. In comparison to the state-of-the-art methods, our algorithm produces reliable HDR reconstructed images for all the cases (lower and highest exposures). Details are well preserved in both zoomed areas and the results are free of artifacts.

Figures 5.12 and 5.14 show the results obtained with Zimmer *et al* [ZBW11], using as reference the middle exposure image of the LDR sequence. Band artifacts are visible on both HDR reconstructed images (Figure 5.12 in the window area and Figure 5.14 on the left hand side).

Figure 5.15 shows a stress case. A child moves his head in front of a window (large over-exposed area). We compare here our results with Granados *et al*'s approach [GKTT13]. Again our approach performs better overall. The reconstruction with Granados *et al*'s method has actually deformed the face of the child and the window content misses details. Our reconstruction recovered the HDR information both for the window and the child's face. However, inpainting edges are visible around the child's head outlining a limitation of our approach. This occurs because of the combination of movement and a large over-exposed area.



(a)

(b)

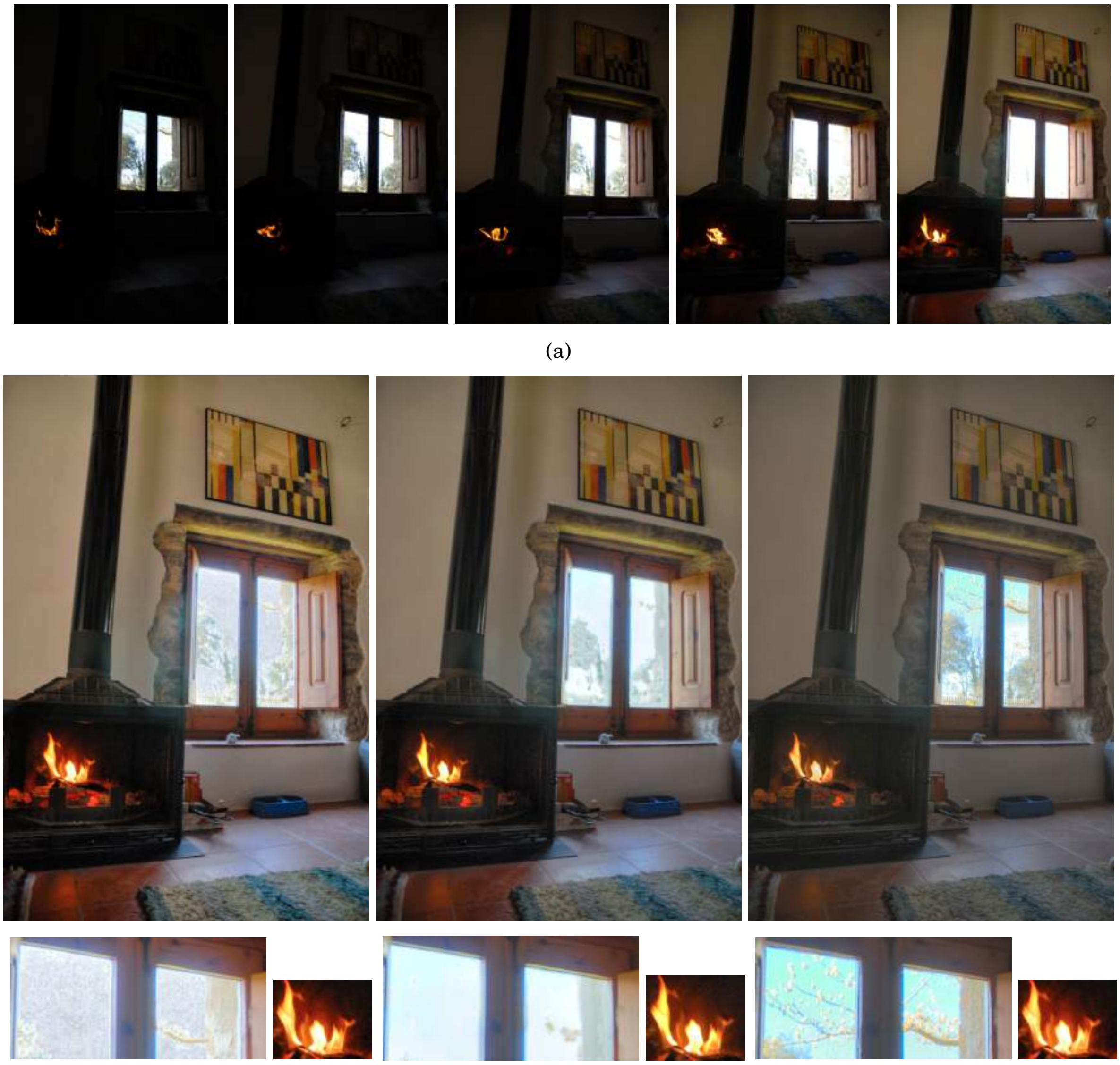
Figure 5.10: LDR sequence with three images taken at shutter speeds of (left) 1/5000, (center) 1/1000 and (right) at 1/180 and with aperture of $f/5.6$. The scene was captured with a moving camera and while the rooster was running. The HDR reconstructed image in b) was obtained using the over-exposed LDR image at the right in a). Notice that all details in the car are successfully recovered

5.4.2 Computational Performances

Unlike state-of-the-art methods, the computational performances of our method depends on the size of the target areas. For bigger target areas more pixels in the contour need to find a nearest neighbor. Larger contours imply that more control points are selected and more coefficients are needed to interpolate the position of a bigger number of pixels to replace the target areas. We have taken the computational costs for reconstructing the HDR image using the high and middle exposures taken as references.

Table 5.1 shows a comparison of the computational times for the largest exposure in each set. In this case only over exposed areas are replaced since there is no better exposure for dark areas in the sequence than the reference itself.

The resolutions of the images used in Table 5.1 are 1296 × 1936 for the "Rooster by Car" and the "Windows" images, and 1350 × 900 for the "Piano Man" image. All tests were executed on a Dell laptop with an Intel(R) Core(TM) i7-2620M 2.70 GHz processor and 8 GB of memory. All algorithms were run



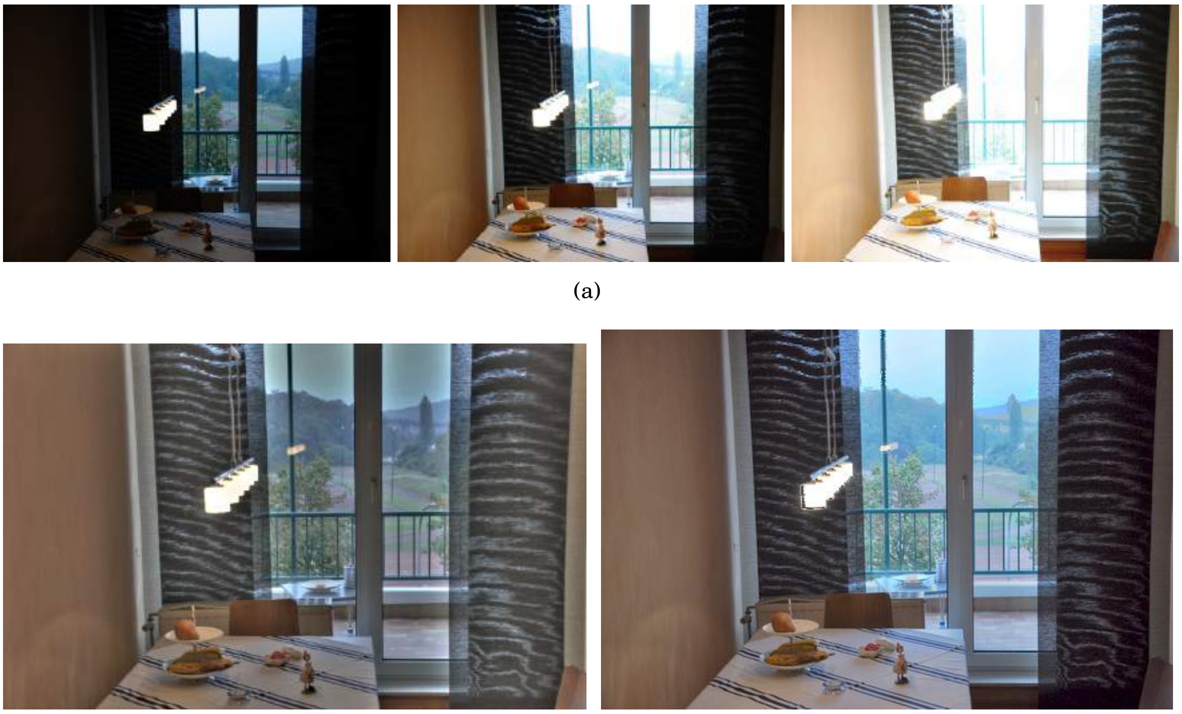
(a)

(b) Sen *et al* [SKY+12]

(c) Hu *et al* [HGPS13]

(d) Ours

Figure 5.11: Comparison with state-of-the-art methods. The used LDR sequence a), is composed of 5 LDR images taken respectively, starting from the left to the right, at shutter speeds of 1/640s, 1/320s, 1/160s, 1/80s, and 1/40s, with an aperture size of f/3.5.



(a)

(b) Zimmer *et al*[ZBW11]

(c) Ours

Figure 5.12: The reconstructed HDR image is obtained using the middle exposure image. From the left to the right, the acquisition shutter speeds correspond respectively to 1/60, 1/13, 1/3 and with aperture size of $f/5.6$.

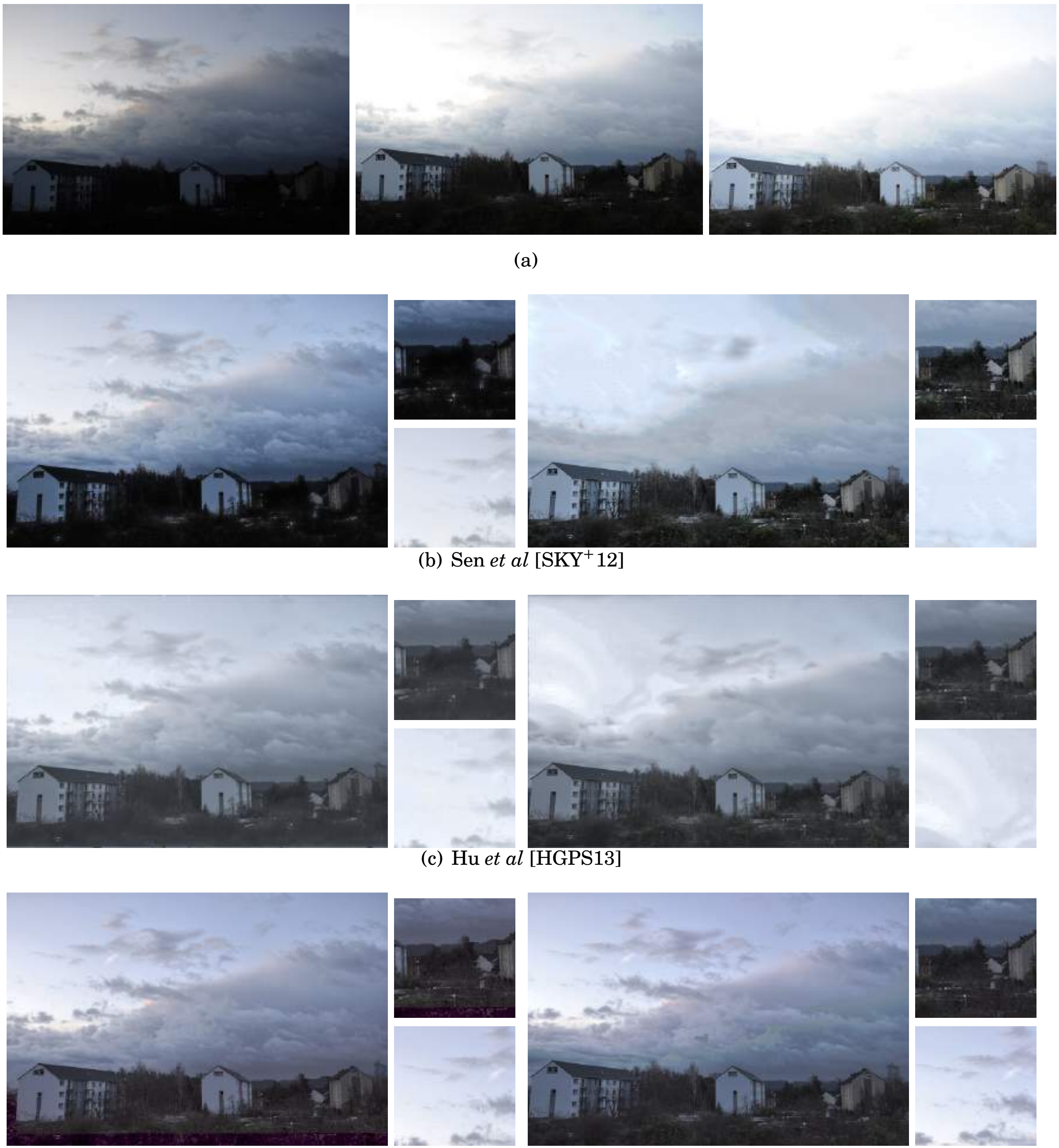
	Sen		Hu		Ours	
	Mid	High	Mid	High	Mid	High
Rooster by Car	618	588	819	793	243	76
Windows	503	577	979	936	254	99
Piano Man	439	576	578	430	347	122

Table 5.1: Run time comparison, in seconds, with state-of-the-art methods taking as reference both the middle exposure and the highest exposure of the set.

using their original code version. Our method outperforms significantly Sen *et al* [SKY$^+$12], which is faster than Hu *et al* [HGPS13]. In particular, our method outperform state-of-the-art methods in about an order of magnitude.

5.4.3 Discussion

The tests show that our method's success relies on the results of the matching step. Mismatches can be produced that may reduce the performances of our selection process. Low textured areas in the contours of the target areas is one of them, when all patches in the contour are similar is unlikely to find proper matches. When replacing dark areas acquired with very short exposures, the contours may be extremely noisy which also represents a challenge for the matching step. One solution to this issue, is to pre-filtering



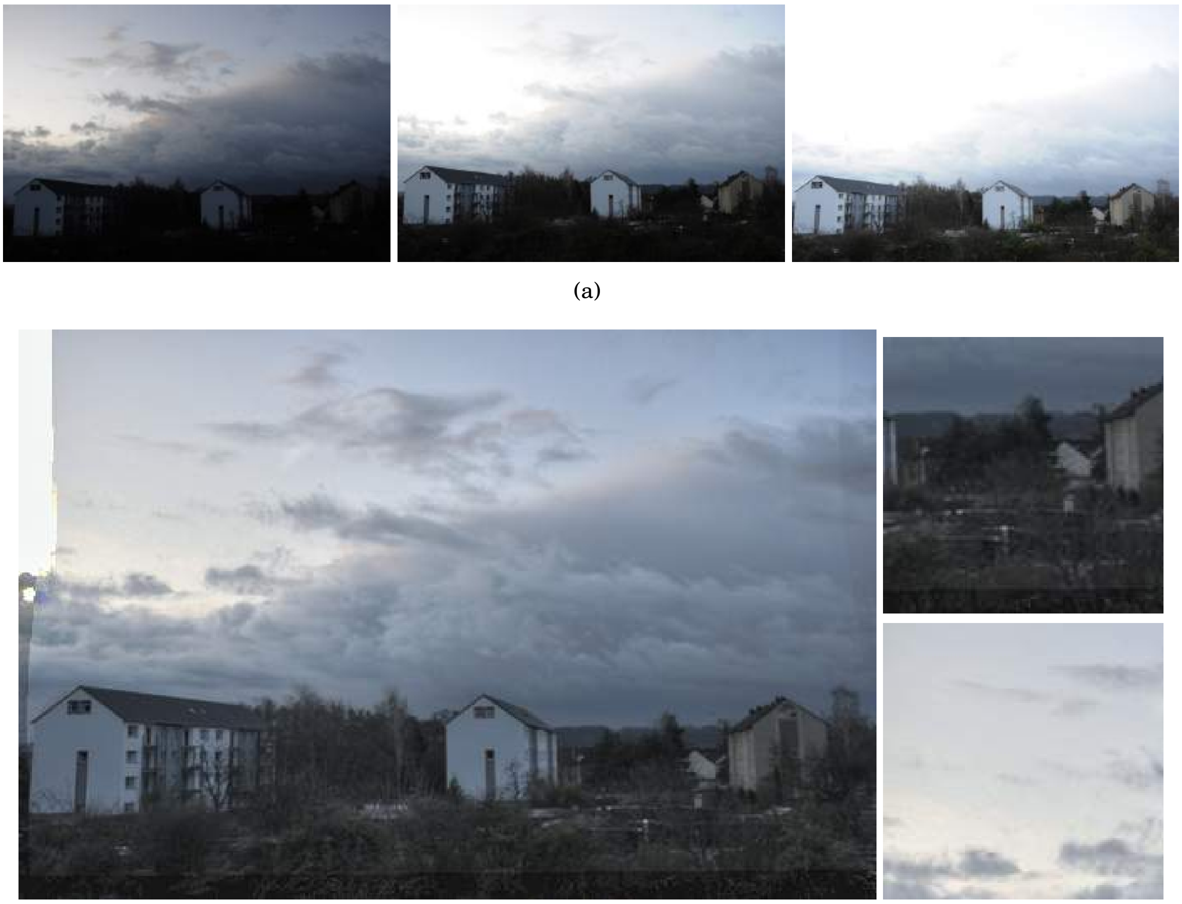
(a)

(b) Sen *et al* [SKY$^+$12]

(c) Hu *et al* [HGPS13]

(d) Ours

Figure 5.13: Comparison with state-of-the-art methods. The used LDR sequence (first row), is composed of 3 LDR images taken respectively, starting from the left to the right, at shutter speeds of 1/250, 1/80 and 1/30, and with an aperture size of $f/8$. The HDR reconstructed images are obatined using as reference the two images with the two extreme exposures (lower and highest).



(a)

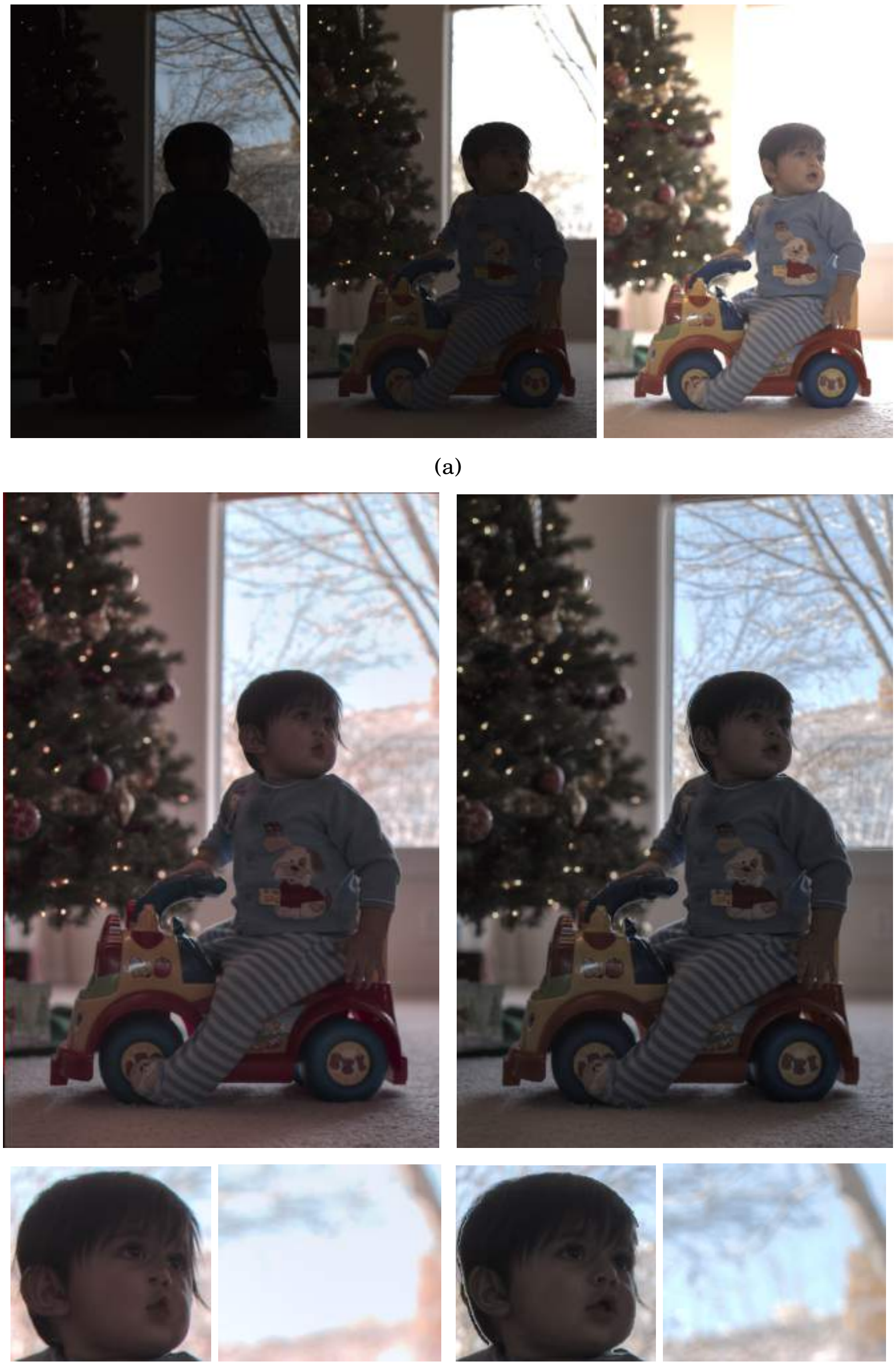
(b) Zimmer *et al.*[ZBW11]

Figure 5.14: Continuation of the figure 5.13. Results from Zimmer *et al.*[ZBW11] are only available for the middle exposure.

the noise of the lower exposed images in the LDR sequence.

5.5 Summary

In this chapter we presented a new fully automatic approach to reconstruct HDR images from a sequence of LDR images taken at different exposure times. There is no restriction on the LDR sequence: images can be misaligned, scene content can be dynamic, and areas under or over exposed can be large. This is an innovative approach when compared to previous approaches because HDR images can be computed using any LDR image as reference. The algorithm has three step. First, target areas where HDR information is missing and its contours are detected. Second, we find matches for the contours in images better exposed. Third, the target areas are filled by interpolating the transformation of control points in the contours. We compared our approach to the state-of-the-art approaches evaluated as the best performing ones. We showed that our solution outperforms them both in the results' quality and the computation times.



(a)

(b) Granados *et al.* [GKTT13]

(c) Ours

Figure 5.15: The reconstructed HDR image is obtained using as reference the best exposure image. In this case is the image in the middle of the LDR sequence. For Granados et al. method this is done automatically. From the left to the right, the acquisition exposure values correspond respectively to $-4, 0, 4$.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

In the previous chapters, a detailed overview of the state-of-the-art techniques was presented regarding multiple exposure HDR acquisition. Three different techniques for HDR alignment and deghosting were proposed: one for images of dynamic scenes acquired from static cameras, the second focused on multiple-exposure mutli-stereo HDR acquisition for auto-stereoscopic displays, and a third one substitutes the global registration approach by an image inpainting based technique for registration and merging of multiple exposure sequence of images for HDR reconstruction.

6.1.1 Dynamic scenes from static cameras

Our method includes detection of areas affected by movement, matching of such areas in a reference image, and recovery of HDR values using the computed matches and the LDR image sequence. Robust results were obtained for scenes where dynamic objects were mainly rigid. We implemented some of the most used algorithms according to the descriptions given by their authors. The selected algorithms were developed in Matlab and a GUI was implemented for supporting the tests. The algorithms were tested using several sets of images from various type of scenes.

Regarding ghost detection, we implemented four approaches for which we assessed their degrees of success. These methods provided acceptable results in general. However, we have noticed that the introduction of a thresholding step on the difference between the predicted pixels with the IMF and the actual values helps to improves the final results. This evaluation finding was very important since the entire process relies on its results.

We applied registration techniques to the HDR reconstruction problem. When the clusters of pixels affected by movement in the ghost mask are closed and well-defined areas, the selection of the target images works properly. The registration step produces the best results when using MI or NCC as cost function, being the latter faster to compute. After the registration step, HDR values are recovered for areas affected by movement and inserted in the HDR image providing better representation for dynamic areas.

6.1.2 Mutliscopic HDR

This thesis presented a framework for auto-stereoscopic 3D HDR content creation that combines sets of multiscopic LDR images into HDR content using image dense correspondences. Image dense correspondences methods used for 2D domain have not be used before for 3D HDR content creation without introducing visible artifacts.

Our novel approach extends the well known Patch Match algorithm, introducing an improved random search function that takes advantage of the epipolar geometry. Also a coherence term is used for improving the matching process. These modifications allow to extend the original approach to work for HDR stereo matching, while improving its computational performances.

We have presented a series of experimental results showing the robustness of our approach, in the matching process, when compared with the original approach and its qualitative results.

6.1.3 In-HDR-painting

The third contribution method of this thesis presented a new fully automatic approach to reconstruct HDR images from a sequence of LDR images taken at different exposure times. Previously imposed restriction on the LDR sequence are reduced: images can be misaligned, scene content can be dynamic, and areas under or over exposed can be large.

The originality of this approach, when compared to state-of-the-art approaches is that the HDR reconstruction quality is not affected by the choice of the reference image in the multi-exposures sequence of LDR images. We have developed a three steps algorithm, first we detect target areas where HDR information is missing. Second, a search step is used to identify the best match between the target areas and the most informative LDR images in the multi-exposures sequence. Finally, the information in the matched area, is used to reconstruct the HDR values using an image inpainting approach.

We compared our approach to the state-of-the-art approaches, showing that our results are of comparable quality when the HDR areas that need to be reconstructed are characterized by small under/over-exposed areas in the corresponding LDR multi-exposures sequence. However, our results are outperforming the stat-of-the-art methods when the under/over-exposed areas are large.

6.2 Future Work

This section discusses new possible exploration directions in future research for each of the main contributions. There are several open problems linked to the techniques studied in this thesis. Dense correspondences for images with different exposures remains unsolved. 2D/3D HDR video using non HDR cameras requires solving the correspondences between consecutive frames. Some of the hardware prototypes on presented so far, acquires multiple exposures of the same scene and requires embedded software to do the HDR merging on real time. There are also open possibilities in applications to environments with extremely high dynamic range, like in spatial imaging, medical applications, or industrial welding.

The following subsections present possible improvements to evolve the techniques presented in this thesis and to make them more robust.

6.2.1 Dynamic scenes from static cameras

It was the first proposed and some recent methods might outperform it for general cases where both the camera and the scene are dynamic. Even though, an updated GPU implementation might provide a fast tool for HDR generation from dynamic scenes acquired from fixed camera.

An important improvement should be conducted on the registration step. The obtained results are good mostly for roughly rigid dynamic objects. For deformable objects this process might fail. This can be addressed directly with an adaptation of our proposed method, by subdividing the dynamic objects and approximate its deformation by matching small patches.

Some artifacts may still be introduced during the merging step. Neighbor areas resulting of combining a different amount of LDR images may contain visible borders due to inaccuracies in the CRF. To avoid such problems, the blending technique presented by Gallo *et al.* [GGC+09] could be considered.

6.2.2 Mutliscopic HDR

In-HDR-inpainting could be adapted to the stereo particularities. The epipolar geometry might help to constrain the search to provide better matches for contours and the interpolation might be guided by an homography obtained from the best matches of the contours.

6.2.3 In-HDR-painting

This method provides promising results to the HDR reconstruction from multiples LDR sequence. However, the proposed solution could be improved to be more robust in several aspects. A soft transition between areas is desired when the under/over exposed areas are not defined regions. When the saturation occurs gradually from properly exposed areas to saturated ones, a threshold over a weighted image might introduce visible artifacts between the original image and the replaced regions.

The interpolation step needs to evolve towards a more robust solution that includes the details of the surrounding areas instead of only the best matches in the contour.

6.3 Contributions

This sections provide a precise list of the publications related to this PhD thesis work:

- **Full high-dynamic range images for dynamic scenes.** Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos and Pere-Pau Vasquez. *Proceedings of the SPIE 8436, Optics, Photonics, and Digital Technologies for Multimedia Applications II.* doi:10.1117/12.922825. Brussels, Belguim. June 2012. [OMLV12] **(Honored with the Best Student Paper Award)**

- **Patch-based registration for auto-stereoscopic hdr content creation.** Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos and Alessandro Artusi. *In HDRi2013 - First International Conference and SME Workshop on HDR imaging.* Porto, Portugal. April 2013.[OMLA13]

- **Génération de séquences d'images multivues hdr: vers la vidéo hdr.** Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos and Alessandro Artusi. *In 27es journèes de l'Association française d'informatique graphique et du chapitre français d'Eurographics.* Reims, France. November 2014. [OMLA14]

- **Multiscopic HDR image sequence generation.** Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos, and Alessandro Artusi. *In Journal of WSCG, 23rd International Conference in Central Europe on Computer Graphics,Visualization and Computer Vision. 23(2):111-120.* Plzen, Czech Republic. June 2015.[OMLA15]

- **Chapter 4. Multi-view HDR video sequence generation.** Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos and Alessandro Artusi. *High Dynamic Range Video: Acquisition, Display and Applications.* Book chapter edited by Frèdèric Dufaux, Patrick Le Callet, Rafal Mantiuk and Marta Mrak. Elsevier Science. March 2016. *(to appear in [OMLA16])*

- **In-HDR-inpainting.** Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos, and Alessandro Artusi. *(To be submitted in a journal.)*

BIBLIOGRAPHY

[ADGM13] Cecilia Aguerrebere, Julie Delon, Yann Gousseau, and Pablo Musé.
Simultaneous HDR image reconstruction and denoising for dynamic scenes.
In *IEEE International Conference on Computational Photography*, pages 31–41, 2013.

[AKCG14] Tara Akhavan, Christian Kapeller, Ji-Ho Cho, and Margrit Gelautz.
Stereo hdr disparity map computation using structured light.
In *HDRi2014 Second International Conference and SME Workshop on HDR imaging*, 2014.

[Aky11] Ahmet Oğuz Akyüz.
Photographically guided alignment for hdr images.
In *Proceedings of EUROGRAPHICS 2011*, Wales, UK, April 2011.

[Anu70] P.E. Anuta.
Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques.
Geoscience Electronics, IEEE Transactions on, 8(4):353 –368, oct. 1970.

[AYG13] Tara Akhavan, Hyunjin Yoo, and Margrit Gelautz.
A framework for hdr stereo matching using multi-exposed images.
In *Proceedings of HDRi2013 First International Conference and SME Workshop on HDR imaging*, Paper no. 8, Oxford/Malden, 2013. The Eurographics Association and Blackwell Publishing Ltd.

[BADC11] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers.
Advanced High Dynamic Range Imaging: Theory and Practice.
AK Peters (CRC Press), Natick, MA, USA, 2011.

[BDA$^+$09] Francesco Banterle, Kurt Debattista, Alessandro Artusi, Sumanta Pattanaik, Karol Myszkowski, Patrick Ledda, and Alan Chalmers.
High dynamic range imaging and low dynamic range expansion for generating hdr content.
Computer Graphics Forum, 28(8):2343–2367, 2009.

[BLV$^+$12] Jennifer Bonnard, Celine Loscos, Gilles Valette, Jean-Michel Nourrit, and Laurent Lucas.
High-dynamic range video acquisition with a multiview camera.
Optics, Photonics, and Digital Technologies for Multimedia Applications II, pages 84360A–84360A–11, 2012.

[Bog00] Luca Bogoni.

Extending dynamic range of monochrome and color images through fusion.
In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 3, pages 7 –12 vol.3, 2000.

[Boi14] Ronan Boitard.
Temporal Coherency in Video Tone Mapping.
PhD thesis, IRISA, Université de Rennes 1, Rennes, France, 2014.

[BRG⁺14] Michel Bätz, Thomas Richter, Jens-Uwe Garbas, Anton Papst, Jürgen Seiler, and André Kaup.
High dynamic range video reconstruction from a stereo camera setup.
Signal Processing: Image Communication, 29(2):191 – 202, 2014.
Special Issue on Advances in High Dynamic Range Video Research.

[Bro92] Lisa Gottesfeld Brown.
A survey of image registration techniques.
ACM Comput. Surv., 24:325–376, December 1992.

[BRR11] Michael Bleyer, Christoph Rhemann, and Carsten Rother.
Patchmatch stereo - stereo matching with slanted support windows.
In *Proceedings of the British Machine Vision Conference*, pages 14.1–14.11. BMVA Press, 2011.
http://dx.doi.org/10.5244/C.25.14.

[BSCB00] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester.
Image inpainting.
In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 417–424, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[BSFG09] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman.
Patchmatch: A randomized correspondence algorithm for structural image editing.
ACM Transactions on Graphics (Proc. SIGGRAPH), 28(3), aug 2009.

[BSGF10] Connelly Barnes, Eli Shechtman, Dan B Goldman, and Adam Finkelstein.
The generalized PatchMatch correspondence algorithm.
In *European Conference on Computer Vision*, sep 2010.

[BVNL14] Jennifer Bonnard, Gilles Valette, Jean-Michel Nourrit, and Céline Loscos.
Analysis of the consequences of data quality and calibration on 3d hdr image generation.
In *European Signal Processing Conference (EUSIPCO)*, Lisbonne, Portugal, September 2014.

[CA06] BRIAN CLARK and ERGUN AKLEMAN.
Time lapse high dynamic range (hdr) photography.
Technical report, Visualization Sciences Program, Department of Architecture Texas A&M University, College Station, Texas, USA, February 2006.

[Can03] Frank M. Candocia.
Simultaneous homographic and comparametric alignment of multiple exposure-adjusted pictures of the same scene.
Image Processing, IEEE Transactions on, 12(12):1485–1494, Dec 2003.

[CBB$^+$09] Alan Chalmers, Gerhard Bonnet, Francesco Banterle, Piotr Dubla, Kurt Debattista, Alessandro Artusi, and Christopher Moir.
High-dynamic-range video solution.
In *ACM SIGGRAPH ASIA 2009 Art Gallery &#38; Emerging Technologies: Adaptation*, SIGGRAPH ASIA '09, pages 71–71, New York, NY, USA, 2009. ACM.

[Cer06] Lukàs Cerman.
High dynamic range images from multiple exposures.
Master Thesis, 2006.

[CHH04] W. R. Crum, T. Hartkens, and D. L. G. Hill.
Non-rigid image registration: theory and practice.
Br J Radiol, 77 Spec No 2:S140–53, 2004.

[CPT04] Antonio Criminisi, P. Perez, and K. Toyama.
Region filling and object removal by exemplar-based image inpainting.
Image Processing, IEEE Transactions on, 13(9):1200–1212, Sept 2004.

[CS01] Tony F. Chan and Jianhong Shen.
Morphologically invariant pde inpaintings, 2001.

[DM97] Paul Debevec and Jitendra Malik.
Recovering high dynamic range radiance maps from photographs.
In *In proceedings of ACM SIGGRAPH (Computer Graphics)*, volume 31, pages 369–378, 1997.

[DPPC13] Frederic Dufaux, Béatrice Pesquet-Popescu, and Marco Cagnazzo.
Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering.
John Wiley & Sons, 2013.

[Est12] Francisco J. Estrada.
Time-lapse image fusion.
In Andrea Fusiello, Vittorio Murino, and Rita Cucchiara, editors, *Computer Vision ‚Äì ECCV 2012. Workshops and Demonstrations*, volume 7584 of *Lecture Notes in Computer Science*, pages 441–450. Springer Berlin Heidelberg, 2012.

[Feh04] Christoph Fehn.
Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv.
In *Proc. SPIE*, volume 5291, pages 93–104, 2004.

[FP10] Yasutaka Furukawa and Jean Ponce.
Accurate, dense, and robust multiview stereopsis.

Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(8):1362–1376, Aug 2010.

[GGC^{+}09] Orazio Gallo, Netasha Gelfand, Wei-Chao Chen, Marious Tico, and Kari Pulli.
Artifact-free high dynamic range imaging.
IEEE International Conference on Computational Photography (ICCP), April 2009.

[GKTT13] Miguel Granados, Kwang In Kim, James Tompkin, and Christian Theobalt.
Automatic noise modeling for ghost-free hdr reconstruction.
ACM Trans. Graph., 32(6):201:1–201:10, November 2013.

[GN03a] Michael D. Grossberg and Shree K. Nayar.
Determining the camera response from images: what is knowable?
IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(11):1455 – 1467, nov 2003.

[GN03b] Michael D. Grossberg and Shree K. Nayar.
High dynamic range from multiple images: Which exposures to combine?
In *ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, Oct 2003.

[Gro06] Thorsten Grosch.
Fast and robust high dynamic range image generation with camera and object movement.
In *Vision, Modeling and Visualization, RWTH Aachen*, pages 277–284, 2006.

[GSL08] Miguel Granados, Hans-Peter Seidel, and Hendrik P. A. Lensch.
Background estimation from non-time sequence images.
In *Graphics Interface*, pages 33–40, 2008.

[Gut12] Benjamin Guthier.
Real-Time Algorithms for High Dynamic Range Video.
PhD thesis, Universität Mannheim, Mannheim, Germany, 2012.

[Har01] Paul Harrison.
A non-hierarchical procedure for re-synthesis of complex textures.
Journal of WSCG, 9th International Conference in Central Europe on Computer Graphics,Visualization and Computer Vision, pages 190–197, February 2001.

[HGP12] Jun Hu, Orazio Gallo, and Kari Pulli.
Exposure stacks of live scenes with hand-held cameras.
In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision ECCV 2012*, volume 7572 of *Lecture Notes in Computer Science*, pages 499–512. Springer Berlin Heidelberg, 2012.

[HGPS13] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun.
Hdr deghosting: How to deal with saturation ?
In *CVPR*, 2013.

[HLL+11] Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee, Youngsu Moon, and Joonhyuk Cha.
Ghost-free high dynamic range imaging.
In *Proceedings of the 10th Asian conference on Computer vision - Volume Part IV* , ACCV'10, pages 486–500, Berlin, Heidelberg, 2011. Springer-Verlag.

[HS88] Chris Harris and Mike Stephens.
A combined corner and edge detector.
In *The Alvey Vision Conference*, pages 147–151, 1988.

[HS09] Heiko Hirschmuller and Daniel Scharstein.
Evaluation of stereo matching costs on images with radiometric differences.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(9):1582–1599, 2009.

[HS11] K. Hirakawa and P.M. Simon.
Single-shot high dynamic range imaging with conventional camera hardware.
In *Computer Vision (ICCV), 2011 IEEE International Conference on* , pages 1339–1346, Nov 2011.

[HSGL11] Yoav HaCohen, Eli Shechtman, Dan B Goldman, and Dani Lischinski.
Non-rigid dense correspondence with applications for image enhancement.
ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2011), 30(4):70:1–70:9, 2011.

[HTM14] Kanita Karaduzovic Hadziabdic, Jasminka Hasic Telalovic, and Rafal Mantiuk.
Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging.
In *HDRi - International Conference and SME Workshop on HDR imaging*, Sarajevo, Bosnia and Herzegovina, 2014.

[htt15] http://www.hdrsoft.com/.
Photomatix 4.2.1, 2015.

[IP97] Homan Igehy and Lucas Pereira.
Image replacement through texture synthesis.
In *Image Processing, 1997. Proceedings., International Conference on* , volume 3, pages 186–189 vol.3, Oct 1997.

[JLW08] Katrien Jacobs, Celine Loscos, , and Greg Ward.
Automatic high-dynamic range generation for dynamic scenes.
IEEE Computer Graphics and Applications, 28:24,Äì33, March-April 2008.

[KAR06] Erum Arif Khan, Ahmet Oğuz Akyüz, and Erik Reinhard.
Ghost removal in high dynamic range images.
IEEE International Conference on Image Procesing, page 2005,Äì2008, October 2006.

[KCS02] Sung Ha Kang, Tony F. Chan, and Stefano Soatto.
Landmark based inpainting from multiple views.
Technical report, UCLA Math CAM, 2002.

[KS04] Sing Bing Kang and Richard Szeliski.
Extracting view-dependent depth maps from a collection of images.
International Journal of Computer Vision, 58(2):139–163, July 2004.

[KSB+13] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B. Goldman, and Pradeep Sen.
Patch-based high dynamic range video.
ACM Trans. Graph., 32(6):202:1–202:8, November 2013.

[KSD+14] Nima Khademi Kalantari, Eli Shechtman, Soheil Darabi, Dan B Goldman, and Pradeep Sen.
Improving Patch-Based Synthesis by Learning Patch Masks.
2014.

[KUWS03] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski.
High dynamic range video.
ACM Trans. Graph., 22(3):319–325, 2003.

[LC09] Huei-Yung Lin and Wei-Zhe Chang.
High dynamic range imaging for stereoscopic scene representation.
In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 4305–4308, 2009.

[Lew95] J. P. Lewis.
Fast normalized cross-correlation.
In *Vision Interface (1995)*, pages 120–123, 1995.

[LHS87] J. Lee, R.M. Haralick, and L.G. Shapiro.
Morphologic edge detection.
Robotics and Automation, IEEE Journal of, 3(2):142–156, April 1987.

[LJ10] Celine Loscos and Katrien Jacobs.
High-dynamic range imaging for dynamic scenes.
In Rastislav Lukac, editor, *Computational Photography, Methods and Applications*, pages 259–281. CRC Press, October 2010.

[LLR13] Laurent Lucas, Céline Loscos, and Yannick Remion.
3D Video from Capture to Diffusion.
Wiley-ISTE, October 2013.

[LRZ+10] Zhengguo Li, S. Rahardja, Zijian Zhu, Shoulie Xie, and Shiqian Wu.
Movement detection for the synthesis of high dynamic range images.
In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 3133 –3136, sept. 2010.

[Mar09] Mateusz Markowski.
Ghost removal in hdri acquisition.
In *Central European Seminar on Computer Graphics*, Budmerice Castle, Slovakia, 2009.

[MFS14] Jaume Rigau Qing Xu Miquel Feixas, Anton Bardera and Mateu Sbert.
Information theory tools for image processing.
Synthesis Lectures on Computer Graphics and Animation, 6(1):1–164, 2014.

[MG10] Stephen Mangiat and Jerry Gibson.
High dynamic range video with ghost removal.
Proc. SPIE, 7798:779812–779812–8, 2010.

[Mid06] Middlebury.
Middlebury stereo datasets.
`http://vision.middlebury.edu/stereo/data/`, 2006.

[MKRH11] Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich.
Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions.
ACM Trans. Graph., 30(4):40:1–40:14, July 2011.

[MN99] T. Mitsunaga and S.K. Nayar.
Radiometric self calibration.
IEEE International Conf. Computer Vision and Pattern Recognition, 1:374–380, 1999.

[MP95] Steve Mann and R. W. Picard.
On Being undigital With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures.
Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology, 1995.

[MPS12] Bernard Mendiburu, Yves Pupulin, and Steve Schklair, editors.
3D {TV} and 3D Cinema.
Focal Press, Boston, 2012.

[MSZ$^+$11] Xing Mei, Xun Sun, Mingcai Zhou, shaohui Jiao, Haitao Wang, and Xiaopeng Zhang.
On building an accurate stereo matching system on graphics hardware.
In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 467–474, Nov 2011.

[MV98] J B Maintz and M A Viergever.
A survey of medical image registration.
Medical Image Analysis, 2(1):1–36, 1998.

[NB03] Shree K Nayar and Vlad Branzoi.
Adaptive dynamic range imaging: Optical control of pixel exposures over space and time.
In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 1168–, Washington, DC, USA, 2003. IEEE Computer Society.

[OBMsC01] Manuel M. Oliveira, Brian Bowen, Richard McKenna, and Yu sung Chang.
Fast digital image inpainting.

In *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON VISUALIZATION, IMAGING AND IMAGE PROCESSING (VIIP 2001*, pages 261–266. ACTA Press, 2001.

[OMLA13] Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos, and Alessandro Artusi. Patch-based registration for auto-stereoscopic hdr content creation. In *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, Oporto Portugal, April 2013.

[OMLA14] Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos, and Alessandro Artusi. Génération de séquences d'images multivues hdr: vers la vidéo hdr. In *27es journées de l'Association française d'informatique graphique et du chapitre français d'Eurographics*, Reims, France, November 2014.

[OMLA15] Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos, and Alessandro Artusi. Multiscopic hdr image sequence generation. *Journal of WSCG, 23rd International Conference in Central Europe on Computer Graphics,Visualization and Computer Vision*, 23(2):111–120, June 2015.

[OMLA16] Raissel Ramirez Orozco, Ignacio Martin, Celine Loscos, and Alessandro Artusi. Chapter 4. multi-view hdr video sequence generation. In F. Dufaux, P.L. Callet, R. Mantiuk, and M. Mrak, editors, *High Dynamic Range Video: Acquisition, Display and Applications*. Elsevier Science, 2016.

[OMLV12] R. Ramirez Orozco, I. Martin, C. Loscos, and P.-P. Vasquez. Full high-dynamic range images for dynamic scenes. *Optics, Photonics, and Digital Technologies for Multimedia Applications II* , 8436(1):843609, 2012.

[Ope15] OpenCV. Opencv 3.0(open source computer vision library). `http://opencv.org/`, 2015.

[OTTE15] Aykut Erdem Okan Tarhan Tursun, Ahmet Oğuz Akyüz and Erkut Erdem. The state of the art in hdr deghosting: A survey and evaluation. In *Proceedings of EUROGRAPHICS 2015*, Zürich, Switzerland, may 2015.

[PcPD$^+$10] Jessica Prévoteau, Sylvia Chalenç con Piotin, Didier Debons, Laurent Lucas, and Yannick Remion. Multi-view shooting geometry for multiscopic rendering with controlled distortion. *International Journal of Digital Multimedia Broadcasting (IJDMB), special issue Advances in 3DTV: Theory and Practice*, 2010:1–11, March 2010.

[PH08] Matteo Pedone and Janne Heikkil. Constrain propagation for ghost removal in high dynamic range images. In *Proc. Third International Conference on Computer Vision Theory and Applications (VISAPP 2008)*, volume 1, pages 36–41, Madeira, Portugal, 2008.

[PK10] F. Pece and J. Kautz.
Bitmap movement detection: Hdr for dynamic scenes.
In *Visual Media Production (CVMP), 2010 Conference on*, pages 1–8, Nov 2010.

[RC11] Shanmuganathan Raman and Subhasis Chaudhuri.
Reconstruction of high contrast images for dynamic scenes.
The Visual Computer, 27:1099–1114, 2011.
10.1007/s00371-011-0653-0.

[Red06] RedCompany.
Read one.
`http://www.red.com`, 2006.

[Res05] Vision Research.
Phantom hd.
`http://www.visionresearch.com`, 2005.

[RKC09] Shanmuganathan Raman, Vishal Kumar, and Subhasis Chaudhuri.
Blind de-ghosting for automatic multi-exposure compositing.
In *ACM SIGGRAPH ASIA 2009 Posters* , SIGGRAPH ASIA '09, pages 44:1–44:1, New York, NY, USA, 2009. ACM.

[RKMS15] Karol Myszkowski Rafal K. Mantiuk and Hans Peter Seidel.
High dynamic range imaging.
In J. G. Webster, editor, *Wiley Encyclopedia of Electrical and Electronics Engineering* . John Wiley and Sons, April 2015.

[RR08] Roshni and Revathy.
Using mutual information and cross correlation as metrics for registration of images.
4(6), June 2008.

[RS84] B. Rezaie and M. D. Srinath.
Algorithms for fast image registration.
IEEE Transactions on Aerospace and Electronic Systems, AES-20(6):716 –728, nov. 1984.

[Ruf11] Dominic Rufenacht.
Stereoscopic High Dynamic Range Video.
PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, Agost 2011.

[RWD$^+$10] Erik Reinhard, Greg Ward, Paul Debevec, Sumanta Pattanaik, Wolfgang Heidrich, and Karol Myszkowski.
High Dynamic Range Imaging, 2nd edition.
Morgan Kaufmann Publishers, San Francisco, 2010.

[SCSI08] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani.
Summarizing visual data using bidirectional similarity.
IEEE Conference on Computer Vision and Pattern Recognition 2008 (CVPR'08), 2008.

[SDBRC13] Elmedin Selmanović, Kurt Debattista, Thomas Bashford-Rogers, and Alan Chalmers. Generating stereoscopic hdr images using hdr-ldr image pairs. *ACM Trans. Appl. Percept.*, 10(1):3:1–3:18, March 2013.

[SDBRC14] Elmedin Selmanovic, Kurt Debattista, Thomas Bashford-Rogers, and Alan Chalmers. Enabling stereoscopic high dynamic range video. *Signal Processing: Image Communication*, 29(2):216 – 228, 2014. Special Issue on Advances in High Dynamic Range Video Research.

[Sha48] Claude E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27:379–423, July 1948.

[SHS$^+$04] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matt Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. High dynamic range display systems. In *Proc. of SIGGRAPH '04 (Special issue of ACM Transactions on Graphics)*, August 2004.

[SKY$^+$12] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B. Goldman, and Eli Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012)*, 31(6):203:1–203:11, November 2012.

[SL12] Simon Silk and Jochen Lang. Fast high dynamic range image deghosting for arbitrary scene motion. In *Proceedings of Graphics Interface 2012*, GI '12, pages 85–92, Toronto, Ont., Canada, Canada, 2012. Canadian Information Processing Society.

[SMA78] M. Svedlow, C. D. Mcgillem, and P. E. Anuta. Image registration: Similarity measure and preprocessing method comparisons. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-14(1):141 –150, jan. 1978.

[SMW10] N. Sun, H. Mansour, and R. Ward. Hdr image construction from multi-exposed stereo ldr images. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Hong Kong, 2010.

[SPS09] Desire Sidibé, William Puech, and Olivier Strauss. Ghost detection and removal in high dynamic range images. In *17th European Signal Processing Conference. EUSIPCO 2009*, Glasgow, Scotland, August 2009. HAL - CCSD.

[SS02] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7,Äì42, May 2002.

[SS12] Abhilash Srikantha and D. Sidibé, Désiré.
Ghost Detection and Removal for High Dynamic Range Images: Recent Advances.
Signal Processing: Image Communication, page 10.1016/j.image.2012.02.001, February 2012.
23 pages.

[ST04] Peter Sand and Seth Teller.
Video matching.
ACM Transactions on Graphics, 23(3):592–599, August 2004.

[Süh08] Karsten Sühring.
H.264/avc reference software.
`http://iphome.hhi.de/suehring/tml/`, 2008.

[SZS03] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum.
Stereo matching using belief propagation.
IEEE Trans. Pattern Anal. Mach. Intell., 25(7):787–800, July 2003.

[Tho05] ThomsomGrassValley.
Viper filmstream.
`http://www.ThomsomGrassValley.com`, 2005.

[TKS06] A. Troccoli, Sing Bing Kang, and S. Seitz.
Multi-view multi-exposure stereo.
In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 861–868, June 2006.

[TKTS11] Michael D. Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen.
A versatile HDR video production system.
ACM Transactions on Graphics, 30(4):41:1–41:10, July 2011.

[TM07] Anna Tomaszewska and Radoslaw Mantiuk.
Image registration for multi-exposure high dynamic range image acquisition.
In Prof. Vaclav Skala, editor, *Proc. Int'l Conf. Central Europe on Computer Graphics, Visualization, and Computer Vision (WSCG)*. University of West Bohemia, 2007.

[UCES11] Hakan Urey, Kishore V. Chellappan, Erdem Erden, and Phil Surman.
State of the art in stereoscopic and autostereoscopic displays.
Proceedings of The IEEE, 99:540–555, 2011.

[VW95] Paul Viola and W.M. Wells.
Alignment by maximization of mutual information.
In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 16 –23, jun 1995.

[War03] Greg Ward.
Fast, robust image registration for compositing high dynamic range photographs from handheld exposures.
Journal of Graphics Tools, 8:17–30, 2003.

[WPA09] Medha V. Wyawahare, Pradeep M. Patil, and Hemant K. Abhyankar.
Image registration techniques: An overview.
2009.

[Yao11] Susu Yao.
Robust image registration for multiple exposure high dynamic range image synthesis.
volume 7870, pages 78700Q–78700Q–9, 2011.

[ZBW11] Henning Zimmer, Andrés Bruhn, and Joachim Weickert.
Freehand hdr imaging of moving scenes with simultaneous resolution enhancement.
Computer Graphics Forum, 30(2):405–414, 2011.

[ZF03] Barbara Zitova and Jan Flusser.
Image registration methods: a survey.
Image and Vision Computing, 21:977–1000, 2003.

[ZFS05] Barbara Zitová, Jan Flusser, and Filip Sroubek.
Image registration methods: a survey.
In *Proceedings of the International Conference on Image Processing*. IEEE, September 2005.