CONTRIBUIÇÕES À RESTAURAÇÃO AUTOMÁTICA DE IMAGENS DE CENAS EM MEIO PARTICIPATIVO

PAULO LILLES JORGE DREWS JUNIOR

CONTRIBUIÇÕES À RESTAURAÇÃO AUTOMÁTICA DE IMAGENS DE CENAS EM MEIO PARTICIPATIVO

Tese apresentada ao Programa de Pós--Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

Orientador: Mario Fernando Montenegro Campos
Coorientador: Erickson Rangel do Nascimento

Belo Horizonte
05 de fevereiro de 2016

PAULO LILLES JORGE DREWS JUNIOR

CONTRIBUTIONS TO THE AUTOMATIC RESTORATION OF IMAGES FROM SCENES IN PARTICIPATING MEDIA

Thesis presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Advisor: Mario Fernando Montenegro Campos
Co-Advisor: Erickson Rangel do Nascimento

Belo Horizonte
February 5, 2016

Drews Junior, Paulo Lilles Jorge

D776c Contribuições à Restauração Automática de Imagens de Cenas em Meio Participativo / Paulo Lilles Jorge Drews Junior. — Belo Horizonte, 2016
xxviii, 137 f. : il. ; 29cm

Tese (doutorado) — Universidade Federal de Minas Gerais— Departamento de Ciência da Computação
Orientador: Mario Fernando Montenegro Campos
Coorientador: Erickson Rangel do Nascimento

1. Computação — Teses. 2. Visão por Computador — Teses. 3. Sistemas de Imagens Subaquáticas – Teses. I. Orientador. II. Coorientador. III. Título.

CDU 519.6*82.10(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Contribuições à restauração automática de imagens de cenas em meio participativo

PAULO LILLES JORGE DREWS JUNIOR

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Mario Fernando Montenegro Campos - Orientador
Departamento de Ciência da Computação - UFMG

Prof. Erickson Rangel do Nascimento - Coorientador
Departamento de Ciência da Computação - UFMG

Prof. Alberto Elfes
Data61 Group - CSIRO

Prof. Flávio Luis Cardeal Pádua
Departamento de Computação - CEFET/MG

Prof. Guilherme Augusto Silva Pereira
Departamento de Engenharia Elétrica - UFMG

Prof. Siome Klein Goldenstein
Instituto de Computação - UNICAMP

Belo Horizonte, 05 de fevereiro de 2016.

This work is dedicated to the memory of my grandfather Afonso Pedrotti, who supported my academic work.

Acknowledgments

First, I would like to thank my parents for all the support they provided me over the years. In addiction, I thank to Isabela who motivated me to the Ph.D..

Next, I would like to thank Prof. Mario Campos and Prof. Erickson Nascimento took their time to share their knowledge and provided a lot of valuable contributions to this work.

I thank to Dr. Alberto Elfes that received me in CSIRO and provided me the OpenROV used in the experiments. And, I would like to thank Dr. Emili Hernández from CSIRO for help me with the experiments using OpenROV, specially the obstacle avoidance tasks. I would also like to thank Prof. Siome Goldenstein for the suggestion to use the Continuous Essential Matrix to estimate the structure-from-motion. Furthermore, I would like to thank Prof. Armando Neto his priceless help in the experimental results with the OpenROV in the Cipo River, and fundamental discussions since the beginning of the present work.

I also thank the staff of C3-FURG, mainly Prof. Nelson Duarte and Prof. Silvia Botelho, for helped me with many bureaucratic issues and provided all support for the development of this work. I thank all of my colleagues from C3-FURG for their assistance with students and lectures during my long period to pursue the Ph.D., and the students/alumni from NAUTEC/FURG. I would like to thank Silvia Botelho, Felipe Moraes and Joel Gaya the cooperation in the beginning of the work, mainly with the UDCP. And, I would like to thank to researchers from FURG and INCT-Mar COI: Vagner Rosa, Osmar Moller Jr., Lauro Calliari, Gilberto Griep, José Henrique Muelbert, Matheus Longaray, Diones Fisher, Felipe Almeida, Milton Araújo, Augusto Cavalcanti, and Ricardo Costa for helped me with the experiments using the ROV-C3 in the Parcel do Carpinteiro reef aboard the Oceanographic Research Ship Atlantico Sul.

I would like to thank my colleagues from VeRLab-UFMG their valuable contribution to this dissertation, specially Levi Ostero, David Saldaña, Omar Vidal, Anderson Tavares, Daniel Balbino and Rafael Colares. I also thank all colleagues from ASL-

CSIRO the wonderful time during my days in Australia, specially Paulo Borges, Pavan Sikka, Ryan Steindl, Tirthankar Bandyopadhyay, David Howard, Jorge Diosdado and Adrián Rechy.

I would also like to thank the staff of PPGCC-UFMG and FURG for putting up with me and answering all my questions.

I would also like to acknowledge the professors graciously agreed to serve on my committee: Prof. Alberto Elfes, Prof. Siome Klein Goldenstein, Prof. Guilherme Augusto Silva Pereira, and Prof. Flávio Luis Cardeal Pádua.

Finally, I would like to thank FURG, CAPES and CNPQ, for the financial support during the development of this Ph.D. dissertation.

“For every complex problem there is an answer that is clear, simple, and wrong.”
(H. L. Mencken)

Resumo

O presente trabalho lida com o problema de restauração de sequências de imagens monoculares adquiridas em meios participativos, ou seja, meios que interferem na propagação da luz. Especificamente, o trabalho foca na restauração automática de imagens adquiridas em cenas subaquáticas ou enevoadas, de modo a permitir o uso apropriado de algoritmos já existentes na área de Visão Computacional.

A metodologia proposta para restauração requer como entrada uma sequência de imagens, isto é, ao menos um par de imagens de uma mesma cena. O método produz como saída uma outra sequência de imagens na qual os efeitos do meio são atenuados e a visibilidade melhorada. Para tal, é feita a estimação automática dos mapas de profundidade da cena e dos coeficientes de atenuação. Tais informações permitem a inversão do modelo e a restauração das imagens. Diferentemente de trabalhos anteriores, o presente trabalho não necessita de equipamentos ou informações adicionais, apenas de uma câmera calibrada.

O presente trabalho explora diversas fontes de informação: temporal, geométrica e ambiental. A informação temporal é obtida usando uma nova formulação de fluxo ótico que permite lidar com meios participativos e não-participativos. A informação ambiental é recuperada utilizando um novo *prior* estatístico que permite obter a transmissão do meio em imagens subaquáticas. As restrições geométricas fornecem as ferramentas necessárias para recuperar a informação de profundidade. Baseado no modelo ótico e em um conjunto de restrições, um novo método estocástico é proposto para estimar o mais importante parâmetro do meio: o coeficiente de atenuação. Esses dados permitem a estimação do mapa de profundidade e a restauração da sequência de imagens utilizando apenas a informação visual fornecida por uma câmera monocular.

A metodologia proposta é avaliada qualitativamente e quantitativamente. Considerando as dificuldades inerentes ao meio, os resultados mostram que o método é capaz de estimar o coeficiente de atenuação e efetuar a restauração. A restauração obteve melhores resultados quando comparado com outras técnicas usando diversas métricas de avaliação.

Palavras-chave: Restauração de Imagens, Visão Computacional, Meios Participativos, Visão Subaquática, Reconstrução Tridimensional.

Abstract

This work deals with the problem of image restoration of monocular sequences of images acquired in participating media, *i.e.* media that interfere with light propagation. Specifically, the proposed work focus on the automatic restoration of images acquired in underwater and hazy scenes, thus enabling the prompt use of legacy Computer Vision.

The proposed restoration process requires a sequence of images, *i.e.* at least a pair of images from the same scene with relative motion, as input and produces as output an image sequence in which the participating medium effects are attenuated and the visibility improved. For reach these purpose, we automatically estimate the depth map and the attenuation parameters. These data enable the inversion of the optical model, and thus to restore the images. Differently from previous works, our method does not need additional equipment or information – only a calibrated camera.

We explored several sources of information: temporal, geometrical and environmental. The temporal information is obtained using a new optical flow formulation that allows us to tackle both non-participating and participating media. The environmental information is recovered using a new statistical prior to obtain the medium transmission in underwater medium. The geometrical constraints provide the tools to recover the depth information. Based on the optical model and a set of constraints, we propose a new stochastic method to estimate the most critical parameter of the medium: the attenuation coefficient. These data allow us to estimate the depth map and restore the image sequence using only the visual information provided by a monocular camera.

Experimental results in real and simulated environments are carried out in order to validate the proposed methodology. Taking into account the inherent difficulties of the participating medium, the results show that our method is able to recover the attenuation parameter and restore the images. The restoration provides better results than other techniques assuming several evaluation metrics.

Palavras-chave: Image Restoration, Computer Vision, Participating Media, Underwater Vision, 3D Reconstruction.

List of Figures

List of Tables

List of Abbreviations

ASL	Autonomous Systems Laboratory
BA	Bundle Adjustment
BP	Bianco's prior
CEM	Continuous Essential Matrix
CL-NL	Classic-NL optical flow method
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DCP	Dark Channel Prior
DEM	Discrete Essential Matrix
FURG	Universidade Federal do Rio Grande
GDIM	Generalized Dynamic Image Model
GOFM	Generalized Optical Flow Model
HS	Horn-Schunk optical flow method
IQA	Image Quality Assessment
LoS	Line of Sight
LK	Lucas-Kanade optical flow method
MDCP	Median Dark Channel Prior
MN	Madjidi-Negahdaripour optical flow method
MRF	Markov Random Field
MSAC	M-estimator SAmple and Consensus
NAUTEC	Intelligent Robotics and Automation Group
OSD	On-Screen Display
RANSAC	RANdom SAmple Consensus
RMSE	Root-Mean-Square Error
SIFT	Scale-Invariant Feature Transform
ROV	Remotely Operated Vehicle
UDCP	Underwater Dark Channel Prior
UFMG	Universidade Federal de Minas Gerais
VeRLab	Laboratory of Computer Vision & Robotics

Contents

Chapter 1

Introduction

In the last years, the fields of computer vision and image processing have been experienced intensive efforts of research. A large number of techniques to extract information from images have been developed [Szeliski, 2010]. Regardless of these advances, almost all of these techniques are based on the premise that scene and camera are immersed in a non-participating medium, *e.g.* the clean air. It is assumed that the light rays travel through the medium without any alteration. However, there are some media that change the intensity and the direction of the light rays, called participating or scattering media.

Several real world problems need to tackle with images acquired in participating media, *e.g* surveillance [Busch and Debes, 1998; Walther et al., 2004; Nery et al., 2005; Hautière et al., 2008a], mapping [Botelho et al., 2009; Hudzietz and Saripalli, 2011; Shihavuddin et al., 2013], autonomous robots and vehicles [Pomerleau, 1997; Ridao et al., 2010; Pavlic et al., 2013; Drews-Jr et al., 2013] to name a few. These applications can take advance from legacy computer vision and image processing algorithms. Indeed, two different approaches may be adopted to deal with images acquired in a participating medium: a) developing new algorithms for each different task that take into account the medium's effects; b) restoring the image by removing the medium's effects that allows to use classical algorithms to solve the task. The latter is adopted in this dissertation because of its generality.

Light is absorbed and scattered by the medium before reaching the camera in a participating medium. These effects become a serious issue because they reduce the overall contrast on images and cause color shifting, directly impacting on the reduction of the visibility. This degradation worsens exponentially with distance making it impossible to acquire meaningful information from images of scenes with large distances. Thus, the main challenge of image restoration methods is to obtain the appearance


(a) (b) (c)

Figure 1.1. Example of images acquired in several participating media: a) dust storm, b) outdoor under fog and c) underwater.

that the scene would have whether it was not immersed in a participating medium. These methods are known in the literature by different names as image and visibility restoration, dehazing, descattering, deweathering, defogging, etc.

The most common participating media are fluids like water, milk, etc.; and poor weather conditions like haze, fog, mist, dust, rain, hail, snow, etc. Some examples of images acquired in participating media are shown in Fig. 1.1. The two most important characteristics that differentiate these media are the particle size and the attenuation rate [McCartney, 1976]. They present a significant impact in the degradation of the acquired image. Although these variety of environments suffer with absorption and scattering effects, most of the methods in the literature are specialized on the problem on fog/mist/haze and underwater because of their relevance. Typically, the underwater medium is more difficult to deal because of the large variability in the particle size and high absorption rates. Accordingly, this dissertation focuses in these two participating media: fog/mist/haze and water.

Underwater images are specially relevant when one considers that oceans cover about 70% of the Earth's surface and contain the most fundamental ecosystems of the planet. They constitute a rich source of biologic and mineral resources that are important to life. Moreover, oceans play a central role in the climatic balance of the planet, transporting heat from the Equatorial regions to Temperate zones by the motion of large water volumes. Although oceans are fundamental to the existence of the human race, they are largely unexplored, mainly because of the inherent difficulties faced by human beings, such as: high pressure, low temperatures and overall hazardous conditions. It is here that computer vision stands out as important tools to overcome these hindrances [Bellingham and Rajan, 2007].

Assuming the optical model, an image taken in participating medium is the result of attenuated scene color and global illumination, where both depend on the distances in the scene. These components generate an inherent bilinearity that makes the inverse problem fundamentally ill-posed, *i.e.* there is no uniqueness or stability of solutions. Typically, methods in the literature assume one of these two unknowns is given or constrained. The present method achieves the restoration by exploring several sources of information contained on the images: temporal, geometrical and environmental.

Although image restoration is not a largely explored field of research in Brazil, the Laboratory of Computer Vision & Robotics (VeRLab) from UFMG is working in this field since 2004, obtaining relevant results in underwater medium. The initial approaches were based on stereo vision [Queiroz-Neto et al., 2004; Queiroz-Neto, 2005; Nascimento, 2008; Nascimento et al., 2009]. Barros [2010] proposed a single image restoration method based on light propagation model and optimization techniques. All of these methods are detailed in Chap. 2. This dissertation follows this research line but dealing with monocular sequences of images.

1.1 Problem Definition

The problem to be faced in this dissertation is a model-based restoration of the image quality without any prior knowledge about the scene using a temporal sequence of images acquired in a participating medium with relative motion between viewpoint by a calibrated camera, *i.e.* the intrinsic parameters are previously known. We understand by image quality the improvement in terms of color fidelity and visibility, and visibility means the distance which one can reliably discriminate an object in the scene. However, we assume medium and illumination are homogeneous, *i.e.* the parameters of the model are constant on the entire three-dimensional space.

1.2 Objectives

The main objective of this dissertation is to develop a method capable of restoring a sequence of images acquired in participating media. We understand by restoration of an image the improvement of the visibility and the image quality. The method should be automatic and only based on color images acquired by a monocular camera. Furthermore, it should not require any additional device or sophisticated equipment such as polarizers. Participating medium is assumed as homogeneous, *i.e.* both illumination and medium characteristics are constant in the scene.

The objective is to estimate depth maps from sequence of images using geometric constraints associated with the information provided by the participating medium. The depth map is an image that contains information relating to the distance of the surfaces of scene from the optical center of the camera. Depending on the research area, the depth map is also known as depth buffer, Z-depth or Z-buffer. We adopted the term depth map.

Since the participating medium modulates the information carried from a scene point to the camera, *i.e.* the degradation is directly related to the distance between scene and camera. Therefore, we can exploit this information to develop a method for recovering scene properties, such as three-dimensional structure and the true scene appearance. Thus, we assumed the problem of restoration as the problem of estimating the medium parameters and the depth map of the scene.

Most restoration methods are based on additional information or single images. The methods based on additional information remains impractical because of the systems' complexity or because they are specific for a particular application. Although, single image methods can get impressive results for some images, they fail in some scenarios because the assumptions are not general for real world applications. Hence, the objective of the methodology proposed in this dissertation is develop a simple and general method to overcome these limitations.

We may think in the problem of image restoration as an "egg-and-chicken" problem. It is necessary to know the depth map to compute the medium transmission and to restore the image. Conversely, estimating the depth map by triangulation is only possible whether the corresponding points are known, which are hard to obtain because the photometric constraint is not valid in images acquired in participating media. Therefore, we aim to formulate the restoration problem as an alternate problem between depth and point matching estimation.

1.3 Contributions and Publications

The main contribution of this dissertation can be summarized as:

- We propose a new method to deal with the restoration of monocular sequences acquired in participating medium. The method uses temporal, geometric and environmental information to achieve the image restoration.

- We propose a new statistical prior for underwater imaging that allow to estimate the medium transmission and restore the images.

- We develop a new generalized formulation for optical flow estimation that allows the estimation in participating and non-participating medium.

- We develop a new method to estimate the attenuation coefficient based in a robust set of constraints.

The results of this work were partly published in international conferences and an international journal:

- Drews-Jr, P. L. J. ; Nascimento, E. R. ; Botelho, S. S. C.; Campos, M. F. M. . Underwater Depth Estimation and Image Restoration Based on Single Images. IEEE Computer Graphics and Application, vol. 36(2), pp. 50-61, 2016, ISSN: 0272-1716.

- Drews-Jr, P. L. J. ; Nascimento, E. R. ; Campos, M. F. M. ; Elfes, A. . Automatic Restoration of Underwater Monocular Sequences of Images. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1058-1064, 2015, Hamburg, Germany, 2015.

- Drews-Jr, P. L. J. ; Nascimento, E. R. ; Xavier, A. ; Campos, M. F. M. . Generalized Optical Flow Model for Scattering Media. In the Proceedings of the 22nd International Conference on Pattern Recognition (ICPR), pp. 3999-4004, 2014, Stockholm, Sweden.

- Drews-Jr, P. L. J. ; Nascimento, E. R. ; Moraes, F. C. ; Botelho, S. S. C.; Campos, M. F. M. . Transmission Estimation in Underwater Single Images. In the Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW) - Workshop on Underwater Vision, pp. 825-830, 2013, Sydney, Australia.

1.4 Organization

This dissertation is organized as follows. After reviewing the optical model of images acquired in participating media and the state-of-art in Chapter 2, Chapter 3 presents the proposed methodology. The experimental framework and the obtained results using simulated and real sequences of images are presented in Chapter 4. Finally, in Chapter 5, the main conclusions and future works are drawn.

Chapter 2

Related Works

Although there are many works trying to characterize the behavior of images acquired in participating media [Middleton, 1952; Preisendorfer, 1976; McCartney, 1976; McGlamery, 1980], two works needs to be distinguished by their innovation for the field of computer vision and image processing. The first one is the work presented by Cozman and Krotkov [1997] which uses a scattering model to compute depth cues. Although this method does not deal with the problem of image restoration, it was one of the first to try to understand what we can do for vision systems in participating media. Another pioneer work in the field was proposed by Oakley and Satherley [1998], where a model-based approach was applied to restore the contrast and visibility of image acquired in participating media. The method outperforms previous methodologies based on histogram equalization, though their method makes strong assumptions about the knowledge of the scene geometry.

Previously, some works like Kocak and Caimi [2005] and Kocak et al. [2008] review the advances in underwater imaging, dealing with the underwater problem in a broad sense, mainly related to new technologies and systems. They highlighted the problem of improving the visibility as one of the main problem in underwater imaging.

We proposed a taxonomy attempting to classify the works in the field of image restoration in participating media. The methods are divided in two high-level classes: only visual image and additional information. The *only visual image* class contains the restoration methods based only on images, and the *additional information* contains the methods that need additional information to restore the images. Furthermore, we identify the individual components and design decision. Fig. 2.1 illustrates the proposed taxonomy that facilitates the classification of the methods and allows comparing them.

Image restoration in participating media is strongly related to the research on op-

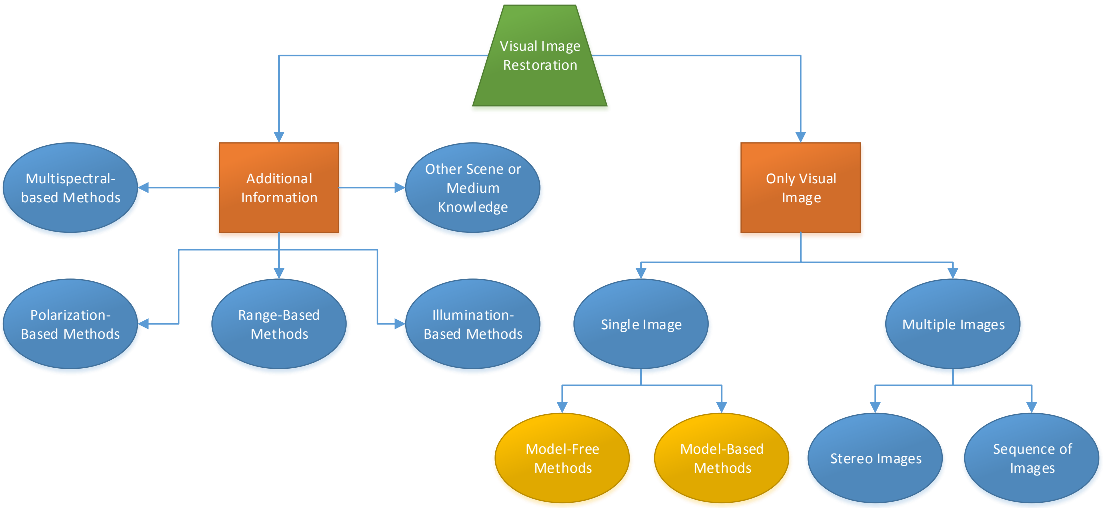


Figure 2.1. Proposed Taxonomy in order to classify visual image restoration methods.

tics [McGlamery, 1980; Jaffe, 1990; Kokhanovsky and Zege, 2004; Grabner and Kvicera, 2014], computer graphics [Sun et al., 2005; Narasimhan et al., 2006; Gutierrez et al., 2008; Darles et al., 2011], remote sensing [Fraser and Kaufman, 1985; Kaufman et al., 2001; Mahiny and Turner, 2007], etc. There are similarities in the models from these works that the restoration methods are based on. We review works from these correlated areas when restored images were obtained in real scenarios and the camera and scene are immersed in participating media.

The following section is addressed to review the optical model of participating media, which is the basis of almost all reviewed works. Furthermore, a unified description of the image formation in participating media is provided.

2.1 Optical Model of Participating Media

Images acquired from scene immersed in participating media may be modeled as a complex interaction between the light, the medium and the three-dimensional structure. One of the most popular models in the literature was proposed by Jaffe-McGlamery [McGlamery, 1980; Jaffe, 1990]. The model is designed as the sum of three main components: Direct illumination (I_d), Forward-scattering (I_{fs}) and Backscattering (I_{bs}), as shown in Eq. 2.1. These effects are illustrated in Fig. 2.2.

$$I = I_d + I_{fs} + I_{bs}. \tag{2.1}$$

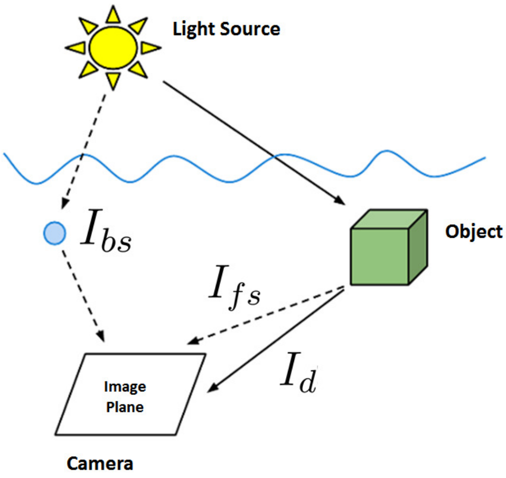


Figure 2.2. Main effects of the interaction between light and participating medium. The direct component I_d captures the fraction of light that reaches the sensor directly from the scene. The forward scattering component I_{fs} represents the reflected light in the scene that suffers scattering in the return path until the camera. Finally, the backscattering effect I_{bs} results from the interaction between the ambient illumination sources and particles dispersed in the medium.

Part of the light that radiates from object is lost because of scattering and absorption. The direct component captures this effect by modeling the fraction of light that reaches the sensor as:

$$I_d = Je^{-\eta d} = Jt, \tag{2.2}$$

where J and d are the scene radiance and the distance between scene and camera, respectively; η is the attenuation coefficient (frequently c in the literature) and t is the medium transmission or only transmission. The term ηd is also known as optical depth, and it is called b in some works.

In foggy scenes, the attenuation coefficient η is only composed of the scattering coefficient, called β [He et al., 2009]. For underwater environments, the absorption represents an inherent property of the medium [Jaffe, 1990], thus the coefficient η is the summation of the absorption, a, and the total scattering, β, coefficients, both being dependent on the wavelength λ, *i.e.* $a(\lambda)$ and $\beta(\lambda)$. The scattering coefficient is typically assumed constant with respect to the wavelength [Narasimhan and Nayar, 2002], since the wavelength dependence is stronger in the absorption coefficient than in the scattering coefficient [Vasilescu et al., 2010, 2011]. The depth values d may be estimated using the medium transmission t, but the constant η is usually unknown. However, it can be measured by using a turbidimeter, a Spectralon [Ahlen, 2003] or estimated by calibration procedures [Negahdaripour et al., 2002; Queiroz-Neto et al., 2004].

The forward scattering component, I_{fs}, also named the glow field, represents the reflected light in the scene that suffers scattering in the return path until the camera. This process generates a blurring effect on the image. The forward scattering is derived from the non-scattering field through convolution with a point-spread function. Eq. 2.3 shows the commonly employed model of the forward scattering [McGlamery, 1980; Zhang and Negahdaripour, 2002; Trucco and Olmos-Antillon, 2006].

$$\begin{aligned} I_{fs} &= I_d * g(x,y|d,G,\eta,B), \\ &= I_d[e^{-Gd} - e^{-\eta d}]\mathcal{F}^{-1}\{e^{-Bd\omega}\}, \\ &= I_d\mathcal{F}^{-1}\{Ke^{-Bd\omega}\}, \end{aligned} \tag{2.3}$$

where B and G are medium parameters. Because η and G are similar in practical issues, it is employed a simplified model with one constant K [Trucco and Olmos-Antillon, 2006]. $\mathcal{F}^{-1}$ indicates inverse Fourier transform, and ω is the radial frequency. The damping factor B is an empirical parameter that is tied to the inherent scattering properties of the medium. Typically, it is similar to the total scattering coefficient β [Negahdaripour et al., 2002]. The forward scattering is commonly ignored in outdoor scenes [Nayar and Narasimhan, 1999; Narasimhan and Nayar, 2002] since the solid angle subtended by the source and camera with respect to each other are small [Middleton, 1952].

The backscattering[1]effect results from the interaction between the ambient illumination sources and particles dispersed in the medium. A simplified model is defined by Eq. 2.4.

$$I_{bs} = A(1 - e^{-\eta d}) = A(1 - t), \tag{2.4}$$

where A is the global light in the scene or the airlight at infinity, *i.e.* at a non-occluded horizon. Other nomenclature are adopted in the literature as A_∞, B_∞ or I_∞. This term is a function that may be assumed as a scalar that depends on the wavelength and the water depth, in the case of underwater. The medium and the illumination conditions also affect this term. The scene radiance J also can be decomposed as a

---

[1] The scattered light may be termed as: backscattering, airlight, spacelight, atmospheric veil or veiling light, but in general, the term is path radiance [Treibitz and Schechner, 2009b]. In the present paper, we adopted the term backscattering.

function of the global light A and the scene albedo ρ, assuming Lambertian surfaces [Oakley and Satherley, 1998; Tan, 2008; Roser et al., 2014].

The constant A can be estimated in different ways: by using a calibration procedure [Queiroz-Neto et al., 2004; Gibson et al., 2012], finding the farthest pixel in the scene [Nascimento et al., 2009], the maximum intensity in white balanced images [Tarel and Hautière, 2009; Zhang et al., 2010], finding the brightest pixel in the image [Narasimhan and Nayar, 2002; Tan, 2008; Fattal, 2008], averaging the 1% brightest pixels [Carr and Hartley, 2009] or estimating the brightest pixel in the dark channel [He et al., 2009, 2011]. For the non-uniform illumination case, the parameter A may be modeled as a function of each pixel, $A(x, y)$, and estimated using a simple calibration process [Treibitz and Schechner, 2006; Sarafraz et al., 2010].

According to Schechner and Karpel [2005], backscattering is the prime reason for image contrast degradation, and for this reason the forward scattering can be usually neglected, even though in underwater scenes. However, the forward scattering is taken into account in some works [Negahdaripour et al., 2002; Trucco and Olmos-Antillon, 2006; Barros, 2010; He et al., 2013], mainly in the underwater context.

Therefore, a simplified model that describes the image formation for participating media can be stated as:

$$I(\boldsymbol{x}) = J(\boldsymbol{x})t(\boldsymbol{x}) + A(1 - t(\boldsymbol{x})), \tag{2.5}$$

where $\boldsymbol{x}$ is the pixel coordinate and $I(\boldsymbol{x})$ is the image obtained in participating media. This simplified version is similar to the model proposed by Koschmieder [Koschmieder, 1924], the most adopted model for foggy images. This model is valid for both scattering and non-participating media. In the case of non-participating medium, the attenuation coefficient is approximately zero, thus the medium transmission is equal to one.

After the estimation of the parameters $t(\boldsymbol{x})$ and A, the radiance can be recovered using Eq. 2.6, based on Eq. 2.5.

$$J(\boldsymbol{x}) = \frac{I(\boldsymbol{x}) - A}{e^{-\eta d(\boldsymbol{x})}} + A = \frac{I(\boldsymbol{x}) - A}{t(\boldsymbol{x})} + A. \tag{2.6}$$

This inverse model allows us to estimate the restored image based on the degraded image, $I(\boldsymbol{x})$, the depth map, $d(\boldsymbol{x})$, and the medium parameters, A and η. Thus, the problem of image restoration may be seen as the problem of estimating the medium parameters and the depth map.

2.2 Restoration Using Additional Information

In the literature of image restoration, several works use additional information. Multispectral images are adopted in some works by using infrared cameras [Schaul et al., 2009; Zhang and Wang, 2011; Feng et al., 2013; Bieganski and Kasinski, 2014] or multispectral cameras [Hadjimitsis et al., 2004; Mortazavi et al., 2012, 2013; Makarau et al., 2014]. Range information from sonar sensors [Schilling et al., 2010; Kaeli et al., 2011] are also applied. Structure light [Negahdaripour et al., 2002; Levoy et al., 2004; Narasimhan et al., 2005; Narasimhan and Nayar, 2005; Fuchs et al., 2008; Gupta et al., 2008] and other illumination-based methods [Nayar et al., 2006; Vasilescu et al., 2010, 2011; Treibitz and Schechner, 2012] are proposed to solve the problem. Some difficulties like calibration procedures, complexity and expensive cost limit their applicability.

Other important additional information may be extracted by using polarizer filter [Schechner et al., 2001, 2003; Schechner and Karpel, 2004; Shwartz et al., 2006; Treibitz and Schechner, 2006; Namer et al., 2009; Treibitz and Schechner, 2009a,b; Sarafraz et al., 2009; Liang et al., 2014] that allows to improve the image quality using post-processing procedures. The use of polarizers is still cumbersome, even though images acquired with them present very good results. The main drawback of this technique is the need to automate the identification of maximum and minimum polarization states.

Finally, some works use explicitly information about the scene like the depth [Tan and Oakley, 2000, 2001; Oakley and Satherley, 1998; Narasimhan and Nayar, 2003b; Kopf et al., 2008; Hautière et al., 2007, 2010; Tarel et al., 2012], or the medium [Grewe and Brooks, 1998; Nayar and Narasimhan, 1999; Narasimhan and Nayar, 2000, 2003a; Ahlen, 2003]. This prior information also limits the practical applicability of these methods.

2.3 Restoration Using Only Visual Images

The problem of image restoration may be tackled using single image or multiple images, *e.g.* stereo or temporal sequence of images. They are described in Sec. 2.3.1 and Sec. 2.3.2, respectively.

2.3.1 Single Image

Most of the recent works are based on single visual image. The idea behind almost all methods is to make assumptions about the medium or the scene. We focused in image restoration methods that are able to deal with space-variant images, *i.e.*


(a) (b)

Figure 2.3. Results obtained by Groszek and Allebach [2005]. (a) Image taken under hazy, and (b) restored image with small improvement in terms of visibility.

the distance between camera and scene varies over the entire image. Typically, the methods are based on the optical model, but there are some recent methods that presents impressive results without considering the model explicitly. Thus, this section is divided into model-based and model-free methods.

2.3.1.1 Model-Free Methods

Typically, this kind of approach is based on image filtering. There are a vast literature on contrast enhancement, an important issue in images acquired in participating media, that includes histogram equalization, contrast stretching, linear mapping, gamma correction, and its variations [Gonzalez and Woods, 2007]. However, they are out of the scope of this dissertation because of their limited capabilities to deal with space-variant images. We highlight only the recent methods that present competitive results.

Groszek and Allebach [2005] applied contrast enhancement on non-sky areas of haze image. The method detects the horizon line using luminance ratio and edge detector. Qualitative results are limited presenting a small improvement in terms of visibility, shown in Fig. 2.3.

The work of Bazeille et al. [2006] proposes an automatic image restoration for underwater images. The method is composed of several successive independent steps which the objective is to enhance the contrast, adjust the colors, suppress noise and correct non-uniform illumination. Classical image processing approach are applied like contrast stretching and some filtering techniques as anti-aliasing, homomorphic filter and anisotropic filter. Qualitative results show improvements, but present difficulties to deal with the backscattering. Furthermore, gradient histogram and robustness criterion [Arnold-Bos et al., 2005] results presented improvement in contrast. Following the same idea, Iqbal et al. [2010] propose successive steps in order to adjust the colors in underwater images. Three steps are proposed: white-balance on RGB color channel

(a) (b) (c) (d) (e) (f)

Figure 2.4. Image restoration results obtained by Ancuti et al. [2010] under hazy condition. (a) Original image, (b-e) restored using the method of Ancuti et al. [2010], Fattal [2008], He et al. [2009], Tarel and Hautière [2009], and Tan [2008], respectively.

and contrast stretching on both RGB and HSI color channels. They obtain visual images, histograms and edge images. However, they are limited to near objects with approximately constant distance to the camera.

Ancuti et al. [2010] proposed a fusion strategy that takes two filtered inputs from the original image: white balanced image and the difference image between original image and luminance average. These inputs are weighted by three normalized weight maps: luminance, chromatic and saliency. In that work, an Image Quality Assessment (IQA) metric [Aydin et al., 2008] has been used for evaluating the results, where the method improves the contrast. Furthermore, visual images are obtained and compared with other single image methods, as shown in Fig. 2.4. The results are visually similar. The evaluation technique proposed by Hautière et al. [2008b] is applied to the obtained results. This work is better detailed and more results are provided in [Ancuti and Ancuti, 2013].

The work of Ancuti et al. [2010] was also extended for underwater images [Ancuti et al., 2011, 2012]. The inputs and weights in the fusion strategy are adapted to the medium. Qualitative results are provided and compared with [Tarel and Hautière, 2009], [Ancuti et al., 2010] and [Schechner and Karpel, 2005]. The results present a good increasing on visibility and contrast. Quantitative results are provided using IQA metric [Aydin et al., 2008] and compared with [Bazeille et al., 2006; Tarel and Hautière, 2009] (Fig. 2.5). Finally, results for state-of-art feature detector and segmentation are provided, showing improvements after the restoration.

2.3.1.2 Model-Based Methods

The work of Trucco and Olmos-Antillon [2006] proposes an underwater image restoration method based on the recovery of the forward scattering parameters. The idea is to obtain the forward scattering component I_{fs}, defined in Eq. 2.3. The inverse filter

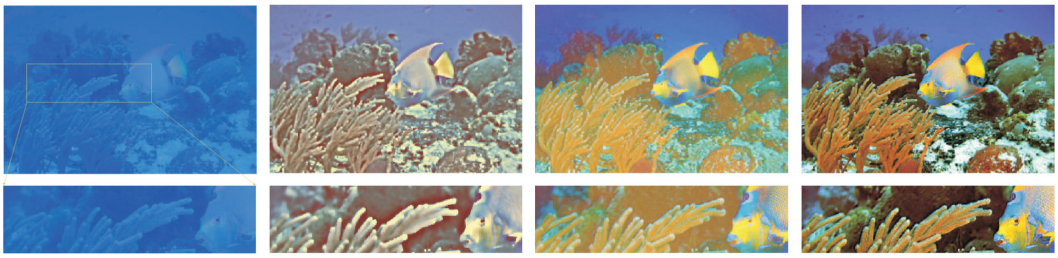


Figure 2.5. Comparative results obtained by Ancuti et al. [2012] in a underwater scene, showing the original image in the left side. Following, the restored images using the method of Bazeille et al. [2006], Tarel and Hautière [2009], and Ancuti et al. [2012], respectively.

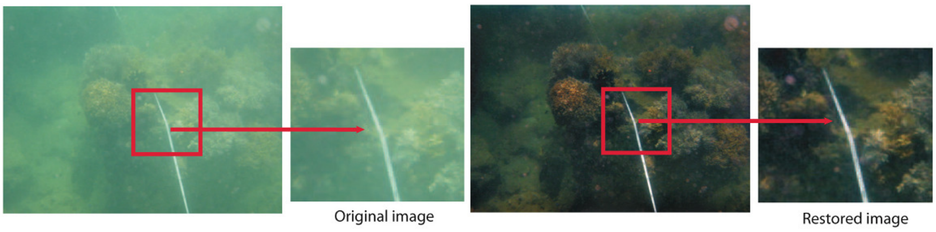


Figure 2.6. Underwater image restoration result obtained by Trucco and Olmos-Antillon [2006].

is initially estimated with typical values in the frequency domain, and optimized using simplex search. Qualitative results are provided (Fig. 2.6), as well as quantitative results in a classification task. The method improves the blur and color fidelity, but does not deal with the backscattering.

Following a similar idea, Barros [2010] proposed a method that deals with forward-scattering and backscattering effects. The restoration is achieved using a broad range of measures such as contrast, blur, edges integrity and noise. The parameters are recovered using a multiobjective goal attainment solver. The qualitative results obtained in simulated and real underwater scenes are expressive. Moreover, a feature detector is employed in order to qualitatively evaluate the method. The method is compared with [Bazeille et al., 2006] and chromatic expansion, with good improvement in terms of color fidelity and number of features. Fig. 2.7 shows an example of image restoration using the method proposed by Barros [2010].

All of the next described works are based on the simplified optical model shown in Eq. 2.5. The work proposed by Oakley and Bu [2007] deals with the contrast loss because of the backscattering effect. The idea is estimating the backscattering component. The main assumption is that the local sample mean of pixel intensities is proportional to the standard deviation. Furthermore, the method considers that

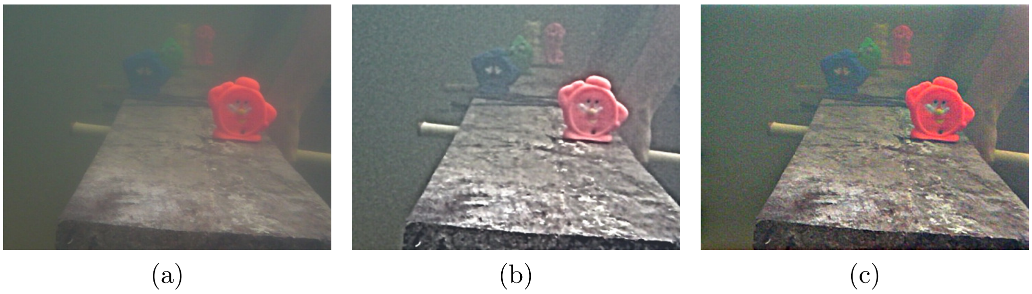

(a) (b) (c)

Figure 2.7. Underwater image restoration results obtained by Barros [2010]. (a) Original image, (b) restored using the method of Bazeille et al. [2006], and (c) Barros [2010] .

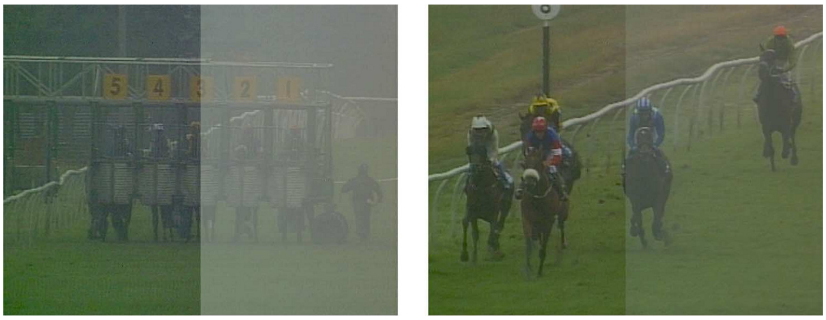


Figure 2.8. Example images obtained by Oakley and Bu [2007], where the processed image is presented in left side, while the unprocessed image is right side.

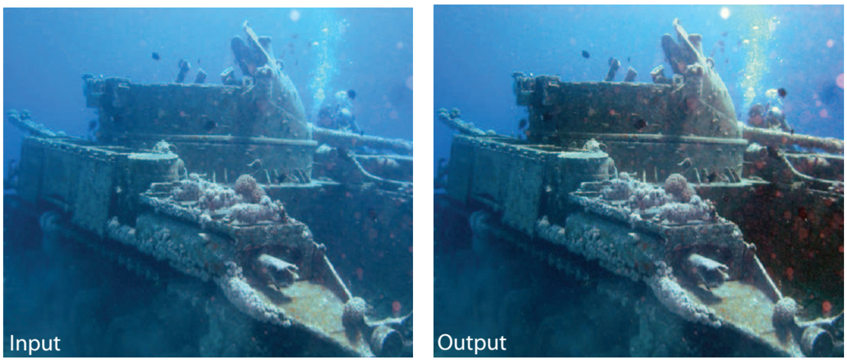
the backscattering is a constant value which is only valid on a limited number of real scenes. Qualitative results are provided for a pair of images taken over the rain with good results, shown in Fig. 2.8. A real-time implementation of the method is proposed in Oakley and Marceta [2011].

In Tan [2008], an optimization formulation based on Markov Random Fields (MRF) [Li, 2009] is applied to maximize the contrast while assuming a smooth layer of backscattering. Qualitative results are obtained with enhanced scene contrasts, but producing some halos near depth discontinuities, as shown in Fig. 2.9.

Fattal [2008] developed a single image method that uses a local window-based operation and a probabilistic graphical model. The proposed model is similar to the adopted here, *i.e* Eq. 2.5, but the scene radiance J is decomposed as a pixelwise product of surface albedo coefficient and the shading factor. The main assumption is the statistical independence between the shading factor that depends on the illumination of the scene, surface reflectance properties, and scene geometry; and the medium transmission that depends on the scene depth and the scattering coefficient β. A Markov Random Field (MRF) [Li, 2009] deals with noise and smooths the transmis-

(a) (b)

Figure 2.9. Image restoration results obtained by Tan [2008]. (a) Original image, (b) Restored image.



(a) (b)

Figure 2.10. Image restoration results obtained by Fattal [2008] in a underwater scene. (a) Original image and (b) the restored image.

sion. Qualitative results are provided in several real scenes that includes hazy, aerial and underwater, as shown in figs. 2.10 and 2.11. Furthermore, comparative results are provided for Schechner et al. [2001], Chavez-Jr [1996] and others imaging processing methods like histogram equalization and unsharp mask, obtaining better results.

A fast restoration method was proposed by Tarel and Hautière [2009]. The method attempts to estimate the backscattering component, I_{bs}, based on the minimum filter over the white balanced image W. Eq. 2.7 defines the estimation of the I_{bs}:

$$I_{bs} = \max(0, \min(p(\mathrm{med}_s\, W - \mathrm{med}_s(|W - \mathrm{med}_s)), W)), \tag{2.7}$$

where p controls the strength of the restoration, and med_s is a special median filter proposed by the authors using a window $s \times s$ that preserves edges and corners. It is assumed that backscattering is proportional to the difference between local average and local standard deviation on W. Qualitative results are provided in haze scenes


(a) (b) (c)

Figure 2.11. Image restoration results in hazy scenes. (a) Original image, (b,c) Restored image obtained by Fattal [2008] and Tarel and Hautière [2009], respectively.

(Fig. 2.11) and compared with single image methods. Quantitative results using the metrics of Hautière et al. [2008b] and a lane-marking extraction application are also provided.

The method of Tarel and Hautière [2009] was also applied to underwater images by Wang and Wu [2010], but presenting color distortion and small improvement in terms of visibility. Lu et al. [2013] and Serikawa and Lu [2014] applied bilateral filtering and automatic color equalization to improve the results on underwater images. Interestingly, they uses four different image quality assessment to evaluate the obtained results: HDR-VDP2 [Mantiuk et al., 2011], JPEG-QI [Wang et al., 2002], SSIM [Wang et al., 2004], Q-MOS and CPU-time. Despite the best results in terms of the metrics, the method presents limited results in terms of visibility and color restoration.

Differently, Carr and Hartley [2009] argue that the distance along the ground to the visible scene point is a monotonically increasing function of the image plane height. The MRF [Li, 2009] is proposed as a regularization tool for improving the transmission estimation. Despite the rough transmission obtained, they obtain impressive qualitative results (Fig. 2.12).

One of the most successful method was proposed by He et al. [2009]. It assumed that every local patch in a image should have at least one color channel component near zero. This fact enables the medium transmission estimation. The method is called Dark Channel Prior (DCP). Eq. 2.8 shows the model to estimate the medium transmission:

$$t = 1 - \omega \min_{\Omega} \left(\min_{cl \in RGB} \frac{I^{cl}}{A^{cl}} \right), \tag{2.8}$$

where I^{cl} and A^{cl} are the image and the global light in the color channel cl. Ω is a local


(a) (b) (c)

Figure 2.12. Image restoration results in a hazy scene obtained by Carr and Hartley [2009]. A foggy image (a) is restored the proposed method using monotonic model (b), and dark channel model [He et al., 2009] (c).

patch and ω is a constant that preserves a small layer of haze in distant objects. This assumption is corroborated by statistical analysis of outdoor natural images acquired in clear day, and inspired by the dark-object subtraction [Chavez-Jr, 1996]. Because of the similarities between image matting equation and the optical model, the authors apply the alpha matting algorithm [Levin et al., 2008] to refine the transmission based on the smoothness of neighboring pixels. Impressive qualitative results are obtained in outdoor scenes (Fig. 2.13) and compared with [Tan, 2008; Fattal, 2008; Kopf et al., 2008]. The main drawbacks of the method are related the depth of objects which are naturally white or light grey, and the computational cost.

Several improvements and variations of the DCP are proposed in the literature. Because of the high computational cost of the method of the soft matting, bilateral filter [Fang et al., 2010], Cross-bilateral filter [Lv et al., 2010], trilateral filtering [Serikawa and Lu, 2014], guided filter [He et al., 2010], adaptive wiener filter [Gibson and Nguyen, 2013], anisotropic diffusion [Tripathi and Mukhopadhyay, 2012], quadtree-based soft matting [Ding and Tong, 2012], multi-scale estimation [Wang et al., 2013], and local patch variation mechanism [Chen et al., 2014] were applied to deal with this limitation of the DCP.

Inspired by DCP, Gibson et al. [2012] proposed the Median Dark Channel Prior (MDCP) by replacing the first minimum in Eq. 2.8 by the median operator. The median operation preserves edge, avoiding the refinement of the medium transmission.

DCP is also applied on underwater images. Chao and Wang [2010] applied the original method without adaptations with limited results on images acquired in a pool. Chiang and Chen [2012] also applied original DCP, however, compensating non-uniform illumination and absorption rates at different wavelength. Fig. 2.14 shows results obtained by both methods.

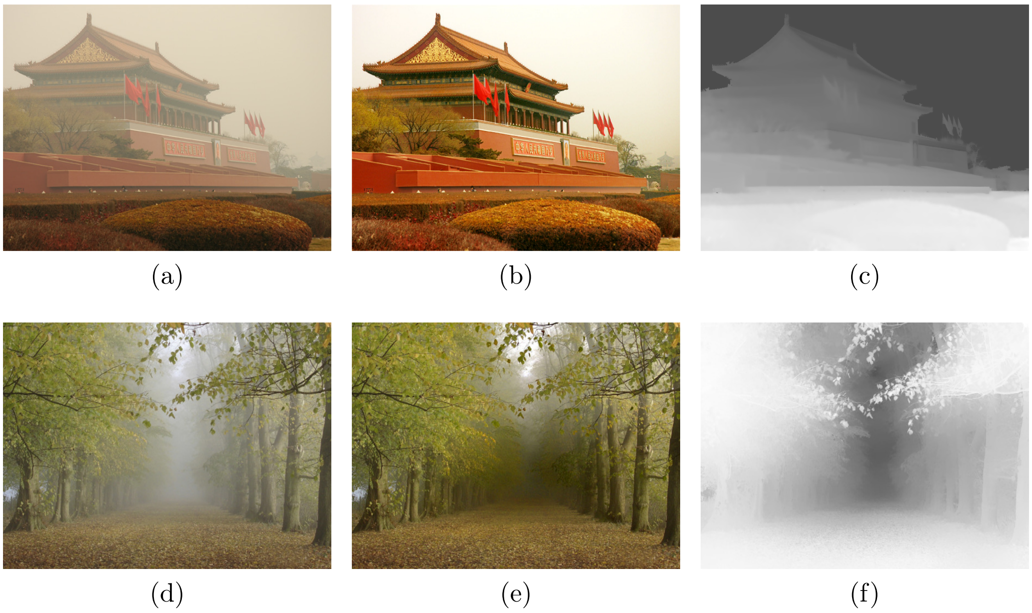


Figure 2.13. Image restoration results in hazy scenes obtained by He et al. [2009]. (a,d) Original images, (b,e) restored images, and (c,f) recovered transmissions.

Figure 2.14. Underwater image restoration results obtained by Chao and Wang [2010] and Chiang and Chen [2012]. (a,c) Original images, (b,d) restored image by Chao and Wang [2010], and Chiang and Chen [2012], respectively.

Carlevaris-Bianco et al. [2010] proposed a new interpretation of the DCP for underwater conditions. Their prior explores the fact that the attention of light in water varies depending of the color of the light, *i.e.*, the wavelength. In underwater medium, the red color channel is attenuated at a much higher rate than the green or blue. The prior is given by Eq. 2.9.

$$\begin{aligned} D(x) &= \max_{x\in\Omega, cl\in R} I^{cl}(x) - \max_{x\in\Omega, cl\in B,G} I^{cl}(x), \\ \tilde{t}(x) &= D(x) + (1 - \max_{x} D(x)). \end{aligned} \qquad (2.9)$$

Differently from the standard DCP, this prior is based on the difference between

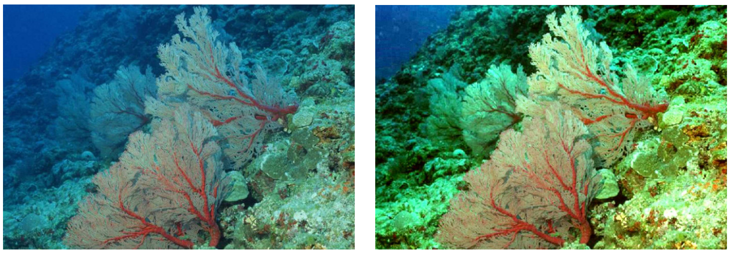

(a) (b)

Figure 2.15. Underwater image restoration results by Carlevaris-Bianco et al. [2010]. (a) Original image, and (b) restored image.

the maximum in the red channel and the other channels, instead of the minimum as in DCP. Qualitative results in underwater scene present a good improvement in terms of color fidelity, but limited in terms of the distance which one can reliably discriminate an object in the scene. Fig. 2.15 shows an example of obtained results.

Codevilla et al. [2014] proposed other adaptation based on the color of the backscattering, *i.e.* airlight in the infinity A. Eq. 2.10 shows the proposed prior, which the inverse of the intensity of the red channel is adopted due to the high absorption rates of this channel, where I^R, I^G and I^B are the intensity of the images in each RGB color channel, respectively. Only a single restored image is provided with promising results in terms of visibility.

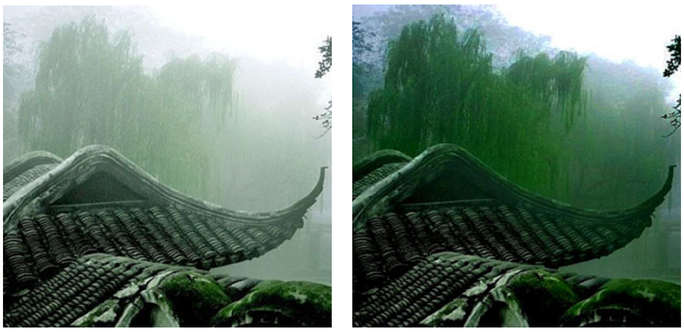
$$\tilde{t} = \begin{cases} \text{If Blue} \quad 1 - \omega \min_{\Omega}\left(\frac{\min(I^B, 1-I^R)}{A}\right), \\ \text{If Green} \quad 1 - \omega \min_{\Omega}\left(\frac{\min(I^G, 1-I^R)}{A}\right), \\ \text{If Ciano} \quad 1 - \omega \min_{\Omega}\left(\frac{\min(I^B, I^G, 1-I^R)}{A}\right). \end{cases} \tag{2.10}$$

Zhang et al. [2010] proposed a restoration method based on rearrangement of the model defined in Eq. 2.5, shown in Eq. 2.11:

$$B = 1 - \frac{I}{A} = t(1-\rho), \tag{2.11}$$

where B is the blurred version of the degraded image, and ρ is the scene albedo. The authors argue that large-scale variations in degraded image are caused by the transmission, t, and small-scale intensity variations are caused by the inverse-albedo $(1-\rho)$. Qualitative results are provided in real and simulated scenes and compared with single image methods, obtaining similar results. Fig. 2.16 shows an example of results obtained by Zhang et al. [2010].

A probabilistic formulation to deal with the inherent ambiguity between scene



(a) (b)

Figure 2.16. Image restoration results by Zhang et al. [2010] in a hazy scene. (a) Original image, and (b) restored image.

(a) (b) (c)

Figure 2.17. Example of image restoration results by Kratz and Nishino [2009] in a hazy scene. (a) Original image, and (b,c) restored image obtained by He et al. [2009] and Kratz and Nishino [2009], respectively. The method of Kratz and Nishino [2009] recovers more detail around dense fog areas such as the rock faces.

albedo and depth was proposed by Kratz and Nishino [2009]. They use MRF [Li, 2009] to factorize the scene albedo and the depth log transform. Moreover, they impose statistical priors on the transformed variables. Statistics priors such as a heavy tail prior on gradients of the scene albedo and piecewise constant prior on the depth were adopted. Qualitative results are provided (Fig. 2.17) and compared with [Schechner et al., 2001; Fattal, 2008; Tan, 2008; He et al., 2009], showing similar results. Nishino et al. [2012] provide details about the method and further results.

Following the idea of modeling the restoration problem as a MRF, Caraffa and Tarel [2013] proposed a decoupled estimation in two steps. Firstly, the global light is estimated considering the knowledge about the scene albedo. Further, the scene albedo is optimized with previous estimation. Quantitative results are presented in a simulated dataset [Tarel et al., 2012]. Comparative evaluation with [He et al., 2009; Tarel and Hautière, 2009; Hautière et al., 2010] presents improvement in terms of mean error. Qualitative results are provided in road scenes presenting improvement in terms

(a) (b)

Figure 2.18. Image restoration results by Caraffa and Tarel [2013] in a hazy scene. (a) Original image, and (b) restored image.

of visibility.

Kim et al. [2011] proposed an interesting approach based on contrast optimization. The method estimates the medium transmission by maximizing two criteria: standard deviation of the luminance component and uniformness of the image histogram. Initially, the best global transmission map is estimated, *i.e.* a single value for the entire image. Furthermore, a local transmission map in a limited window is also estimated. The result is refined in order to avoid halos and preserve edges. Qualitative results are obtained and compared with [Oakley and Bu, 2007; Tan, 2008; He et al., 2009]. The results are similar to those presented by [He et al., 2009], but obtained in a faster way.

Meng et al. [2013] proposed an efficient regularization method exploring the inherent boundary constraint on the medium transmission. This constraint is combined with a L1-norm into an optimization framework to recover the unknown transmission. The lower bound transmission t_b is given by:

$$t_b = \min\left[\max_{cl\in RGB}\left(\frac{A^{cl}-I^{cl}}{A^{cl}-C_0^{cl}}, \frac{A^{cl}-I^{cl}}{A^{cl}-C_1^{cl}}\right), 1\right], \tag{2.12}$$

where C_0^{cl} and C_1^{cl} are two constant vectors that limits the scene radiance in the lower and upper bound, respectively. This equation allows to obtain an estimation of the transmission using $t(x) = \min_{y\in\Omega_x}[\max_{z\in\Omega_y} t_b(z)]$, where Ω_x is a local patch. The transmission estimation is regularized by weighting the neighboring pixel. Good qualitative results are provided and compared with other single image methods. Fig. 2.19 shows some examples of obtained results.


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

Figure 2.19. Image restoration results by Meng et al. [2013] in two participating media. (a-c) Original images, and (b-d) restored images.

2.3.2 Multiple Images

Several methods are proposed to restore image acquired in participating media using more than one image. We divided then in stereo images, *i.e.* acquired by a calibrated stereo pair of camera, and temporal sequence of images.

2.3.2.1 Stereo Sequences

The work of Queiroz-Neto et al. [2004] proposed a methodology to estimate disparity map using stereo images acquired in underwater medium. The model described in Eq. 2.5 is adopted, where two different attenuation coefficient are assumed, one for direct component I_d and another for the backscattering component I_{bs}. Initially, the two attenuation coefficient and the global light A are estimated by an offline procedure. Considering calibrated stereo, the method computed the depth for each point and the restored image. Finally, the method finds the best disparity map based on the brightness difference. Qualitative results were obtained in an indoor fish tank with a high turbidity level (Fig. 2.20).

Nascimento et al. [2009] extended the idea of Queiroz-Neto et al. [2004], but focused in the image restoration by avoiding the offline calibration of the medium parameters. The attenuation coefficient η is automatically estimated. The parameter's estimation is based on the optical model (Eq. 2.5). Therefore, the attenuation parameter is given by:

$$\eta = \frac{\ln(I^l - A) - \ln(I^r - A)}{d^l - d^r}, \tag{2.13}$$

where I^r, I^l, d^r and d^l are the right and left stereo images and their depths to the scene, respectively. Differently from Queiroz-Neto et al. [2004], stereo correspondence is previously estimated using classical methods and refined after restoring both images. Simulated and real results are provided (Fig. 2.21). Moreover, the contrast metric

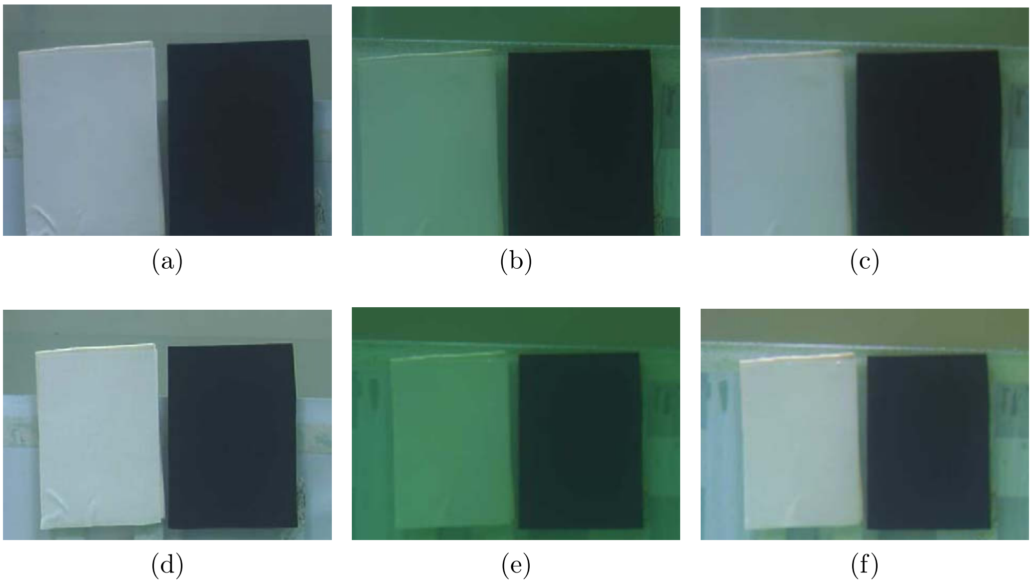


Figure 2.20. Image restoration results by Queiroz-Neto et al. [2004] in a controlled underwater scene. (a,d) Ground truth image, *i.e.* out of water, and (b,e) degraded image of a target under 0.5 m and 1.0 m of water, respectively. Finally, (c,f) restored images.

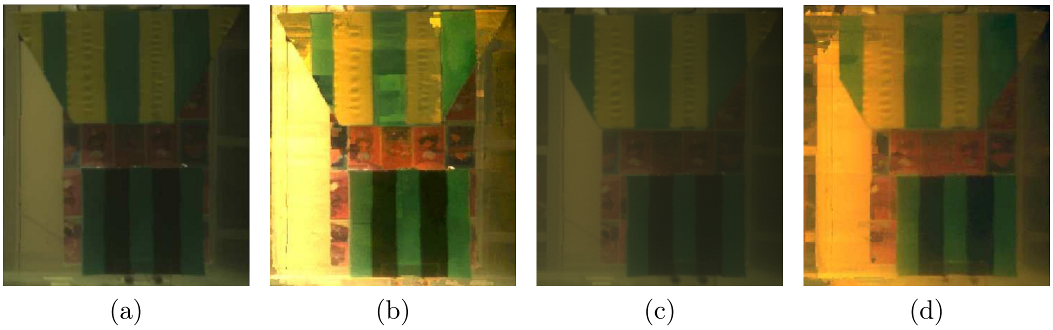


Figure 2.21. Image restoration results by Nascimento et al. [2009] in a controlled underwater scene using dissolved tea. (a,c) Degraded image under two different turbidity levels, and (b,d) restored images.

[Schechner and Karpel, 2005] is evaluated and the method is compared with histogram equalization.

Caraffa and Tarel [2012] proposed a MRF model [Li, 2009] to solve the stereo reconstruction and image restoration in foggy images. The method alternates between image restoration and depth estimation that allows to improve the disparity estimation in untextured areas at short distances. Qualitative simulated results are provided and evaluated in terms of disparity. Restoration results are compared with [Tarel and Hautière, 2009] showing difficulty to deal with discontinuities and long distances (Fig.


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

Figure 2.22. Example of image restoration results by Caraffa and Tarel [2012] in foggy scenes. (a,c) Original images, and (b,d) restored images. The results obtained present good improvement in near regions of the images, but some halos and color distortion are visible in farther regions.

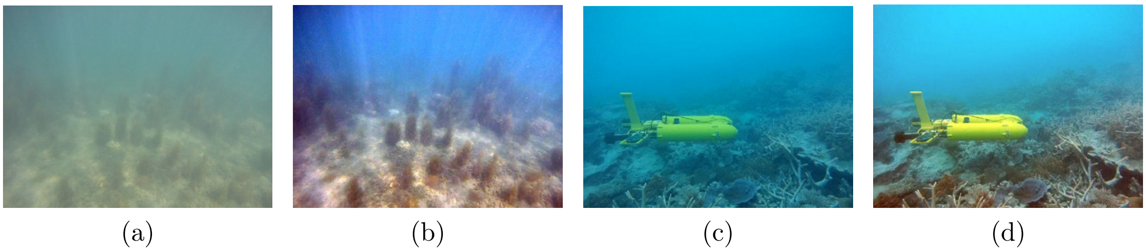

(a) (b) (c) (d)

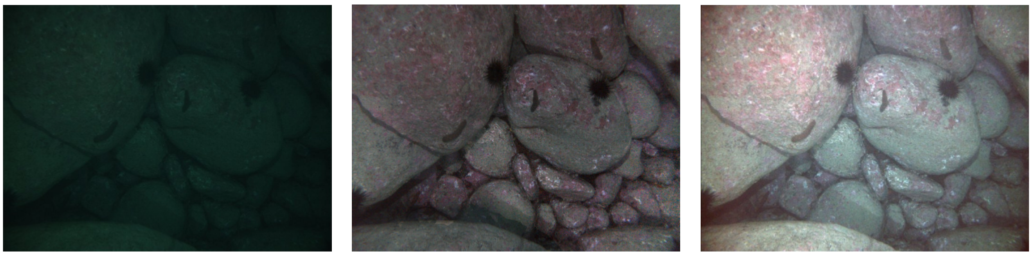
Figure 2.23. Image restoration results obtained by Roser et al. [2014] in underwater scenes. (a,c) Original images, and (b,d) restored images.

2.22).

Recently, Roser et al. [2014] proposed a method for simultaneous image restoration and disparity estimation. The method is divided in two parallel steps that estimate the visibility coefficient, *i.e.* attenuation coefficient and global light, and the stereo matching. Initially, the disparity is coarsely estimated using degraded stereo images, which allows to estimate the parameters and enhancing the images. Secondly, the disparity map and the parameters are refined. The accuracy of the method is evaluated using the images from Middlebury stereo dataset [Scharstein and Szeliski, 2002], after applying an artificial degradation. The method shows good improvement in the disparity estimation. Qualitative results are provided in real underwater scenes acquired using a robot (Fig. 2.23), where color fidelity and visibility are improved.

2.3.2.2 Sequence of Images

The color correction is addressed in Bryson et al. [2012] using a sequence of stereo images. It allows to build a feature-based structure-from-motion model [Mahon et al., 2008]. The adopted model shown in Eq. 2.14 is similar to presented in Eq. 2.5:



(a) (b) (c)

Figure 2.24. Image restoration results obtained by Bryson et al. [2012] in a underwater scene. (a) Original image, and (b,c) restored images using the method of Bryson et al. [2012] and white balance, respectively.

$$I_y(\boldsymbol{x}, \lambda) = m(\boldsymbol{x}, \lambda, d) I_x(\boldsymbol{x}, \lambda) + n(\boldsymbol{x}, \lambda, d), \tag{2.14}$$

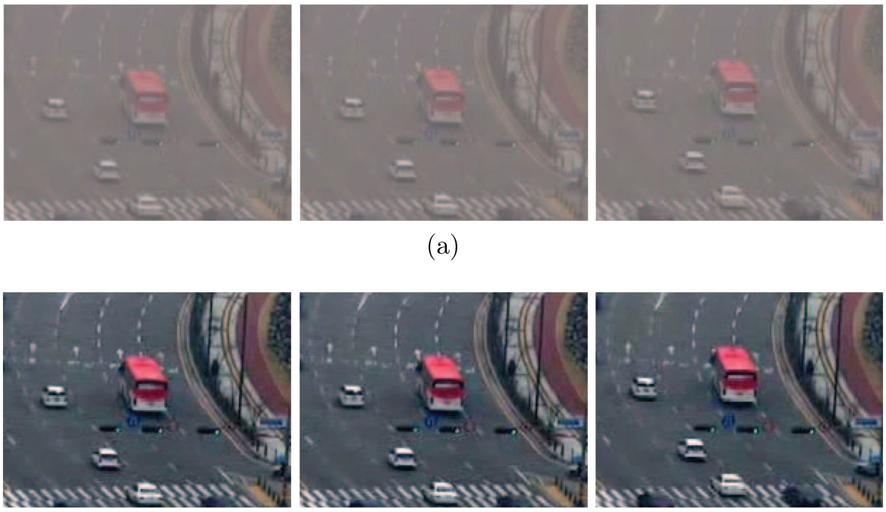
where $I_x(\boldsymbol{x}, \lambda)$ is the intensity of a given pixel $\boldsymbol{x}$ for the wavelength λ in the image I_x. The authors argue this model is able to deal with a variety of effects like artificial illumination, vignetting and range-dependent attenuation in different wavelength. Based on histogram statistics and least square optimization over all sequences, the method compensates these effects. Qualitative results show significant improvement in terms of color consistence, as shown in Fig. 2.24.

A haze removal method based on a sequence of images was proposed by Joshi and Cohen [2010]. The method deals with a set of effects as blur, sampling, alignment error, sensor dust and haze. The work focuses on sequence of images obtained by static camera and scene. Initially, the method performs a global alignment of each image to a reference frame. Based on pixelwise optical flow, a single image is obtained by weighting the sequence. A dehazing method based on DCP [He et al., 2009] is applied which assumes the global light is constant per scan-line. Qualitative results obtained in a sequence of a hazy mountain scene (Fig. 2.25) show improvement in terms of noise and visibility.

Zhang and Wang [2011] proposed a video dehazing method based on DCP [He et al., 2009] and optical flow. The problem is modeled as a MRF [Li, 2009] to smooth and to improve temporal coherence. The method assumes brightness constancy in the classical optical flow estimation is still valid. Qualitative results on real and simulated hazy images are provided with limited improvement in terms of visibility and with some halos in the restored image.

The method of Kim et al. [2011], detailed in Sec. 2.3.1.2, was extended for image sequences [Kim et al., 2012]. The temporal coherence is take into account to estimate

Figure 2.25. Image restoration result obtained by Joshi and Cohen [2010] in a haze scene.



(a)

(b)

Figure 2.26. Restoration results obtained by Kim et al. [2012] in sequence of images acquired from an air polluted scene. (a) Degraded frames, and (b) restored frames.

the medium transmission, beyond the contrast and the histogram uniformness. The temporal coherence is estimated from the difference between transmission maps on consecutive frames. Qualitative results are obtained in two image sequences on road scenes are compared with [Tarel and Hautière, 2009; Kim et al., 2011]. Although the limited results obtained using [Tarel and Hautière, 2009], the improvement compared to [Kim et al., 2011] is imperceptible (Fig. 2.26). Kim et al. [2013] proposed small improvements in the method and provided further results.

2.3.3 Qualitative Evaluation

The most relevant works using only visual images are summarized using a qualitative comparison. Five different properties were concerned, that includes: visibility (**Vis.**),

Table 2.1. Qualitative Comparison from restoration methods using only visual images.

Reference	Vis.	Col.	Perf.	App.	Soft.	Main Assumption	P. Med.	Eval.
[Groszek and Allebach, 2005]	-	-	++	+-	+	Contrast Enhanc.	Haze	Qual.
[Bazeille et al., 2006]	+	+	++	+	++	Seq. Filter	Water	Quan.
[Iqbal et al., 2010]	+	++	++	+	++	Contrast Enhanc.	Water	Qual.
[Ancuti et al., 2010]	++	++	++	++	+	Fusion Filter	Haze	Quan.
[Ancuti et al., 2011, 2012]	++	++	++	++	+	Fusion Filter	Water	Quan.
[Trucco and Olmos-Antillon, 2006]	+-	+	+	+-	+	Forw. Scat. Parameters	Water	Quan.
[Barros, 2010]	++	+	-	+-	+-	Multiobj. Optim.	Water	Quan.
[Oakley and Bu, 2007]	++	+	++	++	+	Local Mean prop. Std.	Haze	Quan.
[Tan, 2008]	+	+-	+	++	+	Maximizing Contrast	Haze	Qual.
[Fattal, 2008]	++	++	+	++	+-	Uncorrelated fields	Haze	Qual.
[Tarel and Hautière, 2009]	++	+	+++	++	+	Airlight prop. Local Patch	Haze	Quan.
[Lu et al., 2013]	++	+-	++	++	+	[Tarel and Hautière, 2009]	Water	Quan.
[Carr and Hartley, 2009]	++	++	++	++	+	Monotonic Distance	Haze	Qual.
[He et al., 2009, 2011]	+++	++	+	+++	++	Dark Channel Prior	Haze	Qual.
[Gibson et al., 2012]	++	++	++	+++	++	Median DCP	Haze	Qual.
[Chiang and Chen, 2012]	+	+	+	++	++	[He et al., 2009]	Water	Qual.
[Carlevaris-Bianco et al., 2010]	+	+	++	++	++	Diff. Max. Red. and Other Chan.	Water	Qual.
[Codevilla et al., 2014]	++	++	+	++	++	Modified DCP	Water	Qual.
[Zhang et al., 2010]	++	++	+-	++	+	Sep. Transm. and Albedo	Haze	Qual.
[Kratz and Nishino, 2009]	++	++	+-	++	+	Simult. Depth and Restoration	Haze	Qual.
[Caraffa and Tarel, 2013]	++	+	+	++	+	Simult. Depth and Restoration	Haze	Quan.
[Kim et al., 2011]	+++	++	++	++	++	Contrast Maximization	Haze	Qual.
[Meng et al., 2013]	++	++	+	++	+	Boundary Constraint	Haze	Qual.
[Queiroz-Neto et al., 2004]	+	+	++	+	++	Simult. Depth and Restoration	Water	Qual.
[Nascimento et al., 2009]	+	+	++	+	++	Automatic Param. Est.	Water	Quan.
[Caraffa and Tarel, 2012, 2014]	+	+-	++	+	++	Simult. Depth and Restoration	Haze	Quan.
[Roser et al., 2014]	++	++	++	+	++	Simult. Depth and Restoration	Water	Qual.
[Bryson et al., 2012]	+-	+++	+	++	+	Statistical Learning	Water	Qual.
[Joshi and Cohen, 2010]	++	++	+	++	++	DCP per line	Haze	Qual.
[Zhang and Wang, 2011]	+	+	++	+	++	DCP + Optical Flow	Haze	Quan.
[Kim et al., 2011, 2013]	++	++	++	++	++	Contrast Max. + Temp. Coherence	Haze	Qual.

color correction (**Col.**), temporal performance (**Perf.**), practical applicability (**App.**) and software simplicity (**Soft.**). Visibility (**Vis.**) is related with the capability of the method to restore the contrast in farthest areas of the image. Color correction (**Col.**) is associated with the ability of the method to improve the colors on the restored image. Temporal performance (**Perf.**) of each method is qualitatively evaluated with the computational cost to restore the image, *e.g* the use of soft matting imposes an extra cost to the method. Practical applicability (**App.**) is a subjective property related with the assumptions and their generality. Finally, software simplicity (**Soft.**) is the subjective property related with the hardiness to implement the method using a high level language like Matlab. For each property evaluated, five different levels are assigned from the best to the worse: "+++","++","+","+-" and"-".

Table 2.1 includes simplified information about the method as the main assumption (**Main Assumption**); the main participating media focused by the method (**P. Med.**); and the evaluation strategy (**Eval.**).

2.4 Discussions

The field of image restoration has been increasing by the impact generated by single image approaches [Tan, 2008; Fattal, 2008; He et al., 2009; Tarel and Hautière, 2009] that present good results using simple and intuitive assumptions.

In terms of applicability, the use of polarizers is a well-known efficient method, but remains unpractical for real world applications [Bonin et al., 2011]. Furthermore, the use of polarizers attenuates the signal, even though in the best polarization state [Treibitz and Schechner, 2009b]. It limits the quality of the restored image. Despite the good results, the methods that rely on specialized hardware are expensive and complex. Moreover, they are hard to implement when automatic acquisition is required.

Single image methods are the most success approach, both in terms of citation and applicability. However, they are limited by the assumptions that can fail in real situations, thus lacking in terms of reliability and robustness. In stereo approaches, solving the correspondence problem is made harder due to the effects imposed by the participating medium.

The present work takes the advantage of single image and multiple images methods to improve the restoration of images acquired in participating media. The present work addresses some issues of the previous multiple image methods to obtain a more reliable method. It includes the temporal coherence using optical flow, the depth map estimation and the estimation of the attenuation coefficient.

Chapter 3

Methodology

The proposed methodology improves the visibility of images acquired in participating media. Assuming the model described in Sec. 2.1, the problem of image restoration may be the problem of estimating the attenuation parameter and the depth map.

Fig. 3.1 illustrates the proposed methodology, where arrows indicate the flow. The orange boxes are the inputs/outputs data, and the blue ellipses are the four main procedures. The contribution of the present work are highlighted by a black contour. In addition to the restored images, our methodology also produces an estimation of the depth maps, which can be used in specific applications such as tridimensional reconstruction.

The four main procedures are:

- Transmission Prior - This procedure provides the initial estimation of medium transmission. It is obtained using a simple but effective image prior called Underwater Dark Channel Prior (UDCP). Most patches in natural images contain pixels that are dark in at least one color channel, in the RGB color channel representation. Using this observation and the optical model, we obtain a raw estimation of the medium transmission;

- Optical Flow - This procedure recovers spatial 2D motion from a sequence of images based on the assumption of brightness constancy in the restored image. We proposed a new formulation that is able to deal with both participating and non-participating media, called Generalized Optical Flow Model (GOFM). The optical flow is crucial for the next procedures;

- Structure from Motion - Extrinsic parameters and the depth maps are estimated in this procedure. Continuous epipolar constraint and bundle adjustment are

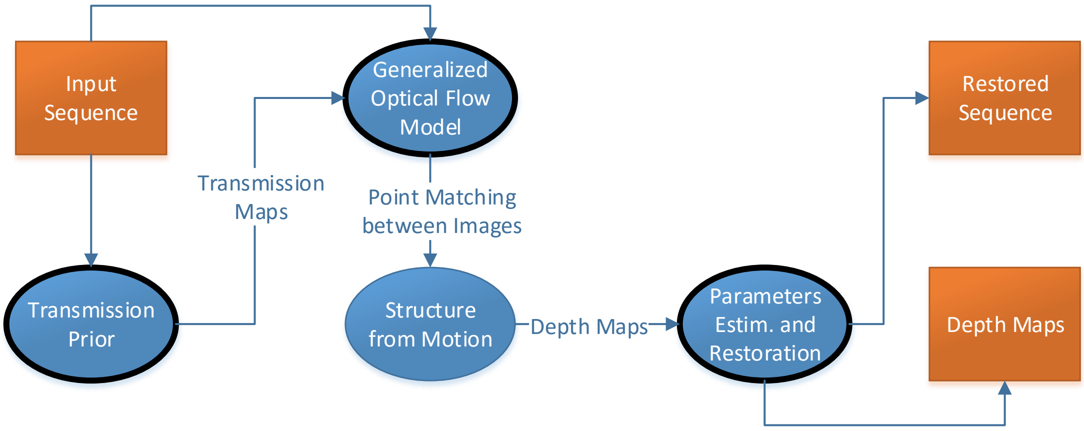


Figure 3.1. Proposed methodology and its main procedures. The orange boxes are the inputs/outputs data, and the blue ellipses are the four main procedures. The contribution of the present work are highlighted by a black contour. We estimate a transmission map based on the input images and priors. This map allows us to compute the optical flow between adjacent frames. Then, depth maps are predicted using structure from motion techniques. Finally, the medium parameters and the restored images are estimated.

applied in order to obtain information about the camera's movement during sequence acquisition. Using triangulation, we estimate three-dimensional points and depth maps;

- Parameter Estimation and Restoration - This procedure estimates the attenuation coefficient based on the optical model that provide a set of constraint. We find the best attenuation coefficient using a stochastic framework based on M-estimator SAmple and Consensus (MSAC). It is also responsible for obtaining the restored image using the inverse model.

3.1 Transmission Prior

Recently, several priors for single images have been proposed, as discussed in Sec. 2.3.1. While these methods have shown good performance for hazy/foggy images, there is still room for improvement when applying to underwater images. One key aspect that needs to be probed further is the transmission estimation, which is based on heuristics that may hold true only for restricted conditions. The next sections details the Dark Channel Prior (DCP) [He et al., 2009] adopted on hazy images, and the proposed extension to underwater images called Underwater Dark Channel Prior (UDCP).

3.1.1 Dark Channel Prior (DCP)

DCP is a statistical prior based on the observation that local patch on clear day images contain pixels whose intensity is very low in at least one color channel. It was firstly proposed by He et al. [2009], and extended by He et al. [2011]. These low intensity in the dark channel is mainly due to three factors, as described in He et al. [2009, 2011]: a) shadows in the images; b) colorful objects or surfaces where at least one color has low intensity and c) dark objects or surfaces.

They collected a large number of outdoor images and built histograms, and with those, they have shown that about 75 percent of the pixels in the dark channel have zero values, and the intensity of 90 percent of the pixels is below 25 in a scale of 8 bits, as shown Fig. 3.3. Some example of images and their DCP images are shown in Fig. 3.2. They cut out the sky regions, and the images are resized so that the maximum of width and height is 500 pixels. Those results provide a strong support to the dark channel prior assumption for outdoor images.

Formal description of the DCP prior defines the concept of dark channel as:

$$J^{dark}(\boldsymbol{x}) = \min_{\boldsymbol{y}\in\Omega(\boldsymbol{x})}\big(\min_{c\in R,G,B} J^c(\boldsymbol{y})\big), \tag{3.1}$$

where J^c is the scene radiance for each color channel and $\Omega(\boldsymbol{x})$ is a local patch centered at $\boldsymbol{x}$. This observation allows us to say that if J is an image acquired in non-participating media, except for the sky region, the intensity of J is low and tends to zero ($J^{dark} \to 0$).

Considering the optical model, it is possible to estimate the transmission based on the DCP. The global light constant A may be estimated in different ways, as described in Sec. 2.1. He et al. [2009] find the brightest pixel in the dark channel. Considering Eqs. 2.5 and 3.1, it is also possible to isolate the transmission in a local patch $\tilde{t}$ using DCP. Applying the minimum operation to both sides of the equation, we may estimate $\tilde{t}$ based on the image I and the global light A. Eq. 3.2 shows the estimation of $\tilde{t}$:

$$\tilde{t}(x) = 1 - \min_{y\in\Omega(x)}\left(\min_{c\in R,G,B}\frac{I^c(y)}{A^c}\right). \tag{3.2}$$

The dark channel prior does not concern sky regions. Fortunately, the color of the sky in a hazy image I is usually very close to the global light A, because the depth $d \to \infty$ and the transmission $t \to 0$. Thus, Eq. 3.2 gracefully handles both sky and non-sky regions.

Transmission $\tilde{t}$ is only an approximation of the transmission. Local patch affects the performance of the transmission estimation, generating artifacts. These artifacts

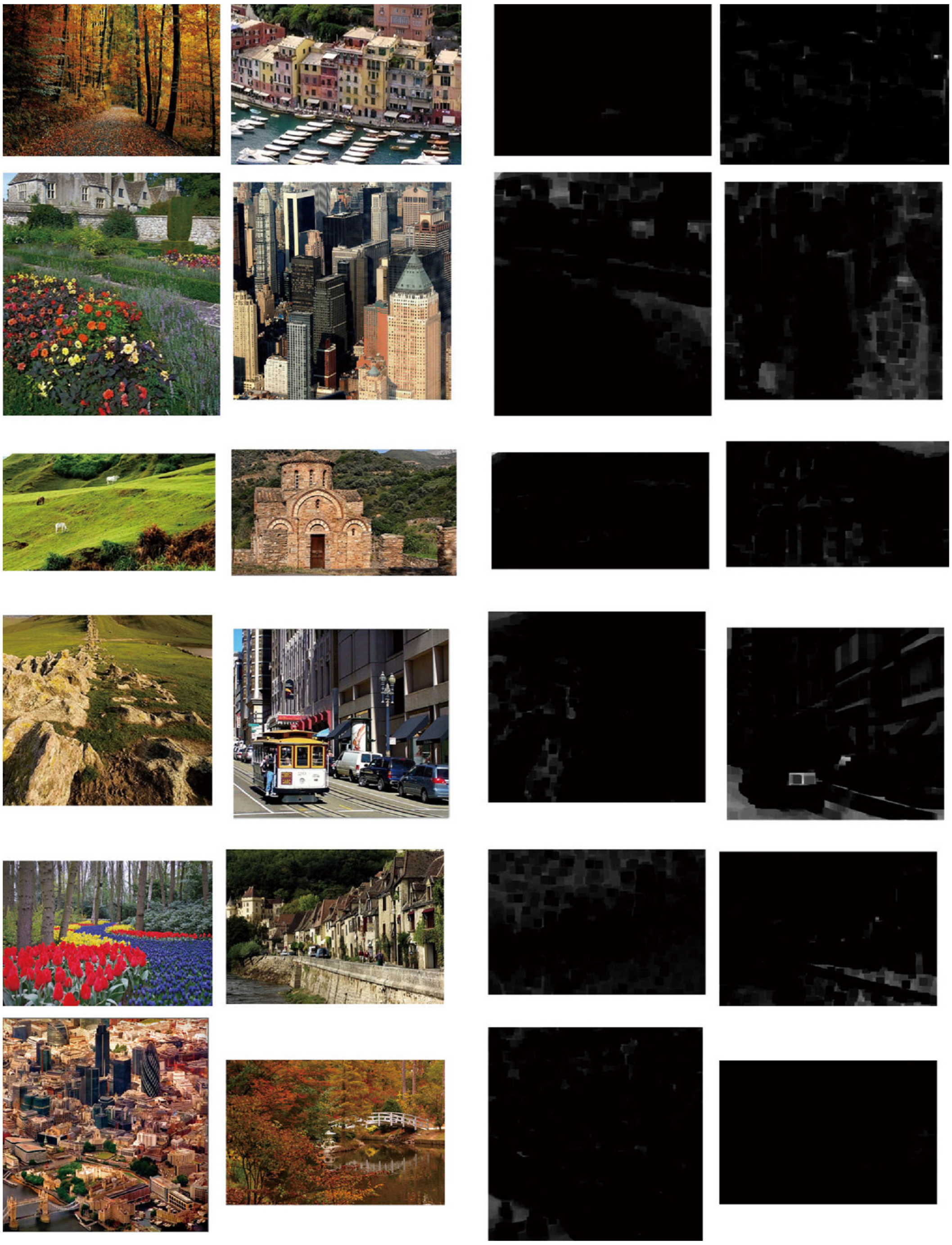
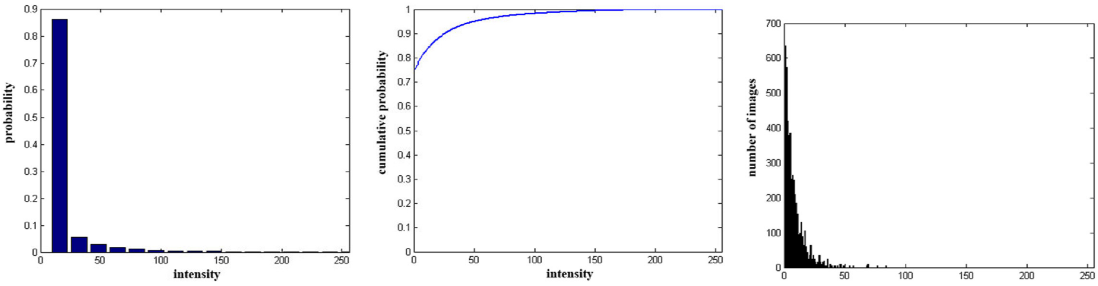


Figure 3.2. Dark channel prior [He, 2011]. Left: outdoor images acquired in non-participating media. Right: the corresponding dark channels. The dark channels are mostly dark. The images are resized so that the maximum of width and height is 500 pixels. Their dark channels are computed using a patch size 15×15.

are caused by the locally-uniform-transmission assumption, which does not hold near the depth edges. Therefore, halo artifacts appear near depth edges. We adopted the



(a) (b) (c)

Figure 3.3. Statistics of the dark channels [He, 2011]. (a) Distribution of the pixel intensity of all of the 5,000 dark channels (each bin represents 16 intensity levels). (b) Cumulative distribution. (c) Distribution of the average intensity of each dark channel. They support the assumption of darkness in at least on color channel.

guided filter [He et al., 2010] to refine the estimated transmission.

3.1.2 Underwater Dark Channel Prior (UDCP)

Because of the impressive results obtained by the work of He et al. [2009] and the similarities in the optical model of a haze image and an underwater image, some works using DCP to restore underwater images were proposed in the literature. One of the first works to apply DCP was Chao and Wang [2010], where the methodology proposed by [He et al., 2009] was directly applied in underwater images. The reported results show a limited number of experiments where the visual quality of the result does not show significant improvement, even for images with small degradation, as shown in Fig. 2.14. However, the authors raised an important issue related to the normalized image (I^c/A^c), that it reduces the effect of color absorption for different wavelength in underwater images.

Similarly, Chiang and Chen [2012] proposed an underwater image restoration method using standard DCP to find the transmission map t. They consider the attenuation coefficient η is known *a priori*. It allows to find the depth map, instead of the transmission. Furthermore, they also proposed a method to estimate the water depth where the image was acquired. Their method obtains significant results for real underwater images, but it is limited by the standard DCP method in underwater images and by the assumption that the image is predominantly blue.

The statistical correlation of a low Dark Channel in images acquired in non-participating medium is not easy to be tested for underwater images. It is due to the difficulty to obtain real images of underwater scene in an out-of-water condition.

However, the assumptions made by He et al. [2009] are still plausible, at least one color channel has some pixels whose intensity are close to zero. This low intensities are due to: a) shadows; b) objects or surfaces where at least one color is low intensity, *e.g* fishes, algae or corals; c) dark objects or surfaces, *e.g.* rocks or dark sediment.

Although the dark channel assumption seems to be correct, some problems arise from the wavelength independence assumption. There are many practical situations where the red channel is nearly dark, which corrupts the transmission estimate by the standard DCP. Indeed, the red channel suffers an aggressive decay caused by the absorption of the medium making it be approximately zero even in shallow waters. Thus, the information of the red channel is undependable.

Fig. 3.4 shows two examples of the behavior of the red channel. The first row shows a typical situation where the red channel is nearly dark, *i.e.* the red component is absorbed. Therefore, it corrupted the transmission estimate by the standard DCP (Fig. 3.4c). In the last row, we show an example that the red channel is approximately independent from the scene depth. Notice that the results obtained by DCP on blue and green channels (last column) are better than the DCP (third column) in both cases.

Considering the limitations in underwater conditions, we propose a new prior that considers just the green and the blue color channels. We named this prior Underwater Dark Channel Prior (UDCP). This phenomenon is mainly related to the high absorption effect in the red channel (Fig. 3.5) which imposes it to be near zero. Therefore, the classical dark channel does not hold the information about depth. This new prior allows us to invert the model and get the medium transmission. The medium transmission and the global light constant provide enough information to restore the images.

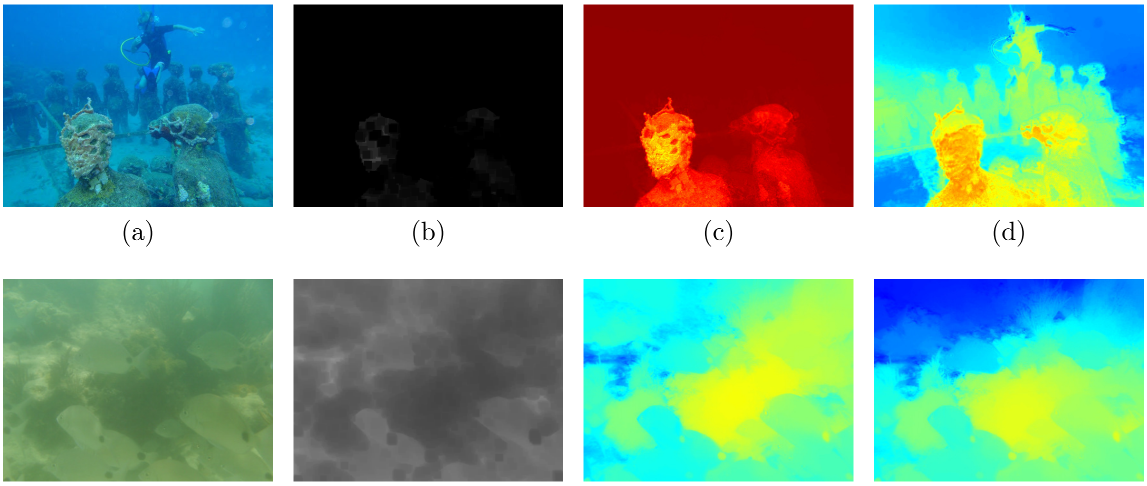
Similarly to DCP, the UDCP formulation is given by:

$$J^{UDCP}(\boldsymbol{x}) = \min_{\boldsymbol{y} \in \Omega(\boldsymbol{x})} (\min_{c \in G,B} J^c(\boldsymbol{y})). \tag{3.3}$$

The global light A is estimated by finding the brightest pixel in the underwater dark channel:

$$A = I(argmin_{\boldsymbol{x} \in I}(\min_{\boldsymbol{y} \in \Omega(\boldsymbol{x})} (\min_{c \in G,B} I^c(\boldsymbol{y})))). \tag{3.4}$$

Considering Eqs. 2.5 and 3.3, we isolate the transmission in a local patch $\tilde{t}$ as in the standard DCP. Applying the minimum operation to both sides, we can estimate $\tilde{t}$ based on the image I and the global light A:



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

(e) (f) (g) (h)

Figure 3.4. Common situation in which the red channel does not respect the wavelength independence and the direct relationship with the depth. Two underwater images are shown in the first column. In the second column, it is shown the minimum of the red channel ($\min_{\boldsymbol{y}\in\Omega} I^R(\boldsymbol{y})$) being quickly absorbed (b) and uncorrelated with depth (f). The third and the last columns show the colorized transmission maps estimated using DCP and UDCP, respectively. The colors vary from red to blue. The credits of the images: a) www.underwatersculpture.com; e) [Ancuti et al., 2012].

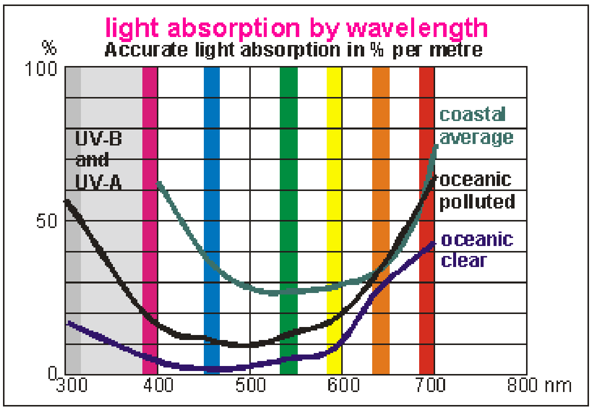


Figure 3.5. Light absorption for each wavelength in several type of waters [Anthoni, 2003]. The wavelengths related to the red color ($\approx 700nm$) suffer a stronger absorption, while the absorption on wavelengths related to blue ($\approx 475nm$) and green color ($\approx 510nm$) are more uniform and weaker.

$$\tilde{t}(\boldsymbol{x}) = 1 - \min_{\boldsymbol{y}\in\Omega(\boldsymbol{x})}\left(\min_{c\in G,B}\frac{I^c(\boldsymbol{y})}{A^c}\right). \tag{3.5}$$

We employ the guided filter [He et al., 2010] to refine the estimated transmission (Fig. 3.6) and define the $\Omega(\boldsymbol{x}) = 15 \times 15$, as proposed by He et al. [2009].

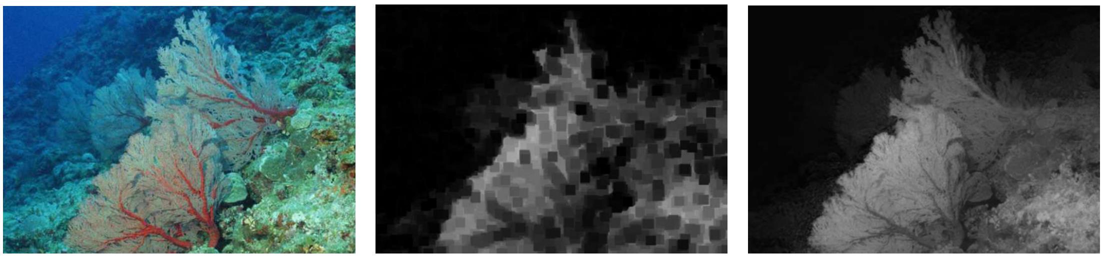

(a) (b) (c)

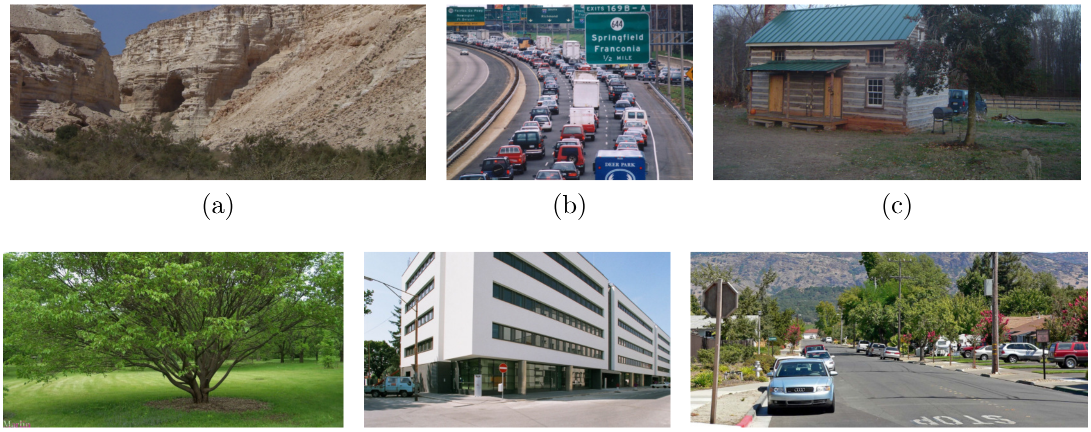
Figure 3.6. Example of Underwater Dark Channel Prior: a) Underwater Image [Chao and Wang, 2010; Carlevaris-Bianco et al., 2010]; b) Transmission $\tilde{t}$; and c) Refined transmission t.

3.1.3 Experimental Verification

We performed an experimental verification to evaluate the assumption of the new prior based on two statements: a) the main assumption of the DCP for outdoor scenes remains valid if only applied to green and blue channels, and b) the behavior of the UDCP histogram in underwater scenes is plausible.

Since He's dataset has not be made publicly available, we created our own following the guidelines proposed by He [2011]. The dataset is composed of 1,022 outdoor landscape images greater than 0.2 Mpixels from the SUN database [Xiao et al., 2010], see Fig. 3.7 for image samples. We divided the dataset into three categories: a) natural scenes, *i.e.* images without any human-made object, comprising of 274 images (first column in Fig. 3.7); b) human-made scenes, *i.e.* images predominantly composed of human-made object, comprising 427 images (second column in Fig. 3.7); and c) hybrid scenes, comprising 321 images (last column in Fig. 3.7). Tipically, underwater scenes tend to be natural. Thus, the dataset of natural scenes is important to identify the sensibility of the prior, *i.e.* whether the darkness of the images using a minimum operator is valid in this context. However, the He's dataset is composed of several scenes that includes human-made scenes. Thus, we classified the dataset into three groups to understand the effects in the darkness due to the characteristics of the scene.

We then compute the distribution of pixel intensities, where each bin contains 16 intensity levels from an interval of $[0; 255]$ (Fig. 3.8). The histograms were obtained by using each division of the dataset, *i.e.* natural, human-made and hybrid scenes, and all images of the dataset. In the Fig. 3.8, each row depicts the results for the minimum operator in a small patch window using only the red, green and blue channels, the DCP (dark channel in all channels) and the UDCP (dark channel in green and blue channels). Each column of the Fig. 3.8 shows the statistics for the three categories,



(a) (b) (c)

(d) (e) (f)

Figure 3.7. Example of outdoor images of our dataset, using images from SUN dataset [Xiao et al., 2010]. Each column shows one of the three categories: natural, human-made and hybrid scenes, respectively.

i.e. natural, human-made and hybrid. The last column shows the results for the full dataset.

Although our dataset and the dataset of He and colleagues are different, they were collected using the same guidelines, and thus, it is expected similarities between them. Indeed, the histograms present similarities but also important differences. The normalized frequency[1] of the first bin (intensities between 0-15) is smaller than presented by He [2011]. They reported $\approx 90\%$ for the first bin in the DCP while the normalized frequency in our dataset is $\approx 45\%$. One can see the highest frequency, $\approx 50\%$, is obtained for histograms of natural scenes (Fig. 3.8m) which is the expected case of typical underwater scenes. The most important observation is related to the significance of each channel for the prior. The lower intensity bins of the blue channel (figs. 3.8i-3.8l) are dominant, mainly in natural scenes. The red channel is still dark, but it is the most equalized histogram in all scenarios. The green channel presents similar behavior. Thus, absence of blue color in the final composition of the scene represents prevalence of this channel in both the DCP and the UDCP.

The second column in Fig. 3.8 shows human-made scenes. One can see a small, but significant, peak in the last bin which represents intensities between the interval of $[239, 255]$. It is an important source of error in the prior. However, the natural and hybrid scenes do not present this peak.

Finally, we can observe a close similarities between DCP and UDCP statistics

---

[1] The frequencies in the histograms are normalized in order to obtain a set of probabilities, *i.e.* the sum is equal to one.

Table 3.1. Correlation coefficient between DCP and the other priors.

	Natural	Human-made	Hybrid	Full
UDCP	0.9999	0.9994	0.9998	0.9998
Red	0.8049	0.9517	0.8465	0.8856
Green	0.7680	0.9230	0.7596	0.8351
Blue	0.9998	0.9976	0.9997	0.9995

in the histograms of Fig. 3.8. We show the Pearson's linear correlation coefficient [Gibbons and Chakraborti, 2010] in Table 3.1, which quantifies these similarities. The correlation coefficient ranges from $[-1;1]$, where a coefficient value close to one indicates that the relationship is almost perfect and negative values show that data are uncorrelated.

One can readily see that there is a strong correlation between DCP and UDCP, which means that both methods are based on similar assumptions about the scene, *i.e.* low intensity dark channel. The value of the correlation coefficient for the blue channel and DCP are approximately equal to one meaning that they are also strongly correlated. In natural and hybrid scenes, the correlation between DCP and green channel is the smallest because of the presence of grass and trees in the scenes, which causes an increase in the intensities of this color channel.

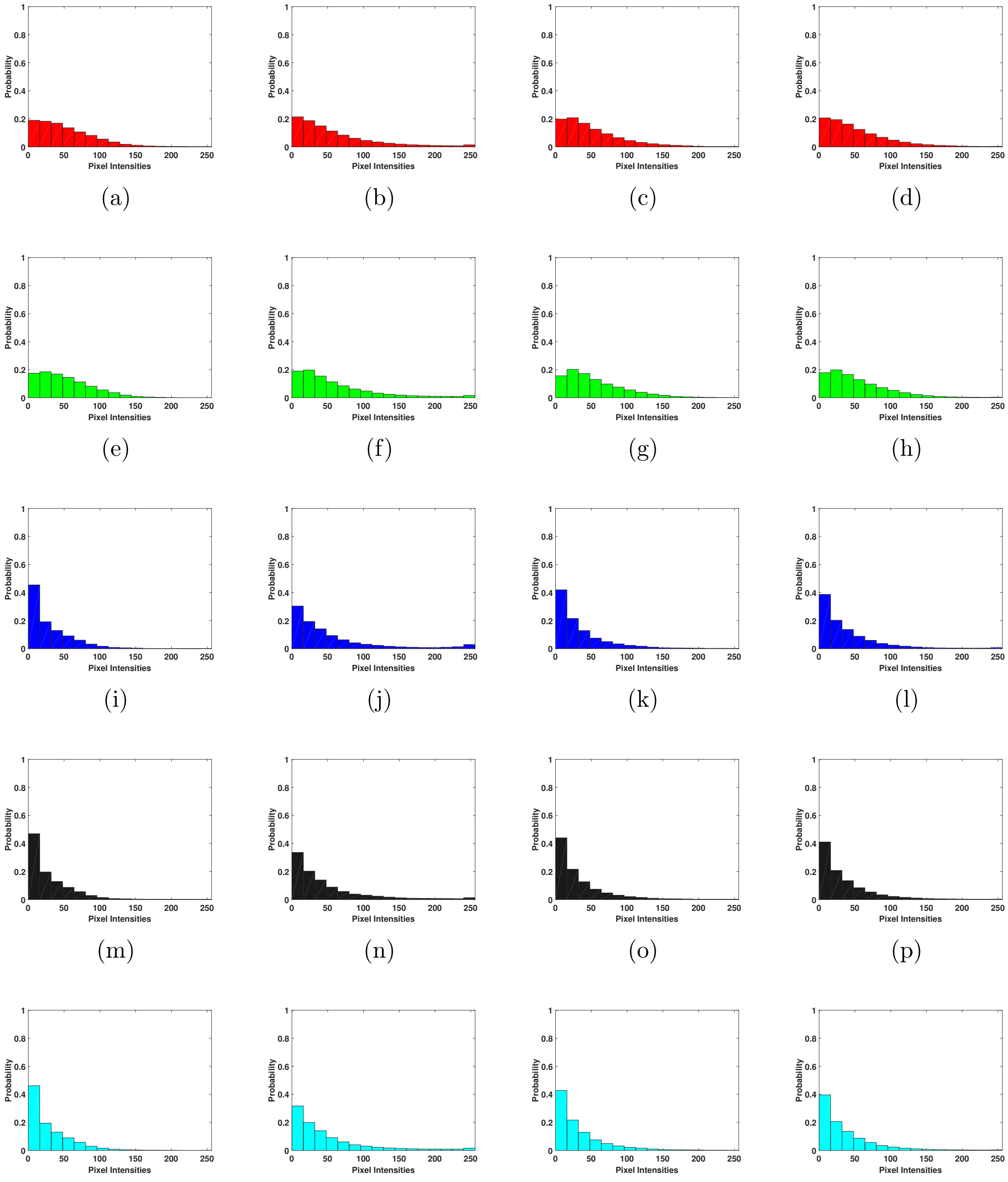
We also created two datasets of underwater images to evaluate the influence of the medium and to verify the UDCP assumptions. It also follows the guidelines of He [2011]. The first dataset (**Semi-Controlled**) was created by extracting the images from a single user of the Flickr website. This dataset contains 65 high quality photos acquired with the same camera and minimal offline post-processing. The images, which include coral reefs, rocks, marine animals, wreck, etc., were acquired during diving activities in several places of the world (thus with different turbidity levels). The horizon was manually cut out and the images were resized so that the maximum width and height is 600 pixels. Figs. 3.9a-3.9d show sample images of the reduced dataset.

The second dataset (**Images On the Wild - IOW**) was obtained by collecting images from several image search engines on the internet on the wild conditions. This dataset is composed of 171 underwater images acquired under diverse media conditions, water depth and scenes, which provides a rich source of information. All the images are approximately homogeneous illuminated that limits the water depth to shallow water. Figs. 3.9e-3.9h show some example of images of this dataset.

Differently from the histograms of outdoor scenes, the dark channel of the red channel is really dark, *i.e.* $\approx 90\%$ of the pixels are in the first bin (Fig. 3.10).

---

[1]



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

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

(q) (r) (s) (t)

Figure 3.8. Statistics of the dark channels produced in our dataset. The dataset is divided in three categories: natural, human-made and hybrid. They are shown in each column, respectively. The last column shows the statistics for the full dataset. The distribution of the pixel intensity of the dark channel are shown in each row, considering only the red, green, blue, DCP and UDCP, respectively.

This agrees with the assumption of UDCP, *i.e.* the highest absorption rate for the red channel. As expected, the dark channel for the blue and the green channels are

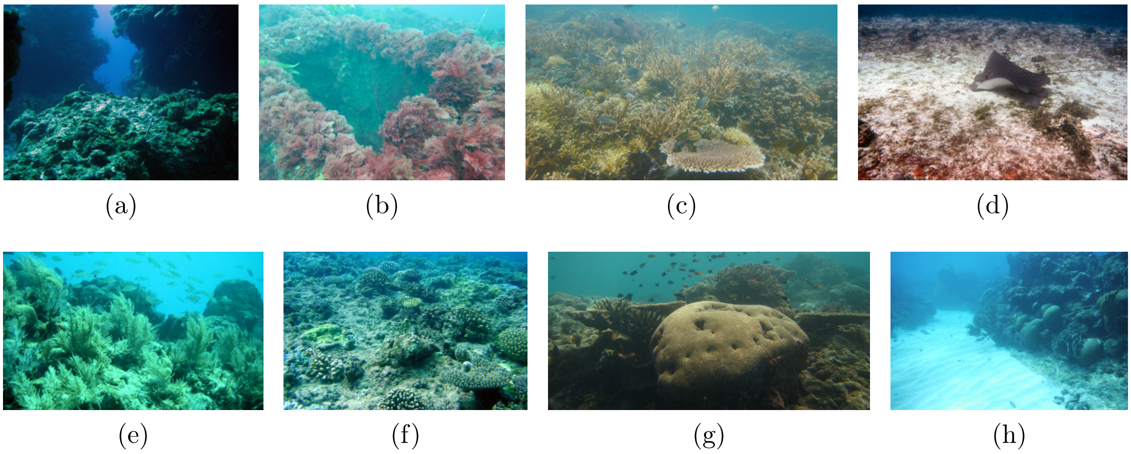
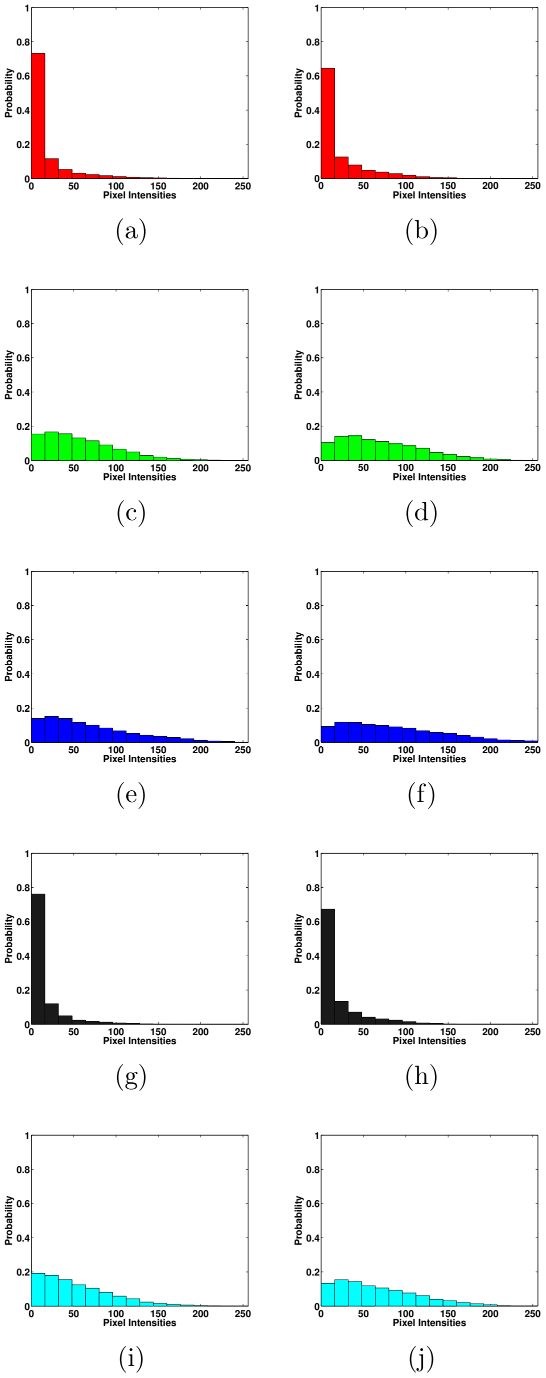


Figure 3.9. Example of underwater images from Semi-Controlled (first row) and IOW datasets (second row). The credits of the images: a-d) courtesy of Rémi Forget; e-f) SUN Dataset [Xiao et al., 2010]; g-h) courtesy of Kristina Maze.

similar but many values cover a broader range due to the effects of the interaction of light with the medium. The histograms of DCP are consistent with what we would expect for non-participating media. Hence, DCP is not able to recover adequately the medium transmission. However, the bin values in the UDCP histograms are more evenly distributed, which indicates that UDCP is better source of information about the medium transmission.

We also evaluated the UDCP using histogram obtained using real underwater videos. Differently from previous analysis, we preserve the horizon on the frames. Three videos with distinct medium properties were employed. They were acquired in shallow waters, but in many depths and regions of the world. They present different colors for the global light (A), the first is approximately white (RGB $[170, 170, 185]$), similar to hazy images, while the others are predominantly green ($[60, 188, 156]$) and blue ($[130, 250, 255]$), respectively. The first video is called **FURG** and it is composed of 4586 frames. This video was acquired by a Remotely Operated Vehicle (ROV) from NAUTEC-FURG in the Atlantic Ocean near the coast of the Rio Grande do Sul, Brazil. The second and third videos were acquired in the Guanacaste Coast Pacific at Costa Rica [Van der Berg, 2012]. They are composed of 210 frames each one and called **Costa Rica 1** and **Costa Rica 2**, respectively. Fig. 3.11 shows some frames from the videos.

The histograms generated for each video are shown in Fig. 3.12. Notice that the shape of the histograms are different from Semi-Controlled and IOW datasets due to the characteristics of the scenes and the temporal sequence, *i.e.* the scene is



(a) (b)

(c) (d)

(e) (f)

(g) (h)

(i) (j)

Figure 3.10. Statistics of the dark channels generated using underwater images in both datasets. The distribution of the pixel intensity of the dark channel, considering only the red, green, blue, DCP and UDCP (each bin represents 16 intensity levels) for each row, respectively. The Semi-Controlled and IOW datasets are shown in the first and the second columns, respectively.

approximately the same. One can see that a small peak in the first bin (low intensity) in the histogram of the FURG video (first column in Fig. 3.12) due to the analogical

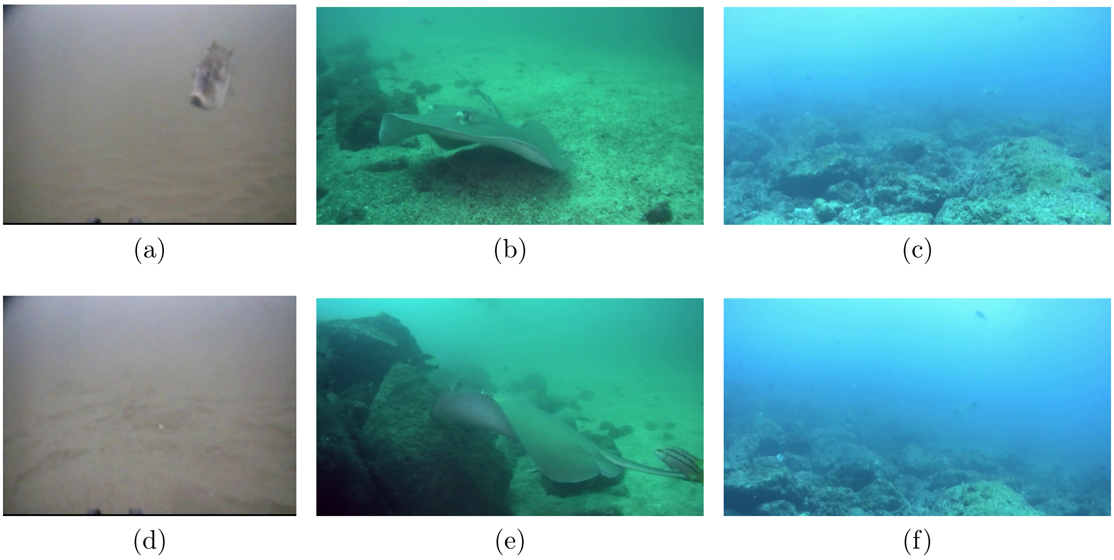


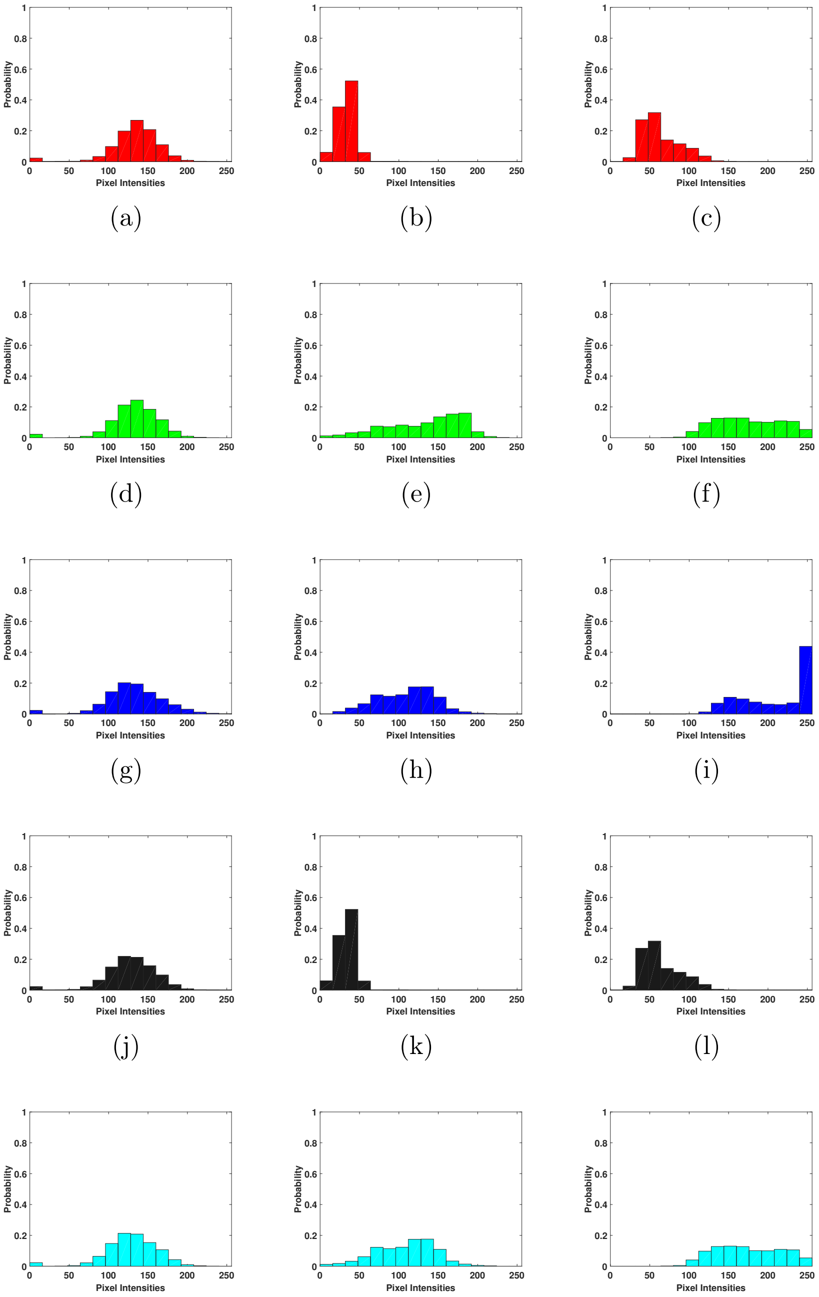
Figure 3.11. Example of frames from the underwater videos used for obtaining dark channel statistics. Frames from FURG video are shown in the first column, the videos Costa Rica 1 and 2 [Van der Berg, 2012] are shown in the next columns, respectively.

camera that produces black lines in the bottom of the frames (Figs. 3.11a and 3.11d), and the black parts of the vehicles that are imaged by the camera. This video presents similar characteristics to a hazy scene. Therefore, all histograms present similar shape, including the DCP and UDCP.

However, the intensities in the red channel in the Costa Rica videos are significantly smaller than the previous histograms. The DCP fails in this case, while the UDCP contains more information about the medium effects. The last video, Costa Rica 2, the intensities of green and blue are saturated, thus the histograms are higher than the normal. Nevertheless, the UDCP is still more informative than DCP.

These experiments showed that the statistics for the UDCP assumption is a more general supposition than the DCP assumption for underwater images. However, these results do not guarantee the quality of the estimated transmission. UDCP and DCP obtain similar histograms on natural scenes, as shown by the correlation analysis. These results indicate that both are based on similar assumptions.

Another important characteristic concerns the blue channel, which in natural scenes tends to be darker than the other channels. The underwater medium is typically blue, thus increasing the intensities of this color channel. This fact corroborates the underwater dark channel assumption.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

Figure 3.12. Statistics of the dark channels generated by using underwater videos. As in the previous histograms, the red, green, blue, DCP and UDCP are shown in each row, respectively. The histograms from FURG video are shown in the first column and the histograms from the videos Costa Rica 1 and 2 are shown in the last columns, respectively.

3.2 Optical Flow

The apparent motion of rigid objects in a scene may be inferred by the motion field obtained from an image sequence acquired by a camera. The motion field is defined as the velocity of moving image points estimated from a sequence of images acquired by a static camera from a moving scene or from a moving camera from a static scene. This projection of the 3-D velocity field onto the 2-D image plane is usually computed from the variation of image brightness between consecutive frames, and it is called optical flow [Horn and Schunck, 1981]. Optical flow estimation is an important stage to recover the spatial motion from a sequence of images, and it has been increasingly used in the past years with great success, despite its dependence on scene texture and illumination constancy. Nevertheless, the typical constant illumination assumption for computing the optical flow quickly breaks down in participating media even for small relative displacements.

In this section, we propose a new optical flow formulation that includes the optical model in the estimation. We called it Generalized Optical Flow Model (GOFM). This model extends the previous one, allowing the optical flow estimation in participating and non-participating media. For non-participating media, the model is simplified to become the Horn-Schunk model [Horn and Schunck, 1981], the most widely employed in the literature [Sun et al., 2010].

3.2.1 Standard Optical Flow

The field of optical flow has experienced impressive advances in the last thirty years [Sun et al., 2010]. The majority of these methods strongly resembles the original formulation proposed by Horn and Schunck [1981], which assumes constancy in the brightness patterns in the image.

Let $I(x, y, t)$ be the image brightness at pixel (x, y) on image plane at time t. Considering that the brightness of every point remains constant, the variation of pixel intensity in time is given by:

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0. \tag{3.6}$$

The partial derivatives can be estimated from the images. Assuming $\frac{dx}{dt} = u$ and $\frac{dy}{dt} = v$, we can rewrite Equation 3.6 as:

$$I_x u + I_y v + I_t = 0. \tag{3.7}$$

Since we have one Equation 3.7 per pixel, it is not possible to solve this system without including an additional constraint. Several constraints can be assumed such as: global smoothness [Horn and Schunck, 1981], local smoothness [Lucas and Kanade, 1981], texture constancy [Arredondo et al., 2004], and multiple color channel [Ohta, 1989].

Although works such as Negahdaripour [1998], Haussecker and Fleet [2001] and Molnár et al. [2010] have proposed to deal with the optical flow estimation in outdoor environments, they did not address the optical model and the relation between brightness and contrast with the depths of the scene. In general, the optical effects are handled as a variation in the illumination, contrast or brightness of a pixel, leaving out the scattering and the absorption effects.

3.2.2 Optical Flow in Participating Media

In order to tackle with the scattering and absorption effects in underwater images to estimate the optical flow, Arredondo et al. [2004] proposed a method that combines intensities and texture. Despite the obtained results in real underwater images, these were shown only for images acquired from scenes near to the camera. Fang et al. [2009] proposed an improvement in the optical flow estimation using neighborhood voting schema. Nevertheless, it does not consider explicitly the optical model.

Negahdaripour and Madjidi [Negahdaripour and Madjidi, 2003; Madjidi and Negahdaripour, 2006] proposed a method to estimate optical flow in underwater environment. In their work, scattering and absorption effects were modeled based on a simplified version of Jaffe-McGlamery model [McGlamery, 1980; Jaffe, 1990] where only the direct component is considered. In fact, the model is only used for getting insights about the adopted color space and the applicability of the model. The authors worked on the HSV color space and extended the model previously proposed in Negahdaripour [1998]. The model is called Generalized Dynamic Image Model (GDIM) and described by:

$$I_x u + I_y v + I_t - I\ m = 0, \tag{3.8}$$

where $I\ m$ is a new radiometric term, which takes into account the brightness variation of a scene point between the two views, m, and the brightness of the image, I. Beyond the velocities u and v, the brightness variation m is also estimated in the optical flow.

Even though this model relaxes the brightness constancy assumption, it does not consider the influence of the distance between scene and camera, which is responsible for a significant impact on the quality of images acquired in participating media.

3.2.3 Generalized Optical Flow Model - GOFM

In spite of advances in optical flow in past years, virtually all proposed methodologies are not catered for participating media, since the main assumption of brightness constancy is not valid in these environments. As described by the model in Eq. 2.5, the brightness in the image $I(x, y, t)$ is not constant because of the effects of the medium.

Nevertheless, one can assume that the radiance $J(x, y, t)$ is approximately constant. Considering this assumption, we can differentiate a generalized model in order to obtain the optical flow from images acquired in participating media. We call this new model Generalized Optical Flow Model (GOFM) and it is given by:

$$\frac{dJ}{dt} = \frac{\partial J}{\partial x}\frac{dx}{dt} + \frac{\partial J}{\partial y}\frac{dy}{dt} + \frac{\partial J}{\partial t} = 0, \tag{3.9}$$

where the radiance $J(x, y, t)$ is given by Eq. 2.6. In this section the variable t is related with time, while the medium transmission is renamed to t_r.

The partial derivatives $\frac{\partial J}{\partial x}$, $\frac{\partial J}{\partial y}$ and $\frac{\partial J}{\partial t}$ are computed as:

$$\begin{aligned}
\frac{\partial J}{\partial x} &= \frac{\partial(\frac{I-A}{t_r} + A)}{\partial x} \\
&= \frac{1}{t_r}\frac{\partial I}{\partial x} + (I - A)\frac{\partial \frac{1}{t_r}}{\partial x}, && (3.10) \\
\frac{\partial J}{\partial y} &= \frac{1}{t_r}\frac{\partial I}{\partial y} + (I - A)\frac{\partial \frac{1}{t_r}}{\partial y}, && (3.11) \\
\frac{\partial J}{\partial t} &= \frac{1}{t_r}\frac{\partial I}{\partial t} + (I - A)\frac{\partial \frac{1}{t_r}}{\partial t}. && (3.12)
\end{aligned}$$

Finally, applying the eqs. 3.10, 3.11, and 3.12 into Eq. 3.9, we obtain the generalized optical flow model defined by:

$$\begin{aligned}
t_i(\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t}) + (I - A)(\frac{\partial t_i}{\partial x}u + \frac{\partial t_i}{\partial y}v + \frac{\partial t_i}{\partial t}) = 0 \\
t_i(I_x u + I_y v + I_t) + (I - A)(T_x u + T_y v + T_t) = 0, && (3.13)
\end{aligned}$$

where the image derivatives can be estimated as proposed by Sun et al. [2010], $t_i = t_r^{-1}$ is the inverse of the scene transmission and T_x, T_y and T_t are the inverse medium transmission derivatives.

In non-participating media, the attenuation coefficient is approximately zero, thus the medium transmission is a unitary matrix and the derivatives T are zero. Therefore, the model obtained is a generalization of the brightness constancy model proposed by Horn and Schunck [1981], defined in Eq. 3.7.

In order to improve the numerical stability of the model, we have used $t_i = e^{\eta d(x,y,t)} = e^{D(x,y,t)}$. Thus, the new function is estimated as $D(x,y,t) = -\log t_r(x,y,t)$. By using this assumption, the final version of the model is given by:

$$(I_x^c u + I_y^c v + I_t^c) + (I^c - A^c)(D_x^c u + D_y^c v + D_t^c) = 0, \tag{3.14}$$

where the derivatives of function $D(x,y,t)$ in relation to x, y and t are defined as D_x, D_y and D_t, respectively. The index c represents each channel in the RGB color channel.

In participating media, specially in underwater environments, the spectral characteristics of the medium can modify the three channel contents significantly, to the point that this can provide valuable information for the optical flow estimation [Madjidi and Negahdaripour, 2006].

3.2.4 Implementation and Optimization

This section details the implementation and optimization that enable GOFM formulation to provide competitive results. The accuracy of optical flow estimation algorithms has been improving steadily as evidenced by results on the Middlebury optical flow benchmark [Baker et al., 2011]. However, the typical formulation is the classical work of Horn and Schunck [1981]. This formulation is still performing surprisingly well when combined with modern optimization and implementation techniques [Sun, 2013]. Thus, a good optical formulation depends on mainly four factors, as shown in Fig. 3.13. The main contribution is the our new objective function (GOFM) highlighted by a black contour. However, the other parts are also important, thus we adopted state-of-the-art methods.

The classical model cannot be solved pointwise, since the number of parameters to be estimated is larger than the number of linearly independent equations. This indeterminacy is called the aperture problem: there is no sufficient information to recover the optical flow at one point by only looking at first order derivatives of the image intensity [Meinhardt-Llopis et al., 2013]. Color information provides valuable information, mainly in participating medium that presents different absorption rates for each wavelength, such as underwater environment. Since the contribution of the color information in hazy scenes is limited, other constraint needs to be applied to solve the aperture problem. We adopted two classical methods which are the global smoothness constraint [Horn and Schunck, 1981] and the local smoothness constraint [Lucas and Kanade, 1981].

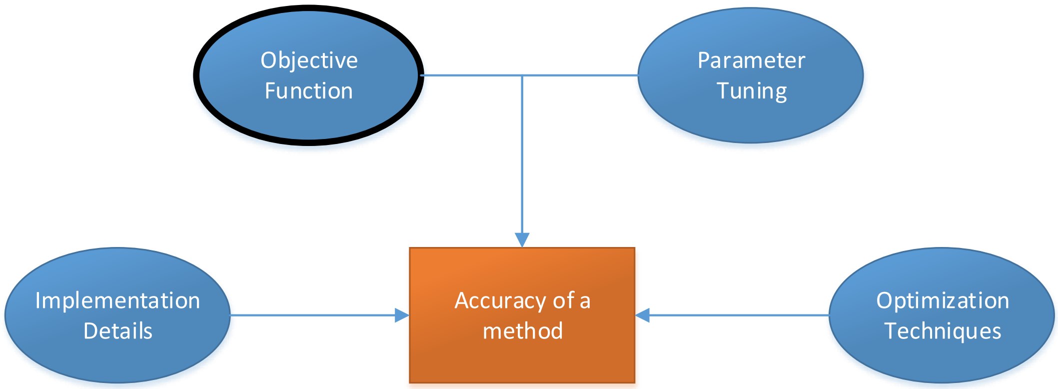


Figure 3.13. The accuracy of an optical flow formulation depends on mainly four factor: objective function, implementation details, parameter tuning and optimization techniques. The main contribution is the our new objective function (GOFM) highlighted by a black contour. Image adapted from Sun [2013].

Following the four factors of the Fig. 3.13, we also adopted a set of steps to obtain a good implementation based on GOFM. Two optical flow approaches are derived from the methodology. Firstly, the method is implemented in the simplest way without additional optimization and multiple scale representation, since the main objective is to compare the GOFM formulation with the state-of-art in participating medium under similar conditions, instead of the best possible performance. The second optical flow model is based on the recent work of Sun [2013], which evaluated the state-of-art methods in order to identify the most important characteristics from them.

The first approach is based on the classic method of Barron et al. [1994] that provides many useful and simple suggestions about the practical aspects of the optical flow estimation problem. The local smoothness constraint [Lucas and Kanade, 1981] was adopted because of the simplicity and inspired by the state-of-art in underwater optical flow [Madjidi and Negahdaripour, 2006]. The optimization technique is a simple least-squares solution to the system of equations. Furthermore, we smoothed all images using a Gaussian kernel with different values of σ in order to reduce noise [Barron et al., 1994]. The first-order derivatives of the image functions are estimated using five-points kernel. Finally, the flow field is post-processed removing the boundary pixels and applying median filter, as proposed by Sun et al. [2010].

A more accurate approach is also proposed. The recent work of Sun et al. [2014] that proposes a modern implementation of the Horn-Schunk methodology is the basis of this approach. That modern implementation obtains surprisingly competitive results on the Middlebury optical flow benchmark [Baker et al., 2011], although the oversmoothing and sensibility to outliers. Therefore, we optimize the energy function as described by Horn and Schunck [1981]:

$$E(u,v) = \sum_{c \in RGB} \sum_{\boldsymbol{x} \in I} (\mathcal{E}_d^2 + \alpha^2 \mathcal{E}_s^2), \tag{3.15}$$

where $\mathcal{E}_d^2$ is the rate of change of scene radiance brightness defined by Eq. 3.14, α^2 is the weighting factor, and $\mathcal{E}_s^2$ is the spatial penalty function that measures the global smoothness in the velocity field given by:

$$\mathcal{E}_s^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2. \tag{3.16}$$

The optimization technique is still a simple least-squares solution to the system of equations, but in a global sense instead of a local optimization. Furthermore, the optimization is performed using a standard incremental multi-resolution technique [Black and Anandan, 1996] to estimate flow fields with large displacement. Initially, the optical flow is estimated at the coarse level. This flow is applied to warp the second image toward the first at the next finer level. Moreover, a flow increment is calculated between the first image and the warped second image. This process is repeated for all levels. The number of levels of the pyramid is adaptively determined so that the top level has a width or height of around 20 to 30 pixels. The standard deviation of the Gaussian filter is set to be $\frac{1}{\sqrt{2d}}$, where d is the downsampling factor as proposed by Sun et al. [2014] ($d = 0.8$). At each pyramid level, we perform ten warping steps to compute the flow increment.

The data term is linearized around the current flow estimate at each iteration. The first-order derivatives of the image functions are estimated using five-points kernel, and warping the second image and its derivatives toward the first using the current flow estimate by bicubic interpolation. For pixels moving out of the image boundaries, we set both their temporal and spatial derivatives to zero. After each warping step, we apply a 5 × 5 median filter to the newly computed flow field in order to remove outliers.

These optimizations are a good trade-off between accuracy and temporal performance [Sun et al., 2014]. Using these optimizations, the modern implementation of Horn-Schunk [Sun et al., 2014] outperforms recent methodologies in some challenger datasets as MPI [Butler et al., 2012] and KITTI [Geiger et al., 2012].

3.3 Structure-from-Motion

This procedure presents a well-defined objective: providing depth maps from a pair of images based on structure-from-motion techniques. The camera calibration is previously known, and the correspondences between points are previously estimated using optical flow. In this section, we detail the adopted methodology which is based on continuous essential matrix, triangulation and bundle adjustment.

3.3.1 Camera Calibration and Intrinsic Parameters

The intrinsic camera parameters model the projection of points in the camera's coordinate system onto the image plane [Hartley and Zisserman, 2003]:

$$\boldsymbol{x}_H = K\boldsymbol{X}, \tag{3.17}$$

where $\boldsymbol{x}_H = (x, y, w)^T$ are homogeneous image pixel coordinates, and $\boldsymbol{X} = (X, Y, Z)^T \in \mathbb{R}^3$ is a point in the camera frame. The camera calibration matrix $K \in \mathbb{R}^{3\times3}$ is given by:

$$K = \begin{bmatrix} f_x & \gamma & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.18}$$

where the focal length is $f = (f_x, f_y)^T$, γ is the skew factor, and $p = (p_x, p_y)^T$ is the principal point. The skew factor generally is zero [Hartley and Zisserman, 2003]. The linear relation modeled by Eq. 3.17 is called pinhole camera model or perspective model. Typically, camera system is composed of a lens system that shows distortion. It can be modeled as a non-linear function of the projected image coordinates [Zhang, 2000]. The coefficients of the distortion function are counted among the internal parameters.

Furthermore, camera parameters are also composed of extrinsic parameters that describe the pose of the camera in the 3D world. They are given as a matrix $C = [R|t] \in \mathbb{R}^{3\times4}$, where $R \in \mathbb{R}^{3\times3}$ is the rotation matrix and t is the translation matrix.

The combination of internal and external parameters given the projection matrix $P = KC = K[R|t]$, also called camera matrix, that models the complete relation between points in the 3D world and 2D image points. Thus, we can redefine the relation expressed in Eq. 3.17 to deal with points $\boldsymbol{X}_H^{world}$:

$$\boldsymbol{x}_H = P\boldsymbol{X}_H^{world} = K[R|t]\boldsymbol{X}_H^{world}. \tag{3.19}$$

The process of finding internal and/or external parameters is called camera calibration. We consider as a calibrated camera that whose intrinsic parameters are previously known. The calibration method of Zhang [2000] is still widely applied because of its robustness and simplicity. It is adopted in the present work using the implementation of Bouguet [2013].

For haze/fog scenes, the medium does not change a lot the parameters of the camera. However the camera calibration in underwater images presents some specificity. Cameras immersed in underwater environments are usually confined in underwater housing filled with air, viewing the scene through a piece of glass. In case of this glass being a flat port, the light rays entering the camera housing are refracted twice, once at the water-glass interface and again at the glass-air interface. The double refraction causes the usually straight rays of light to bend and change direction depending on the interface incidence angles.

Treibitz et al. [2008] show that the perspective camera model might fail underwater because the rays do not intersect in one common center of projection. Nonetheless, the perspective camera model is the most applied for underwater images because of its simplicity [Brandou et al., 2007; Johnson-Roberson et al., 2010], approximating the refractive effect by calibrated parameters. Lavest et al. [2000] showed that a camera calibrated below water approximates refraction with focal length and radial distortion. Sedlazeck and Koch [2012] showed that the principal point and camera pose absorb some of this model error in addition to focal length and radial distortion. Thus, we adopted the perspective camera model, but calibrating the camera in the underwater medium, *i.e.* the images of the classic chessboard are acquired in underwater environment. As described by Lavest et al. [2000], it is possible to calibrate a camera in the air and obtaining the same calibration parameters of the camera immersed. However, their method is not adopted in the present work because of its complexity. This procedure may be useful for deep-water applications.

3.3.2 Pose Estimation and Triangulation

The estimation of the extrinsic parameters of the camera consists of determining the rotation R and the translation t of a calibrated camera, given a set of correspondences points between image points. Although there exist algorithms to estimate pose from uncalibrated camera [Koch et al., 2000; Lhuillier and Quan, 2005], the focus here is on the case where the intrinsic parameters are previously known.

The discrete epipolar constraint is a well studied tool to obtain the extrinsic parameters of the calibrated cameras by means of the Discrete Essential Matrix (DEM).

However, the continuous epipolar constraint is the adequate tool for the case of correspondences obtained by optical flow. As detailed in Sec. 3.2, our approach is based on a modified optical flow method. Thus, the Continuous Essential Matrix (CEM) is adopted. Following the nomenclature adopted by Ma et al. [2006], the continuous epipolar constraint is defined as:

$$\begin{aligned} \boldsymbol{u}^T\hat{v}\boldsymbol{x}_H + \boldsymbol{x}_H^T\hat{w}\hat{v}\boldsymbol{x}_H = 0, \\ \boldsymbol{u}^T\hat{v}\boldsymbol{x}_H + \boldsymbol{x}_H^T s\boldsymbol{x}_H = 0, \end{aligned} \tag{3.20}$$

where $\boldsymbol{x}_H$ is the image point, $\boldsymbol{u}$ is the image motion field that can be approximated by the optical flow vector. The skew-symmetric matrix $\hat{w}$ and $\hat{v}$ represent the body velocities w and $v \in \mathfrak{se}(3)$, respectively. The $\mathfrak{se}(3)$ is the Lie algebra associated to the Lie group called special Euclidian group $SE(3)$ [Murray et al., 1994]. The special Euclian group allows to model rotations and translations. The symmetric epipolar component s is defined as $0.5(\hat{w}\hat{v} + \hat{v}\hat{w}) \in \mathbb{R}^{3\times3}$. This equation suggests some redundancy that explains different forms of the continuous epipolar constraint in the literature [Soatto et al., 1996]. We have adopted the form described in Ma et al. [2006]. Its structure is substantially different from the DEM and cannot be derived as a first-order approximation.

The CEM is defined as:

$$E' = \left\{ \begin{bmatrix} \hat{v} \\ 0.5(\hat{w}\hat{v} + \hat{v}\hat{w}) \end{bmatrix} \middle| w \in \mathbb{R}^3, v \in \mathbb{S}^2 \right\} \subset \mathbb{R}^{6\times3}. \tag{3.21}$$

As v may be recovered only up to a constant scale, the matrix E' belongs to the normalized continuous essential matrix with v is normalized to be 1 [Ma et al., 2006]. More details about the continuous epipolar geometry and its geometric interpretation is found in Soatto et al. [1996] and Ma et al. [2006].

Armangue et al. [2001] provide a comparative study between algorithms to estimate the fundamental matrix. Following this study, the most accurate methods for real image are the algorithms for robust fitting of models because of their capability to deal with data outliers. Due to the similarity between the fundamental matrix and our problem of estimation of the CEM, we adopted the MSAC approach [Torr and Zisserman, 2000] and the eight-point linear algorithm for estimating the CEM matrix [Ma et al., 2006]. It is detailed in Appendix A.2. Furthermore, a comparative study between the CEM and DEM is also performed in the Appendix A.3.

After recover the body velocities v and $w \in \mathfrak{se}(3)$, we determine the rotation matrix R and the translation vector t of the camera. Following the mapping described in Murray et al. [1994] and Soatto et al. [1996]:

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} = \exp \begin{pmatrix} \hat{w} & v \\ 0 & 0 \end{pmatrix}. \tag{3.22}$$

The Rodrigues' formulas provide a convenient way of computing the exponential mapping to the Rotation Matrix R:

$$R = I + \frac{\hat{w}}{||w||} \sin(||w||) + \frac{\hat{w}^2}{||w||^2}(1 - \cos(||w||), \tag{3.23}$$

where I is the identity matrix. The translation vector t is obtained from the linear velocity v as defined in Soatto et al. [1996]:

$$t = \frac{1}{||w||}((I - R)\hat{w} + ww^T)v. \tag{3.24}$$

Since E' and $-E'$ satisfies the continuous epipolar constraint, both $(w, \pm v)$ are possible solution for a given set of optical flows. We solved this ambiguous solution by enforcing the positive depth constraint [Ma et al., 2006], after triangulation of the points and applying the chirality constraint [Hartley and Zisserman, 2003].

After the triangulation method (detailed in Appendix A.4), we obtained a set of 3D points. The rigid body transformation computed using MSAC-based CEM is better than provided using MSAC-based DEM, as described in details in the Appendix A. However, the estimation is still sensible to noise and misestimation of the optical flow. Thus, we performed a classical Bundle Adjustment (BA) optimization [Triggs et al., 2000; Hartley and Zisserman, 2003] to compute a reliable set of 3D points (detailed in Appendix A.5).

We compute the distances between 3D points and optical center of each camera, which enable us to obtain the distance in the Line of Sight (LoS) for each pixel. Although the correspondences between points are dense, they are prone to outliers and missing values. The missing values are due to the correspondence to be occluded or out of the image. Therefore, the depth maps present some discontinuity and holes. We applied two algorithms to solve these problems. Firstly, the inpainting approach [Bertalmio et al., 2000], or image interpolation, is used to fill the gaps in the generated depth maps. After that, we applied the guided filter method [He et al., 2010] to smooth the depth maps and to improve the edges discontinuities.

3.4 Parameter Estimation and Restoration

This procedure of the methodology presents two main objectives: estimating the attenuation coefficient and restoring the image.

The image restoration follows the model defined in Eq. 2.6. The direct component $I_d = J(\boldsymbol{x})t(\boldsymbol{x})$ is close to zero when $t(\boldsymbol{x})$ is small. In this case, the recovered scene radiance $J(\boldsymbol{x})$ is prone to noise. Thus, we restrict the transmission $t(\boldsymbol{x})$ by a lower bound t_0, *i.e.*, we preserve a small amount of scattering in strongly degraded regions. The final restored image $J(\boldsymbol{x})$ for each color channel is given by:

$$J(\boldsymbol{x}) = \frac{I(\boldsymbol{x}) - A}{\max(t_0, t(\boldsymbol{x}))} + A, \tag{3.25}$$

where a typical value of t_0 is 0.1 [He et al., 2009]. For the underwater image, t_0 is defined on $[0.05; 0.1]$ depending on the participating medium.

The estimation of the medium transmission is based on the knowledge about the depth map of the image, but it is also required the knowlegde of the attenuation coefficient. We proposed two different methods for this estimation. We followed the idea of an automatic online estimation proposed by Nascimento [2008] where underwater images acquired from a stereo pair are used for estimating the medium parameters. However, the present work employ a monocular pair of images.

In the first method, we proposed a new interpretation from the optical model, Eq. 2.5, that allows to estimate the attenuation coefficient, η, based on the depth maps in two consecutive frames. We call this method as **MSAC-Single**. Assuming the same 3D point in the scene and a small displacement between images, the model is given by:

$$\begin{aligned} I^t - A &= (J - A)e^{-\eta d}, \\ I^{t+1} - A &= (J - A)e^{-\eta(d+\Delta d)}, \end{aligned} \tag{3.26}$$

where I^t and I^{t+1} are the image acquired in a participating medium in the time t and $t+1$, respectively. d is the depth in time t, and Δd is the depth variation in the LoS between time t and $t+1$. Dividing both equations, we obtained the model:

$$\begin{aligned}
\frac{I^{t+1}-A}{I^{t}-A} &= \frac{e^{-\eta(d+\Delta d)}}{e^{-\eta d}} = e^{-\eta\Delta d}, \\
\ln\frac{I^{t+1}-A}{I^{t}-A} &= -\eta\Delta d, \\
\eta &= -\frac{1}{\Delta d}\ln\frac{I^{t+1}-A}{I^{t}-A}. \qquad (3.27)
\end{aligned}$$

The problem can be reduced to line fitting because the attenuation coefficient is assumed as constant for the entire image. In this case, we are interested in the slope of the line. The estimation of correspondence between points, *i.e.* optical flow, and the depth maps are subject to noise and error, thus the line fitting needs to be robust to outliers.

We adopted a line fitting approach based on the MSAC [Torr and Zisserman, 2000]. MSAC is an iterative method to estimate parameters from a set of observed data which contains outliers, similarly to the classic RANSAC [Fischler and Bolles, 1981]. Differently from the original formulation of RANSAC that the rank is based on the cardinality of the consensus set, MSAC evaluates the quality of the consensus set calculating its cost function. The approach adopted is similar to the MSAC-based algorithm described in Sec. 3.3.2. We select randomly one image point and fit the attenuation coefficient using Eq. 3.27. This procedure is repeated N times, and the best attenuation coefficient is chosen, *i.e.* the coefficient that represents the largest number of point. The value of N is typically estimated based on a false alarm rate ϵ, *i.e* the probability we never pick a good minimal sample set, as described by Zuliani [2008]:

$$N = \frac{\log\epsilon}{\log\left(1-\left(\frac{Nb_I}{Nb_t}\right)^{k}\right)}, \qquad (3.28)$$

where Nb_I is the number of inliers in the set, Nb_t is the total number of input data, and k is the minimal sample set dimension. This equation assumes $Nb_I \gg k$ and it is a valid approximation whether we assume that all the elements in the dataset have the same probability of being selected. As the number of inliers could be small, this equation tends to a large number of iterations. As previously reported, the number of iterations should be *over-optimistic* [Tordoff and Murray, 2005] since it is difficult to obtain a fair enough estimation of the rate of false alarm. Thus, we defined the N as $10,000$ iteration. The practical experience shows the method usually converges in less than $1,000$ iteration.

Another important parameter of the method is the δ (Eq. 3.29) that defines whether a input data is a inlier in the estimated model, *i.e.* the maximum tolerated distance from the model. We estimated δ using the classical assumption of Gaussian noise $\mathcal{N}(0, \sigma_{MSAC})$ in the data. The σ_{MSAC} is a parameter that needs to be set in the method. Furthermore, the Chi squared probability threshold for inliers P_{inlier} is also required, *i.e.* the error generated by a true inlier contaminated with Gaussian noise. P_{inlier} defines the probability that a point whose squared error is less than δ is an inlier. The δ is defined as [Zuliani, 2008]:

$$\delta = \sigma_{MSAC}\sqrt{F_{\chi^2}^{-1}(P_{inlier})}, \tag{3.29}$$

where $F_{\chi^2}^{-1}$ is the inverse cumulative distribution function associated with a χ^2 random variable. Since these parameters are hard to tuning and typically unknown, we empirically defined the value of as $P_{inlier} = 0.99$ and $\sigma_{MSAC} \in [0.05; 0.2]$. The MSAC method is repeated for each RGB color channel.

This approach is highly dependent on the depth maps and the correspondence between pixels. However, both are sensitive to noise and misestimation. Therefore, this approach suffers with a high rate of outliers that compromise its performance. Our idea is to impose more constraint to improve our capability to deal with outliers.

The new proposed constraint is inspired in the work of Roser et al. [2014] based on the Black patch assumption. They assume that the 1% of lowest intensity pixels in each color channel for all depths are originated from black objects in the color channel. Indeed, the object does not need to be black but dark in the specific color channel, *e.g.* a red coral is "black" in the blue and green channel assuming a RGB color channel. Fig. 3.14 illustrates the idea. Although this assumption is related with the dark channel assumption, it is slightly different because it is valid for a small portion of the pixels instead of all patches on image.

Based on this assumption, we proposed the second method called **MSAC-Multiple**. It imposes two constraints:

$$\begin{aligned} I^{t+1} &= A(1 - e^{-\eta(d+\Delta d)}), \\ I^{t} &= A(1 - e^{-\eta d}), \quad \{\forall I | J \approx 0\}. \end{aligned} \tag{3.30}$$

Assuming the 1% of lowest pixels in each color channel are usually concentrate on the lowest depths, so they do not carry on much information about the medium. Roser et al. [2014] uniformly selects the lowest intensity point in all depth bins, named

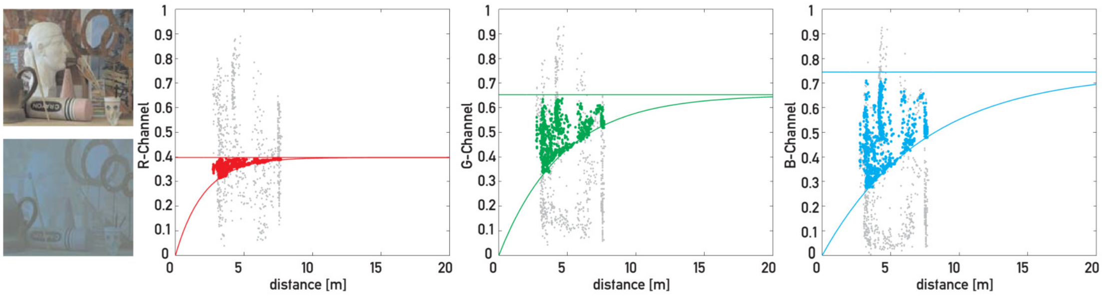


Figure 3.14. Illustrative diagram provided by Roser et al. [2014] about the behavior of the Black patch assumption. The effects of the medium are simulated on a image obtained from the Middlebury dataset [Scharstein and Szeliski, 2002]. The pixels in the original image are shown in gray, while the pixels after simulation are shown in the respective colors, *i.e.* red, green and blue. The range of scene point are independent from scene distance and possess a wide range of values (gray pixels). The deviation of observed scene radiances in simulated images decreases with scene distance, and depends on the global light (horizontal line) that limits the intensities.

here as **MSAC-Multiple Uniform**.

We assume that the black patch are distributed on all scene depths, however the dark intensity of these pixels are not due to the effects of the medium, mainly the backscattering. Therefore, the strategy is to select the 1% of lowest pixels on the full range of the scene depth. The uniform selection [Roser et al., 2014] could not be adequate due to the exponential behavior of the attenuation coefficient. The uniform distribution tends to select the same number of points in small and large depths. Therefore, we select the pixels based on the natural distribution of the points. We sort the pixels based on their respective depth. For each set of 1% of the total of pixels in the sorted set, we select the lowest intensity pixel. This approach is named **MSAC-Multiple Sort**. This procedure is performed for each color channel. Both approaches are evaluated in Sec. 4.3.1 with similar performance. Thus, we adopted the **MSAC-Multiple Sort** by the aforementioned limitation related to the exponential behavior.

The new error function to be minimized for each color channel is defined as:

$$E = \sum \omega_1 (I^{t+1} - A(1 - e^{-\eta(d+\Delta d)})) + \omega_2 (I^t - A(1 - e^{-\eta d})) + \omega_3 \left(\frac{I^{t+1} - A}{I^t - A} - e^{-\eta(\Delta d)} \right), \tag{3.31}$$

where the weights ω_1, ω_2 and ω_3 define the importance of each term in the error function. It is worth to note that this function is valid only for black pixels, *i.e.* $J \approx 0$.

In the present work, $\omega_1 = 1.0$, $\omega_2 = 1.0$ and $\omega_3 = 0.5$. These values were obtained using experimental estimation.

This function is non-linear and complex to be optimized in terms of η. Furthermore, the number of outliers could be large due to the difficult to obtain depth and correspondence between pixels in participating medium. Roser et al. [2014] adopted a trust region approach [Coleman and Li, 1996] for minimizing the function. Differently from them, we adopted an optimization based on MSAC [Torr and Zisserman, 2000], as previously detailed. The error function is based on Eq. 3.31. The sample subset is composed by a single 3D point that projects two image pixels on the time t and $t+1$. The attenuation parameter for this point is obtained on using a weighted sum of the three constraints, *i.e.* Eq. 3.27 and 3.30. The method is repeated for each RGB color channel.

One key point is related to the estimation of Δd, further detailed in the Sec. 3.3. The depth maps may be obtained up to a single universal scale due to the use of monocular cameras. The same result is valid to the difference of depth maps. Therefore, the proposed algorithm allows us to recover the attenuation coefficient up to scale, as detailed in Eq. 3.32:

$$\begin{aligned} \eta' &= -\frac{\ln(I^{t+1} - A) - \ln(I^t - A)}{\lambda \Delta d}, \\ &= \frac{\eta}{\lambda}, \end{aligned} \tag{3.32}$$

where λ is the universal scale and η' is the attenuation coefficient up to scale. This estimated coefficient compensates the scale factor of the depth values. Thus, the model remains the same.

Chapter 4

Experiments and Results

The proposed methodology is divided in four procedures. Thus, several experimental results using simulated and real data are provided to validate each procedure. Data are obtained using two different approaches: embedding the degradation model inside a rendering software and applying the optical model in images from public datasets. We also evaluate in real sequences acquired in hazy and underwater scenes.

The next sections present the results obtained for each procedure of the methodology.

4.1 Transmission Prior

In this section, we provide results using UDCP in underwater environment. Firstly, we evaluate the UDCP capability to estimate the medium transmission (Sec. 4.1.1). We apply UDCP to restore images in Sec. 4.1.2. Differently from the complete methodology proposed in this work (Fig. 3.1), the UDCP does not require a calibrated camera thus increasing the applicability to a large set of images available on the internet.

4.1.1 Transmission Estimation

Three different transmission estimation methods based on the DCP prior are compared with the UDCP in this section. Standard DCP was applied in underwater images, as proposed by Chao and Wang [2010] and Chiang and Chen [2012]. MDCP [Gibson et al., 2012] was also applied to underwater images, but using refinement method as proposed by He et al. [2009]. To the best of our knowledge, we are the first work to demonstrate the MDCP's performance in underwater environment. We also evaluate the Bianco's prior (BP) [Carlevaris-Bianco et al., 2010]. These methods are detailed in Sec. 2.3.1.2.

Initially, we show results obtained by applying the methods on **real images** that we have captured, as well as images acquired by other authors (Fig. 4.1). The captured images by us were acquired in the Brazil's Southeast Coast by using the underwater vehicle Seabotix LBV300-5 equipped a color camera, at approximately 35km from the coast and has depths ranging from 12m to 20m. For these images, the global light is well balanced in all wavelengths because of the characteristics of the water and the shallow water depth, *i.e.* the red channel suffers small absorption rate; some images were acquired with On-Screen Display (OSD) as the one in Fig. 4.1u. Furthermore, the images contain part of the vehicle structure in the top left corner. Resolution of the captured images is 640×480 pixels.

For all methods, we adopted a local patch size is $\Omega = 15 \times 15$ pixels and $\omega = 0.95$, similar to He et al. [2009]. Finally, transmission maps were refined using a soft matting algorithm [Levin et al., 2008].

In the first column of Fig. 4.1, we show the original images with the global light, A, estimation. Yellow pixels are the 0.1% brightest pixels in the dark channel. Red circle represents the largest intensity pixel among these pixels. Orange circle is the estimation using the UDCP. Finally, purple circle is the estimation based on the BP, where the farthest point from the camera is automatically chosen.

Global light estimation requires a completely haze-opaque region on the image, which usually occurs in a region above the horizon in forward looking images. The method proposed by He et al. [2009] fails in the estimation as shown in Fig. 4.1f, being detected on a rock. Using the UDCP, the global light is detected in haze-opaque areas. Lastly, the estimation based on BP fails in the figs. 4.1a and 4.1k.

In Fig. 4.1, the refined transmissions t are shown in the second, third, fourth and fifth for standard DCP, UDCP, MDCP, and BP, respectively. Results are colored in order to improve the visualization, where hot colors, *i.e.* reddish colors, represent closer points and cold colors, *i.e.* bluish colors, represent distant points.

Notice that the standard DCP fails in the transmission estimation in some cases. This can be mainly seen in figs. 4.1b and 4.1l. This failure is due to the low intensity of the red channel. Thus, the standard DCP is mostly composed of the red channel information. Hence, both works [Chao and Wang, 2010; Chiang and Chen, 2012] employ standard DCP that may fail in this typical situation. In Fig. 4.1g, the transmission seems overestimated because of the bad estimation of the global light.

UDCP and MDCP results are presented in third and fourth columns in Fig. 4.1. They are similar as it can be seen in figs. 4.1c and 4.1d. Notice that the methods cor-

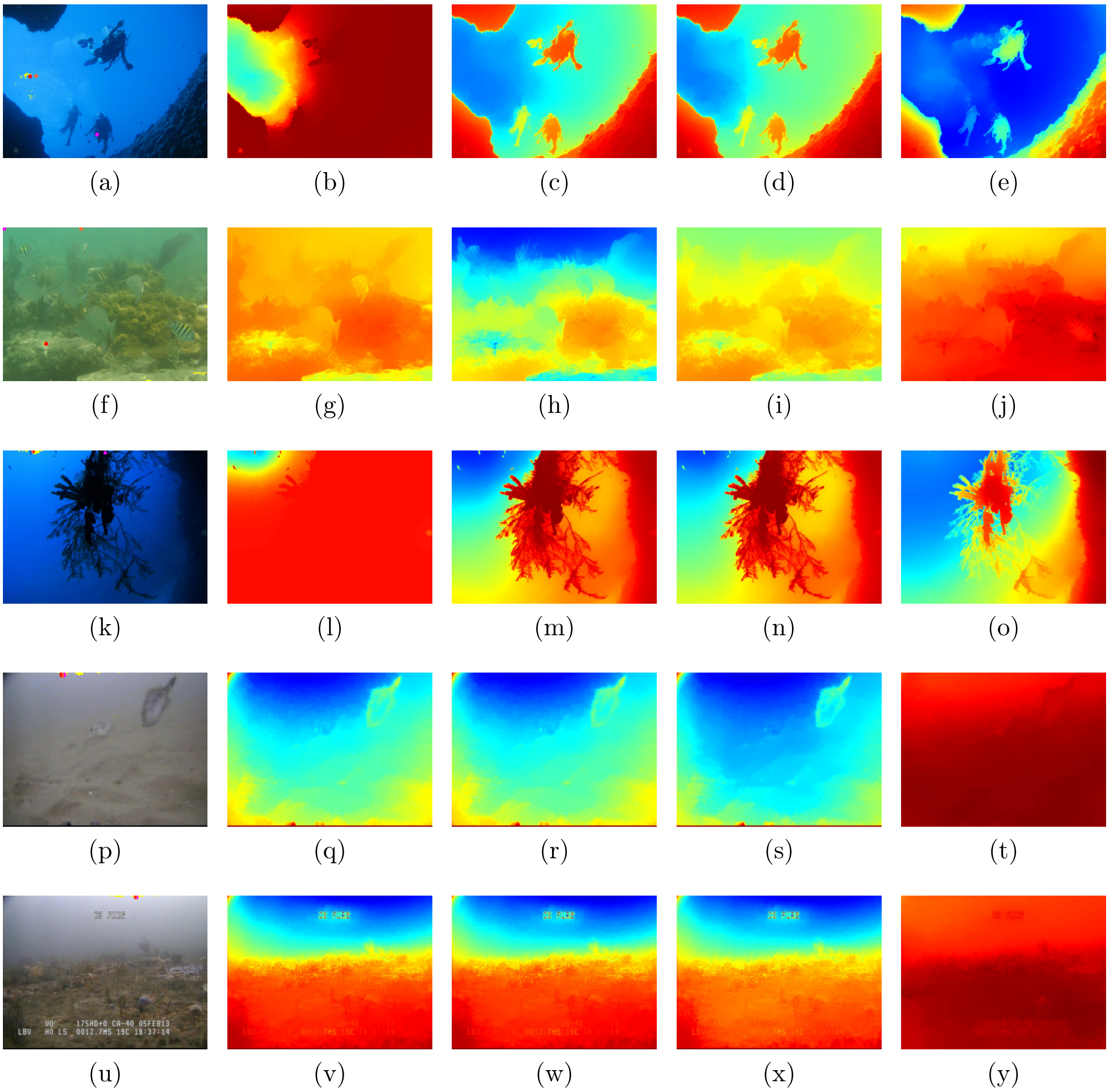


Figure 4.1. Evaluation of the transmission estimation methods: First column - real underwater images and global light estimation, using DCP, in red, UDCP, in orange, and the BP, in purple. We also provided colorized refined transmission t using standard DCP, in the second column, the UDCP, in the third column, the MDCP, in the fourth column, and the BP, in the last column. Credits of the images: a) courtesy of Nick Roberts; f) Additional Materials of Ancuti et al. [2012]; k) freeunderwaterimages.com.

rectly identify the distance relation from the camera to the divers and rocks. Moreover, figs. 4.1m and 4.1n are also similar. Similarities occur because of the low intensities of the red channel and the similar intensities in the green and blue channel, thus the median and the minimum of the blue and green channels are similar. The largest difference between UDCP and MDCP is due to the red channel intensities, as shown in figs. 4.1h and 4.1i. In the last two rows, there is a small difference in the low intensity values of the transmission.

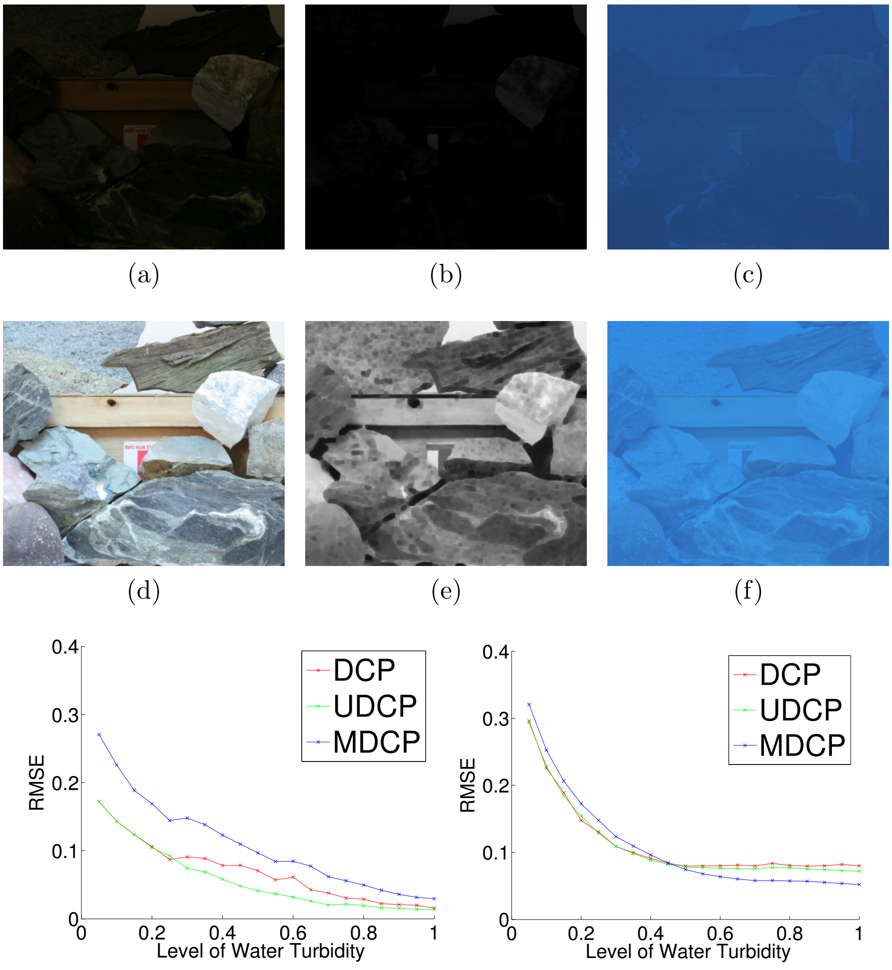
The last column of Fig. 4.1 shows the results obtained by BP. Transmission maps are well estimated in figs. 4.1e and 4.1o. However, the transmission of the scene is overestimated in Fig. 4.1j. The last results are the worst obtained by the BP, where the method fails to estimate the transmission (see figs. 4.1y and 4.1t).

Quantitative results were obtained using images after applying the optical model in images from public datasets. We adopted images from the Middlebury Stereo Dataset - Rocks1 [Scharstein and Szeliski, 2002] under different illumination conditions, where the disparity maps are available. The optical model (Eq. 2.5) was applied in both images in order to simulate the underwater effects. The attenuation coefficient η and the global light A was defined as proposed in Mobley [1994] and Schechner and Karpel [2005], respectively. It allows us to simulate an oceanic water under variable chlorophyll concentration, defined by a constant C [Mobley, 1994]. Depth map was obtained using previous knowledge of the stereo baseline and focal distance. We multiplied these values by a 1.5 factor to increase the global distance and depth variation. Therefore, scene is far from the camera $[\approx 4\text{m}; \approx 6.5\text{m}]$. We define typical values $C = [0.05; 1.0]\text{mg.m}^{-3}$ in simulation [Mobley, 1994]. Ground truth transmission was selected as the green channel transmission because of the small global Root-Mean-Square Error (RMSE) due to the fact that the methods based on DCP are only able to estimate a single transmission. RMSE was applied as metrics to evaluate the quality of evaluated methods.

Fig. 4.2 shows the results using simulation. Figs. 4.2a and 4.2d show the images obtained from the dataset. Figs. 4.2b and 4.2e show DCP results using all color channels, where one may see that the assumption of darkness is almost true for the first case, but invalid for the second case. Figs. 4.2c and 4.2f illustrate the images using $C = 0.5\text{mg.m}^{-3}$. Both graphics, in figs. 4.2g and 4.2h, show the quantitative results. The results obtained using BP are discarded to improve the visualization. BP obtains $RMSE \approx [0.7; 0.9]$ for both cases.

Fig. 4.2g shows the results for a low illumination image (Fig.4.2a). In this case, one may see that the UDCP obtains the smaller error in almost all turbidity level. The largest errors were obtained for smaller values of C because slightly degraded images provide little information about medium transmission. When the level of turbidity increases, the method is able to estimate the transmission until a limit where the information about the scene is lost. In this case, the transmission is almost null for the ground truth and estimation. Therefore, the RMSE is small.

Similarly, Fig. 4.2h shows the results for a high illumination image (Fig. 4.2d).



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 4.2. Quantitative evaluation of the transmission estimation methods: First column - Images obtained from Middlebury Stereo Dataset under different illumination conditions; DCP of the original images, in the second column; example of simulated underwater images using $C = 0.5\text{mg.m}^{-3}$, in the third column; quantitative results using RMSE, in the last row.

In this case the premises of DCP fail, although all DCP-based methods are able to provide a good estimation. Notice that the overall RMSE is larger than previous one. It is due to the main premise of the DCP-based method, *i.e.* $J \approx 0$, is false in this high illumination condition. MDCP obtains the smaller error for large turbidity level. For smaller levels, the DCP and the UDCP obtain similar results.

4.1.2 Image Restoration using UDCP

We also obtained results for underwater image restoration using UDCP in **real images**. The estimated medium transmission is applied in the Eq. 3.25 to obtain the restored image. As described in Sec. 4.1.1, we compared the UDCP with the results obtained by DCP [He et al., 2009], MDCP [Gibson et al., 2012], and BP [Carlevaris-Bianco et al., 2010].

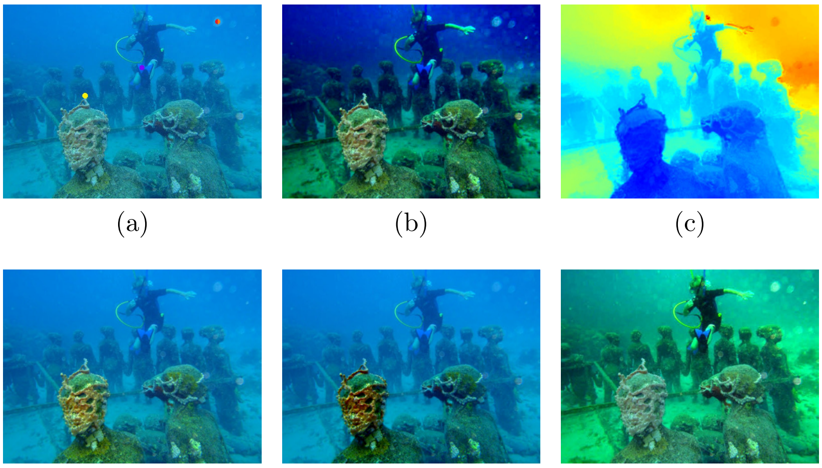
Our evaluation is based on qualitative and quantitative analysis on **real images**. Figs. 4.3 and 4.4 show the qualitative results for two sample images from the underwater dataset collected from the internet (IOW dataset detailed in Sec. 3.1.3). Fig. 4.5 shows the sample images from three underwater videos that we captured. We acquired these videos in a coral reef at the Brazilian Northeast Coastal area at the depth of approximately 10m. They are composed of 150, 138 and 610 frames, and the sample images of these videos are shown in figs. 4.5a, 4.5b and 4.5c, respectively.

In the quantitative evaluation, we adopted the restoration score τ proposed by Hautière et al. [2008a] to analyze their method for weather-degraded images. Three different indexes are defined in the τ metric: e, $\overline{r}$ and s. The value of e evaluates the ability of a method to restore edges, which were not visible in the degraded image, but are visible in the restored image. The value of $\overline{r}$ measures the quality of contrast restoration; a similar technique was adopted by Schechner and Karpel [2005] to evaluate restoration in an underwater medium. Finally, the value of s is obtained from the proportion of the number of pixels which are saturated (black or white) after applying the restoration method but were not before. These three indexes allow us to estimate an empirical restoration score $\tau = e + \overline{r} + (1 - s)$ [Hautière et al., 2008a], where larger values mean better restoration. Table 4.1 shows the obtained results.

One example of these experiments is depicted in Fig. 4.3, which shows the original image, Fig. 4.3a, the restored image, Fig. 4.3b, and the colorized depth maps, Fig. 4.3c, obtained using UDCP approach. We colorized the depth maps to aid the visualization, where bluish colors represent closer points and reddish colors, represent points that are further away. The colors are the inverse of the medium transmission due to the behavior of the exponential term, *e.g.* large depth obtains medium transmission near zero.

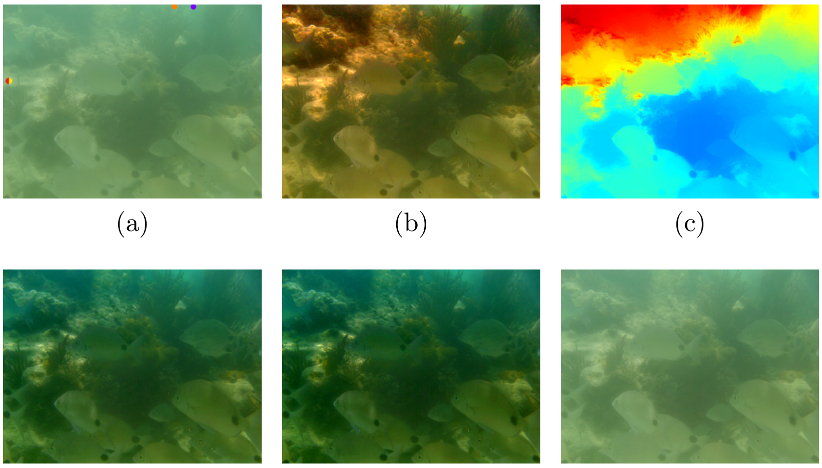
Figs. 4.3 and 4.4 also show the results obtained by applying the methods proposed by other authors, *i.e.* DCP, MDCP, and BP. We show the underwater images with the back scattering light estimation in figs. 4.3a and 4.4a. The estimation of the backscattering constant obtained by UDCP seems to be the most plausible, *i.e.* near the farthest point in the image. The other methods fail in the estimation in at least one of the images. They identify the backscattering light in bright surfaces of the scene instead of the farthest point.

The restored images by UDCP, figs. 4.3b and 4.4b, show that there was an improvement as far as contrast and color fidelity are concerned. Table 4.1 shows the


(a) (b) (c)

(d) (e) (f)

Figure 4.3. Single image restoration of a sample image from the IOW dataset (Sec. 3.1.3): (a) Underwater image with regions where the backscattering constant was estimated, using UDCP (orange patch), DCP (red patch), MDCP (yellow patch), and the BP (purple patch). Restored images using UDCP (b), DCP (d), MDCP (e), and BP (f). Colorized depth maps obtained using UDCP is shown in (c).


(a) (b) (c)

(d) (e) (f)

Figure 4.4. Single image restoration of another sample image from the IOW dataset (Sec. 3.1.3): (a) Underwater image with regions where the backscattering constant was estimated, using UDCP (orange patch), DCP (red patch), MDCP (yellow patch), and the BP (purple patch). Restored images using UDCP (b), DCP (d), MDCP (e), and BP (f). Colorized depth maps obtained using UDCP is shown in (c).

estimation of the UDCP obtaining the larger values in terms of τ metric. In Fig. 4.3, the improvement in the estimation of the ocean floor of the scene is noticed for

Table 4.1. Quantitative evaluation of the single image underwater restoration methods using the τ metric [Hautière et al., 2008a] and its components e and $\bar{r}$. We show the results for the sample images in figs. 4.3, 4.4 and 4.5, and the average of the extended dataset (samples images in figs. 4.3 and 4.4 - detailed in Sec. 3.1.3) and the videos of the Fig. 4.5. The best results are highlighted in bold letters.

	UDCP			DCP			MDCP			BP		
	e	$\bar{r}$	τ	e	$\bar{r}$	τ	e	$\bar{r}$	τ	e	$\bar{r}$	τ
Fig. 4.3a	**1.56**	**1.96**	**4.52**	0.06	1.16	2.22	0.04	1.10	2.14	0.75	1.75	3.50
Fig. 4.4a	**56.03**	**3.40**	**60.43**	38.42	2.02	41.44	53.92	2.05	56.97	0.88	1.13	3.01
Average	**1.28**	1.42	**3.68**	0.60	1.18	2.77	0.67	1.09	2.75	0.43	**1.62**	3.01
Fig. 4.5a	**699.82**	4.79	**705.61**	635.29	4.65	640.95	635.57	**5.15**	641.72	0.79	1.06	2.85
Average	**3455.2**	3.99	**3460.2**	1495.0	3.93	1500.0	1937.6	**4.17**	1942.8	0.73	1.06	2.79
Fig. 4.5b	**2.49**	**2.33**	**5.82**	2.45	2.04	5.49	2.37	1.43	4.80	0.15	1.06	2.21
Average	**29.07**	**1.95**	**32.02**	20.23	1.86	23.09	21.39	1.70	24.09	0.24	1.06	2.30
Fig. 4.5c	**5.37**	**4.44**	**10.81**	4.98	3.58	9.56	5.31	2.67	8.98	0.18	1.05	2.23
Average	**85.46**	3.91	**90.37**	64.11	**4.44**	69.56	67.80	4.31	73.11	0.30	1.06	2.36

the image restored using UDCP. The improvement provided by DCP and MDCP is imperceptible because both methods are not able to recover the depth map in a correct way.

The UDCP method obtains a significant improvement, as shown in Fig. 4.4. Conversely, the BP, Fig. 4.4f, presented limited results in terms of restoration. The values in Table 4.1 also confirm this fact. We can explain it because the BP underestimates the attenuation coefficient, limiting the quality of the map. This is due to the behavior of the red channel, which is not completely absorbed. The results obtained by standard DCP, Fig. 4.4d, and MDCP, Fig. 4.4e, are also related to this fact, since both methods are able to provide good results for restoration.

Table 4.1 also shows the results for the videos illustrated by sample images in Fig. 4.5. UDCP is better than the others in τ metric, especially due to its advantage in the e metric [Hautière et al., 2008a]. UDCP also achieves good contrast restoration, as showed by the $\bar{r}$ metric. However, the contrast is slight better for BP in the extended dataset. Similar behavior occurs in the average of the videos of the Fig. 4.5 for DCP and MDCP. One can see that the edge restoration metric e presents large values for video associated to Fig. 4.5a. It is because the number of edges in the underwater image is small, assuming the parameters adopted by Hautière et al. [2008a]. As can be seen, the increasing provided by the restoration is large producing large values of the e metric.

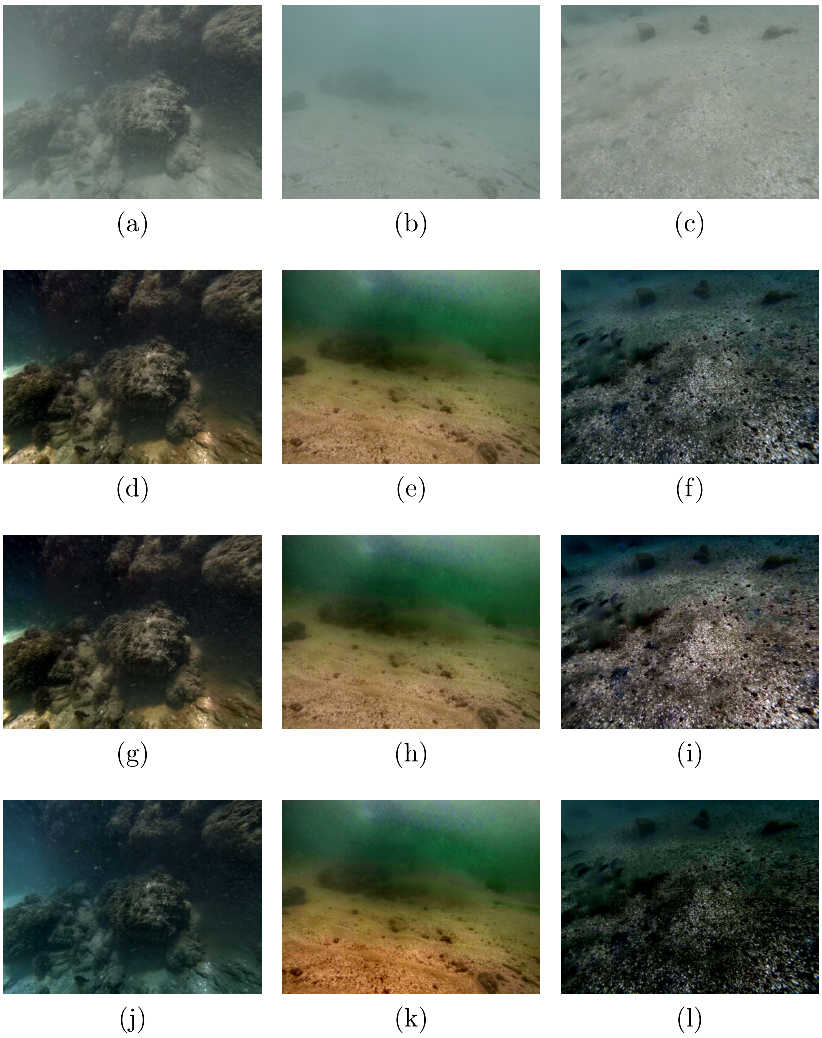


Figure 4.5. Single image restoration of three underwater videos acquired in Brazilian Northeast Coastal area. The first row shows three sample images for each video. The restoration results for these sample images obtained by the standard DCP, UDCP and MDCP are shown in the second, third and fourth rows, respectively.

4.2 Optical Flow

We performed several experiments to evaluate the robustness of the GOFM using simulated and real image sequences. As described in Sec. 3.2.4, we instantiate two different approaches to estimate optical flow using GOFM. The first aims to compare the proposed approach with classical approaches and the state-of-art in terms of underwater optical flow. The second one is more accurate and based on modern implementation.

Quantitative evaluation follows the same approach proposed by Baker et al. [2011], where the absolute endpoint error is estimated along with the average angular error proposed by Barron et al. [1994]. Endpoint error is estimated in pixels while the angular error is estimated in degrees.

4.2.1 Comparative Evaluation

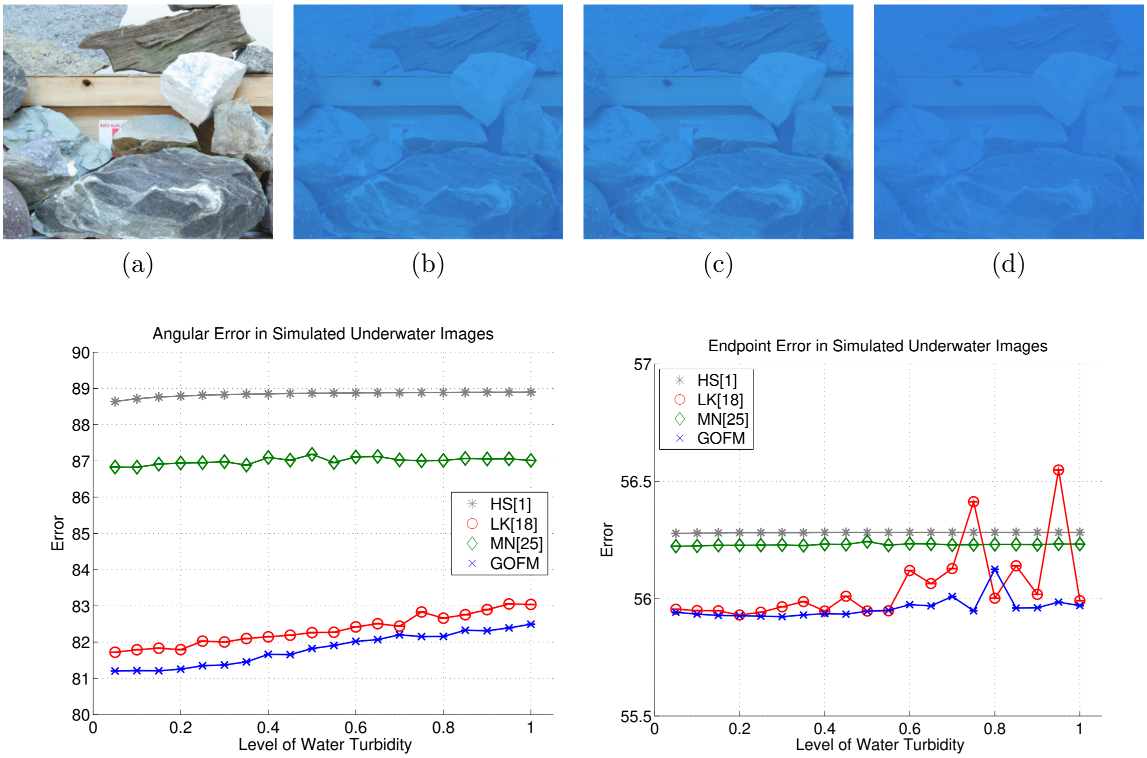
The first approach is developed in a simpler way to make an evaluation study. Considering the model of the Eq. 3.7, both global and local smoothness constraints approaches are evaluated using the classical Horn-Schunk (HS) [Horn and Schunck, 1981] and the Lucas-Kanade (LK) [Lucas and Kanade, 1981] methods, respectively. The method proposed by Madjidi and Negahdaripour (MN)[Negahdaripour and Madjidi, 2003; Madjidi and Negahdaripour, 2006] was chosen to compare because of its relevance in optical flow estimation in participating media. For the sake of a fair comparison, all methods are implemented using the simplest model without optimization and multiple scale representation, instead of the best possible performance.

Ground truth estimation was obtained using the method proposed by Sun et al. [2014] in the non-participating media images, when available. This method is one of the best ranked in the Middlebury Optical Flow Dataset [Baker et al., 2011]. Additionally, a manual refinement was performed in order to improve the ground truth.

Same parameters are adopted for all methods to compare them, *i.e.* Gaussian kernel with $\sigma = 1.5$ for the simulated images or $\sigma = 2.5$ for the real images, and median filter with kernel size of 5×5 pixels. These parameters are defined using the evaluation made by Barron et al. [1994] and Sun et al. [2010]. In the case of Horn and Schunck [1981], we adopted the parameters values as defined by the authors, *e.g.* $\alpha^2 = 1$. For the other methods, we adopted a local window of 11×11 pixels [Madjidi and Negahdaripour, 2006].

Simulated results were obtained using a pair of images from the Middlebury Stereo Dataset - Rocks1 [Scharstein and Szeliski, 2002], where the disparity maps are available. Similar to the qualitative results of Sec. 4.1, we applied the optical model (Eq. 2.5) in both images in order to simulate the underwater effects.

Fig. 4.6 shows the results obtained through the simulated sequence. Figs. 4.2d and 4.6a show the image from the Middlebury Stereo Dataset - Rocks1 [Scharstein and Szeliski, 2002] before we applied the underwater effect. Three simulated images are shown in figs. 4.6b-4.6d, where the water turbidity level was defined by $C = 0.05\text{mg.m}^{-3}$, $C = 0.5\text{mg.m}^{-3}$, and $C = 1\text{mg.m}^{-3}$, respectively. A quantitative evaluation between all four methods is performed on the figs. 4.6e and 4.6f. Average angular error shows the advantage of the proposed methodology in comparison to the HS [Horn and Schunck, 1981] and the MN [Madjidi and Negahdaripour, 2006],



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

(e) (f)

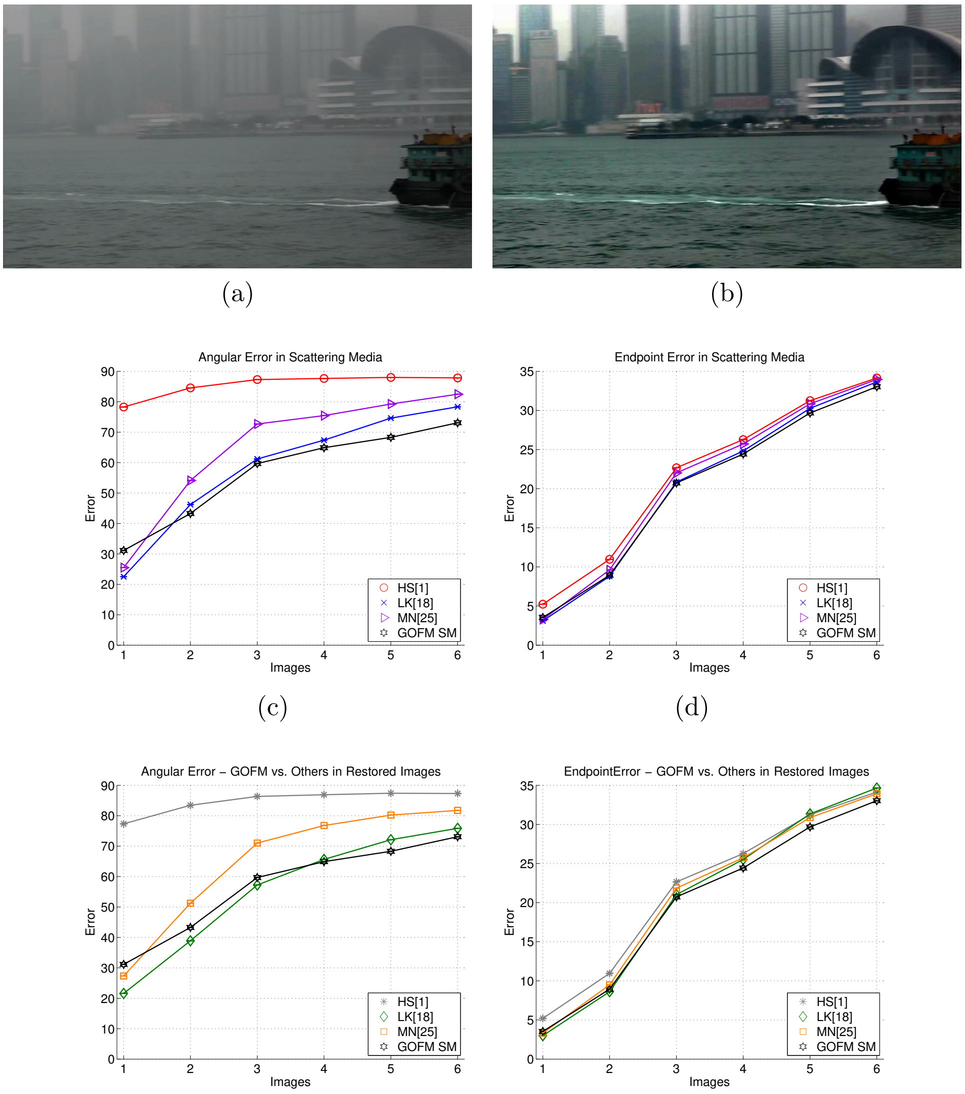
Figure 4.6. Quantitative results in simulated sequence: (a) Original Image from Middlebury Stereo Dataset - Rocks1 [Scharstein and Szeliski, 2002]; (b-d) Simulated images where the turbidity level are $C = 0.05\text{mg.m}^{-3}$, $C = 0.5\text{mg.m}^{-3}$, and $C = 1\text{mg.m}^{-3}$, respectively; and (e,f) Average angular, in degree, and endpoint error, in pixels, for all evaluated methods, where GOFM obtains the smallest error.

although all methods increase the error proportionally to the turbidity level. Similar to the angular error, the endpoint error provided by LK [Lucas and Kanade, 1981] and the proposed method (GOFM) are comparable, with a small advantage to the latter. All methods obtain results that are similar in terms of endpoint error. Because of the small depth variation in the image sequence, the improvement of the proposed method becomes limited. Nevertheless, GOFM shows better results than the others, even in this situation. Furthermore, the large movement between images affects the performance of all methods because a single scale estimation is employed.

We also employed **real image sequences** to evaluate the first approach. Firstly, an outdoor sequence under fog conditions acquired by He et al. [2011] was used for comparing our method with several standard optical flow approaches (a frame of the sequence is shown in Fig. 4.7). Initially, we randomly selected a frame from the sequence, and then we estimate the ground truth between this restored frame and a restored sequence of six consecutive frames by using the previously detailed methodology. Therefore, we computed the optical flow using HS [Horn and Schunck, 1981], LK [Lucas and Kanade, 1981] and MN [Madjidi and Negahdaripour, 2006] for the sequences, original and restored by He et al. [2009]. Figs. 4.7a and 4.7b show a sample from the original image sequence and the restored one. Notice that the restoration were obtained by He et al. [2009], because our implementation of their method obtains worse results. It is due to the authors restored the full image, *i.e.* both in resolution and size, and without video compression. However, they make available only a smaller version.

Figs. 4.7c and 4.7d show the average angular and the endpoint errors using the original image sequence. For a small displacement between images, the results are similar with a small advantage to LK [Lucas and Kanade, 1981]. However, GOFM becomes dominant and provides a smaller error when the displacement increases (*e.g.* fast motions), mainly in terms of angular error. One may see that the results obtained in restored image sequence by HS [Horn and Schunck, 1981], LK [Lucas and Kanade, 1981] and MN [Madjidi and Negahdaripour, 2006] are worse than GOFM, even though it is applied in the scattering image (figs. 4.7e and 4.7f). Error for restored sequence is slightly smaller than the participating media sequence. In terms of endpoint error, GOFM presents the smallest error in all sequences.

Underwater **real images** are also obtained as shown in Fig. 4.8. A sample image from the dataset is shown in Fig. 4.8a. Optical flow fields are estimated using LK [Lucas and Kanade, 1981], MN [Madjidi and Negahdaripour, 2006] and HS [Horn and Schunck, 1981], figs. 4.8b, 4.8c and 4.8d, respectively. Consecutive frames are selected from the image sequence for which the movement is approximately perpendicular to image plane with a small counterclockwise rotation in this axis. This movement makes easier to understand the result. Optical flow estimation provided by our method is shown in figs. 4.8e and 4.8f, where the model is applied to the intensity image and the color image. LK [Lucas and Kanade, 1981] and MN [Madjidi and Negahdaripour, 2006] are prone to error in the top region of the image because of the presence of haze in the sequence. However, HS [Horn and Schunck, 1981] and our method produce more realistic results. Notice that HS fails in the bottom-right region, where some vectors point to the right, when the correct would be to the left. Color information improves the optical flow density, as it can be perceived in figs. 4.8e and 4.8f. Moreover, the fish movement is correctly estimated by our method (red box in figs. 4.8e and 4.8f). HS [Horn and Schunck, 1981] is unable to estimate it.



(a) (b)

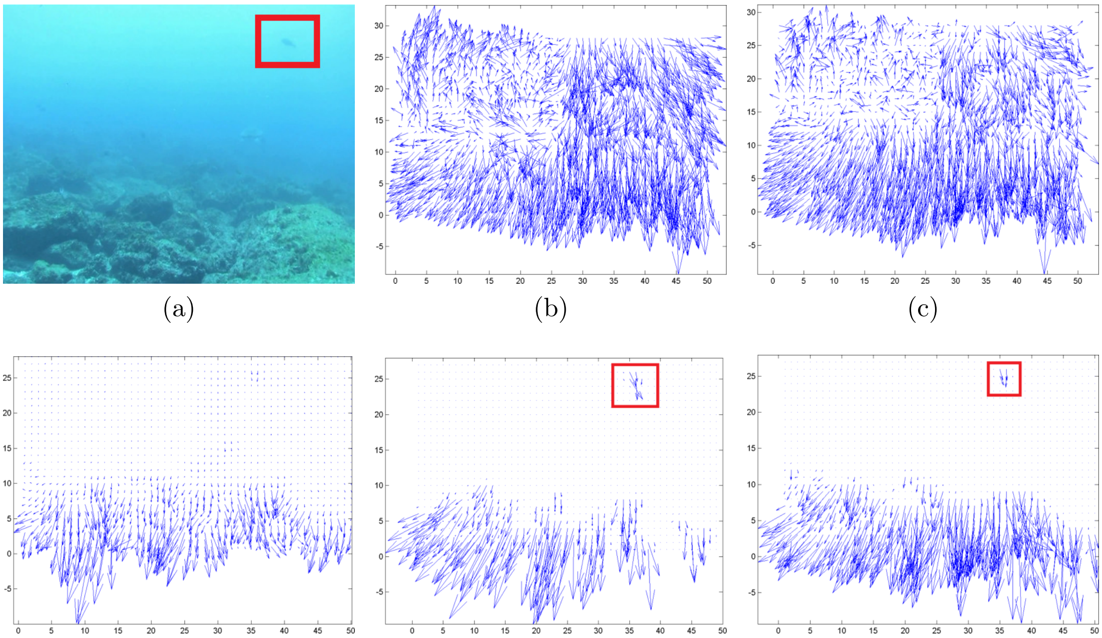
(c) (d)

(e) (f)

Figure 4.7. Experiments with real image sequence under fog condition obtained from He et al. [2011]: (a,b) Average angular, in degree, and endpoint, in pixels, errors for all evaluated methods employing the foggy image sequence; and (c,d) Average angular and endpoint errors, where GOFM is applied to the foggy image sequence, *i.e.* scattering media (SM), and the other methods are applied to the restored version of the image sequence. Horizontal axis represents the number of the frame used for estimating the optical flow, where the reference frame is always the image 0; while the vertical axis represents the estimated error.

4.2.2 Accuracy Evaluation

Differently from previous section, the focus here is evaluating the accuracy of GOFM by using a modern implementation. Hence, it is compared with the Classic-NL (CL-NL) and the modern implementation of Horn-Schunk method (HS), both proposed by



(a) (b) (c)

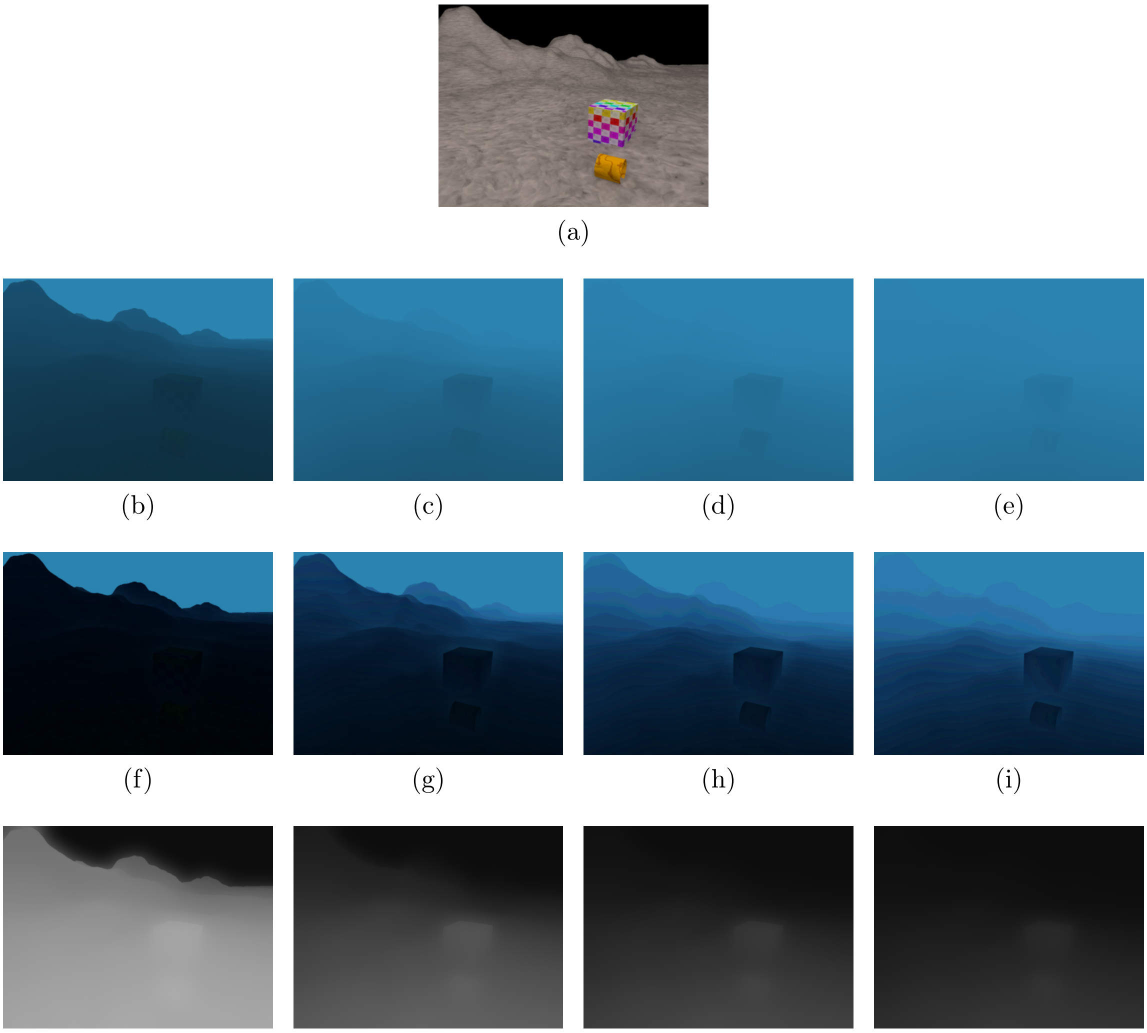
(d) (e) (f)

Figure 4.8. Real image sequence in underwater scene with images acquired in Guanacaste Coast Pacific - Costa Rica [Van der Berg, 2012]: (a) Sample image with a fish marked by a red box; (b,c,d) Optical flow fields obtained using LK [Lucas and Kanade, 1981], MN [Madjidi and Negahdaripour, 2006] and HS [Horn and Schunck, 1981], respectively; and (e,f) Optical flow fields obtained using GOFM applied on the intensity image and the color image, respectively. In these fields, one may see that the proposed method is able to estimate the fish's movement, even under strong haze.

Sun et al. [2014]. Results for both methods are obtained using degraded images, *i.e.* in participating medium, and the restored version by UDCP (R). In fact, this section evaluates the second approach.

Simulated underwater images were created using a modified version of the open-source raytracer POVRay 3.2 [Persistence of Vision Pty. Ltd., 2004], where an underwater scene is built inspired by the work of Sedlazeck and Koch [2011]. Optical model is employed in the rendering process, and it is built upon three coefficients: the absorption coefficient $a(\lambda)$, the scattering coefficient $\beta(\lambda)$ and the global light A. Typical discrete wavelength, λ, values are chosen for the RGB colors in the rendering process: 700nm for red, 530nm for green and 470nm for blue.

Similar to the simulated images from Middlebury Stereo Dataset [Scharstein and Szeliski, 2002], described in Sec. 4.2.1, the other two coefficients are calculated using typical values for oceanic waters [Mobley, 1994]. Both coefficients depend on the wavelength λ and the chlorophyll concentration C. Ground truth is obtained by applying the method Classic-NL [Sun et al., 2014] on the images without the effects of the participating medium.



(a)

(b) (c) (d) (e)

(f) (g) (h) (i)

(j) (k) (l) (m)

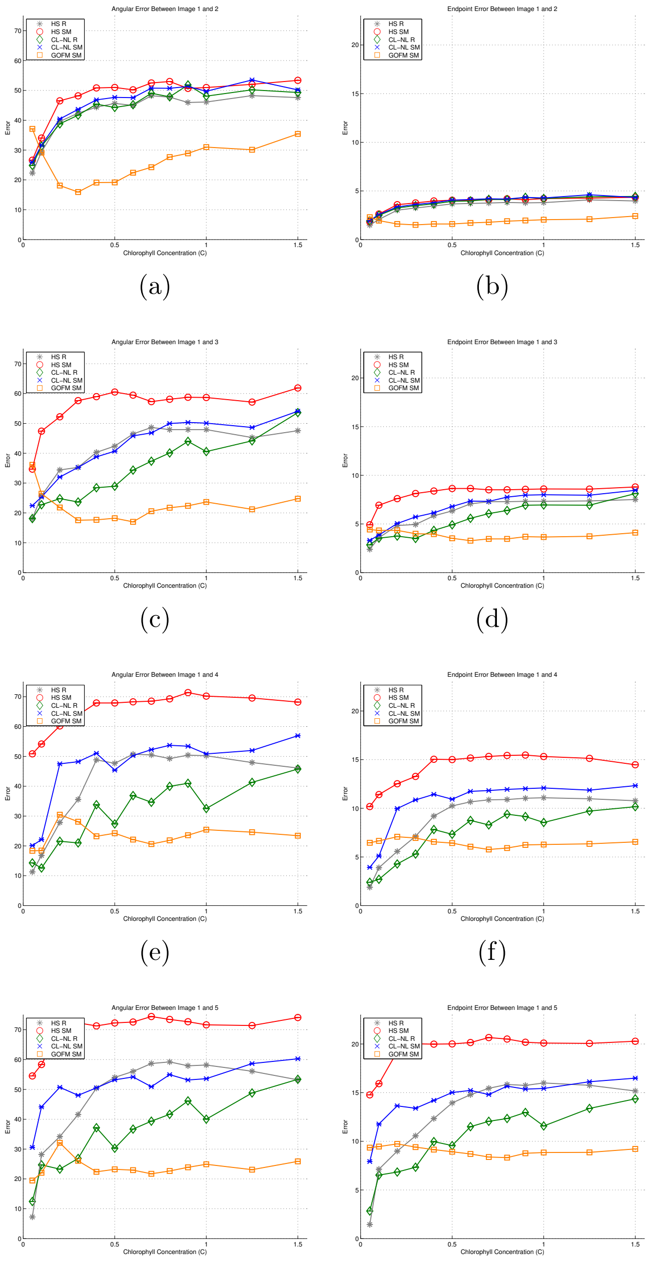
Figure 4.9. Simulated sequence obtained using a raytracer simulator: (a) Frame obtained without the effects of the participating medium, (b-e) Frame from underwater scene with different medium parameters, where the chlorophyll concentrations C are equal to 0.05mg.m^{-3}, 0.5mg.m^{-3}, 1mg.m^{-3}, and 1.5mg.m^{-3}, respectively. (f-i) Restored images using UDCP approach for each degradation level and their estimated medium transmission (j-m), respectively.

Fig. 4.9 shows a frame from the simulation (Fig. 4.9a) which is composed of five frames. During the simulation, we vary the chlorophyll concentration constant in a range of typical values: $[0.05, 1.5]\text{mg.m}^{-3}$ [Mobley, 1994]. Degraded images are shown in figs. 4.9b-4.9e, where C is defined as 0.05mg.m^{-3}, 0.5mg.m^{-3}, 1mg.m^{-3}, and 1.5mg.m^{-3}, respectively. For each degraded image, the restored images using UDCP are presented (Figs. 4.9f-4.9i). One can see that UDCP produces improvement in terms of visibility, but lacks in color fidelity. Restored image obtained applying UDCP for the smallest simulated chlorophyll concentration presented small improvement. It is because UDCP obtains a single medium transmission which reduces the scattering but does not restore the colors. Scene disappears for larger values of chlorophyll concentration, thus becoming impossible to restore. Limits of restoration is the focus of the study conduced by Treibitz and Schechner [2009c].

Fig. 4.10 shows average angular and endpoint errors for each method using a simulated image sequence. Results are presented by increasing the baseline, *i.e.* simulating several speeds of camera's movements. The first frame is used as reference and the optical flow is estimated in relation to the subsequent frames. Average angular error and endpoint error for the optical flow estimated between the first and second frames are presented in figs. 4.10a and 4.10b. Because of the small amount of information provided by the transmission map in the smallest level of chlorophyll concentration, GOFM presents slightly worse results. However, it outperforms the other methods when evaluated in higher level of turbidity, mainly in terms of angular error. Similar behavior is obtained for a larger baseline (Figs. 4.10c-4.10h). Notice that GOFM is less sensible to the increasing of level of chlorophyll concentration, presenting an approximately linear behavior in endpoint error.

Similarly to Sec. 4.2.1, we also tested the methods using **real image** sequences. They are two outdoor sequences under fog conditions acquired by He et al. [2011]. They were used for comparing our method with state-of-art optical flow techniques. Initially, we randomly selected a frame from the complete sequence, and then we estimate the optical flow between this selected frame and the next ten frames. Optical flow estimated by Classic-NL [Sun et al., 2014] using the restoration provided by He et al. [2009] is our ground truth. Furthermore, we computed the optical flow using GOFM, modern Horn-Schunk (HS) [Sun et al., 2014], and Classic-NL (CL-NL)[Sun et al., 2014] for the original and restored frames using our implementation of DCP. As previously reported, the restored results by He et al. [2009] are better than ours because of the knowledge about the full image, *i.e.* both in resolution and size, and without video compression.

Fig. 4.11 shows results for the same sequence presented in Fig. 4.7, but now we are using the second proposed approach, *i.e* the accurate implementation of GOFM. Notice that the restoration using UDCP produces some halos (Fig. 4.11c). Average angular and endpoint errors for the optical flow estimated between frame one and next frames are presented in figs. 4.11d and 4.11e, respectively. Endpoint error is similar to the presented by Classic-NL, both using restored (R) and degraded images (SM). However, the smallest average angular error are presented by GOFM. Results obtained from modern HS [Sun et al., 2014] are worse than other methods. Furthermore, they



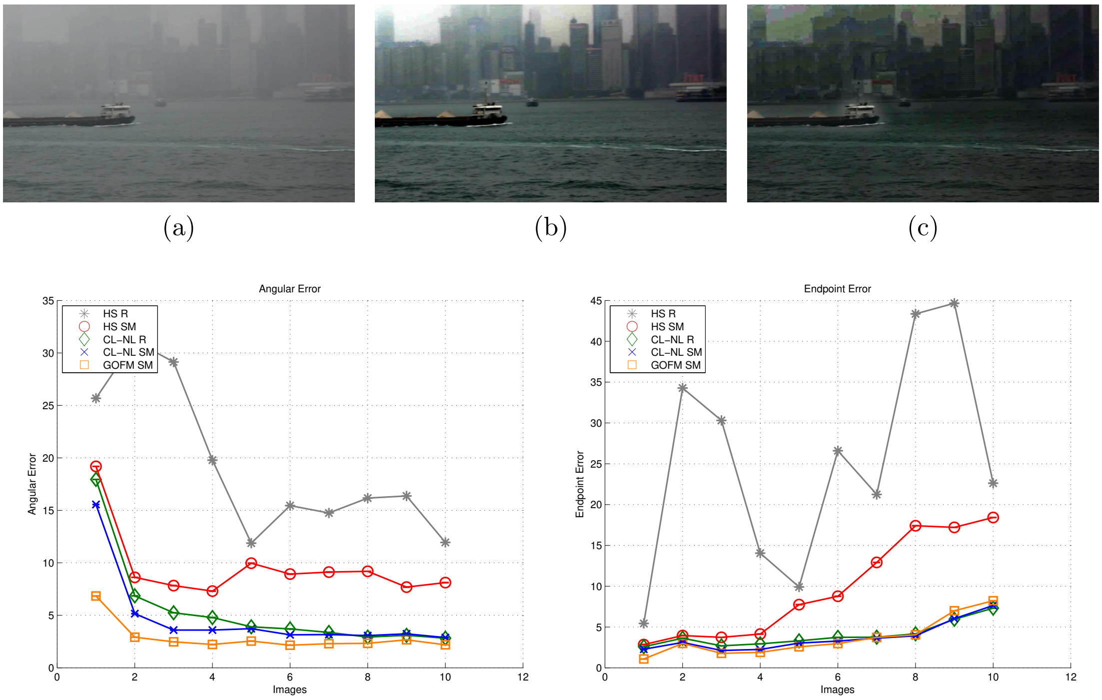
(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 4.10. Quantitative results in simulated sequence obtained using a ray-tracer simulator by increasing the baseline in each row, *i.e.* optical flow between frames one and two (a-b), frames one and three (c-d), etc. Average angular and the endpoint errors are presented in each column, respectively. GOFM is compared with the Classic-NL (CL-NL) and the modern implementation of Horn-Schunk (HS), both proposed by Sun et al. [2014]. These methods are applied in scattering medium images (SM), and the restored version obtained by UDCP (R). GOFM outperforms other methods, mainly for higher chlorophyll concentration.



(a) (b) (c)

(d) (e)

Figure 4.11. Optical flow evaluation using real image sequence in a hazy scene obtained by He et al. [2011]: (a) Sample frame of the sequence; (b,c) restored image of (a), provided by He et al. [2011] and by our implementation. Average angular error (d), in degree, and endpoint error (e), in pixels, are presented. GOFM is compared with other two modern methodology: Classic-NL (CL-NL) and the modern implementation of Horn-Schunk method [Horn and Schunck, 1981] (HS), both proposed by Sun et al. [2014].

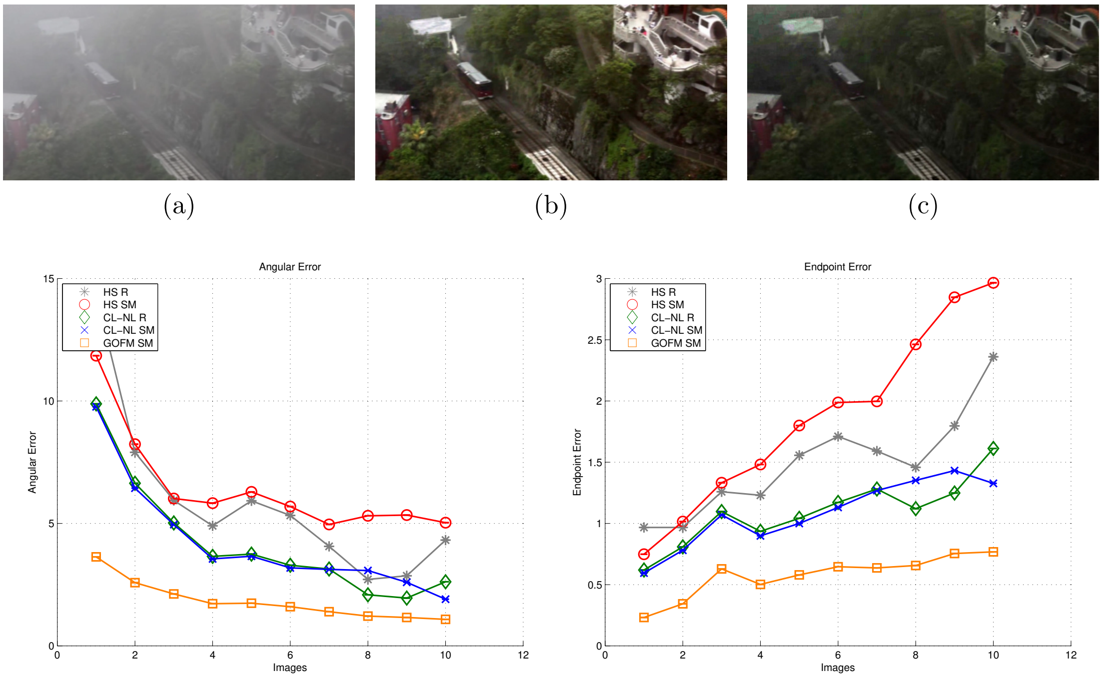
present a higher variance when increasing the baseline.

Fig. 4.12 shows the results for another sequence provided by He et al. [2011]. Average angular and the endpoint errors are presented in figs. 4.12d and 4.12e. Both errors obtained from GOFM are significantly smaller than the other methods. Results obtained from modern HS [Sun et al., 2014] remain the worst.

4.3 Parameters Estimation and Restoration

Finally, we obtained results for parameters estimation, restoration and structure-from-motion procedures. They are obtained using **simulated** and **real** datasets, all of them with qualitative and quantitative evaluation.

We compared our approach with our proposed method for single image, called UDCP, (Sec. 3.1.2) and two classical enhancement techniques: histogram equaliza-



(a) (b) (c)

(d) (e)

Figure 4.12. Optical flow evaluation using real image sequence in another hazy scene obtained by He et al. [2011]: (a) Sample frame of the sequence; (b,c) restored image of (a), provided by He et al. [2011] and by our implementation. Average angular error (d), in degree, and endpoint error (e), in pixels, are presented. GOFM is compared with other two modern methodology: Classic-NL (CL-NL) and the modern implementation of Horn-Schunk method [Horn and Schunck, 1981] (HS), both proposed by Sun et al. [2014].

tion [Szeliski, 2010] and Contrast-Limited Adaptive Histogram Equalization (CLAHE) [Zuiderveld, 1994].

Quantitative results are obtained using the metric proposed by Hautière et al. [2008a], similarly to Sec. 4.1.2. This metric was developed for weather-degraded images, but it was also applied to underwater images due to their similarities. It defines three different indexes: e, $\bar{r}$ and s. These three indexes allow us to estimate a restoration score $\tau = e + \bar{r} + 1 - s$ [Hautière et al., 2008a], where larger values mean better restoration.

We also obtained quantitative results by matching Scale-Invariant Feature Transform (SIFT) [Lowe, 2004] descriptors. Thus, evaluating the ability of the descriptor to identify and match features from a raw and restored image. This metric was also adopted in Barros [2010] and Ancuti et al. [2012] to evaluate restoration methods. For each pair of images, we show the number of keypoints detected in both images and the number of correct matches. This number is computed applying the approach proposed by Lowe [2004], *i.e.* the best candidate match for each keypoint is found by identifying its nearest neighbor in the previous image. Lowe [2004] removes outliers by comparing the distance of the closest neighbor to that of the second-closest neighbor. We defined a more restrictive ratio value than employed by Lowe [2004] in order to avoid outliers, thus providing a more reliable value to evaluate the restoration methods. We use the ratio of closest to second-closest neighbors of each keypoint equals to 0.4, while the implementation of Lowe [2004] uses 0.6.

For the simulated images, we also show the quantitative results obtained by using the Image Quality Assessment (IQA) measures where three state-of-the-art metrics are employed [Zhang et al., 2011; Zhang and Li, 2012]. IQA methods are well established in the literature and aim to use computational models to measure the image quality consistently with subjective evaluations. The IQA metrics compare two images using values in the interval $[0, 1]$, where one means the best quality while zero means the worst quality. The first metric, called FSIM [Zhang et al., 2011], is based on the fact that the human visual system responds to an image mainly according to its low-level features, specifically the phase congruency and the gradient magnitude. Secondly, Zhang et al. [2011] also proposed an extension called FSIMc that uses the chrominance information in the YIQ color space. The third metric adopted is called SR-SIM [Zhang and Li, 2012]. This metric is based on the spectral residual visual saliency and the hypothesis that an image's visual saliency map is closely related to its perceived quality. This last method is only applicable to grayscale images.

4.3.1 Simulated Results

One of the challenges with validating our methodologies is the difficulty of acquiring images of participating medium with reliable ground truth. To tackle with this problem, we chose to validate our method by using the sequence *freiburg1_xyz* from the RGB-D SLAM dataset [Sturm et al., 2012], which includes the sequence of images and their respective depth maps. Furthermore, they provide the calibration parameters which allow to apply structure-from-motion and to simulate the participating medium effects.

We perform artificial degradation of the images by simulating the effects of an underwater camera at a depth of 5m. We simulated the attenuation coefficient η and the global illumination A as proposed in Mobley [1994], similarly to described in Sec. 4.1.1 and 4.2.1. We assumed a strong turbidity with the medium being contaminated with a chlorophyll concentration $C = 2.0\text{mg.m}^{-3}$ [Mobley, 1994]. Since the depth maps provided by the dataset are not perfect, we also performed inpainting [Bertalmio et al., 2000] and guided filtering [He et al., 2010].

Fig. 4.13 shows the inputs. Firstly, two frames from the dataset are selected to be used in the simulation (figs. 4.13a and 4.13e), each one with 640 × 480 pixels. Depth maps provided by the dataset are shown in figs. 4.13b and 4.13f. Notice that the depth maps are prone to errors and present many missing values (dark blue pixels). The depth maps are shown in color to improve the visualization, where bluish colors represent closer points and reddish colors represent distant points. The colors are the inverse of the medium transmission due to the behavior of the exponential term, *i.e.* large depth obtains medium transmission near zero.

Results after applying inpainting algorithm [Bertalmio et al., 2000] and guided filtering [He et al., 2010] are presented in figs. 4.13c and 4.13d for the depth map of Fig. 4.13b, and figs. 4.13g and 4.13h for the depth map of Fig. 4.13f. Images after simulation of participating medium effects are shown in figs. 4.13i and 4.13j. Notice that the space variant effect generated by the simulation, where the chair in the farthest region presents larger decreasing in visibility than the keyboard, which is closer to the camera.

Results of the structure-from-motion method by using the simulated images of Fig. 4.13 are shown in Fig. 4.14. The correspondences between points are obtained by using the accurate approach of GOFM. The selected flow is shown in Fig. 4.14e. Figs. 4.14a and 4.14c present the estimated depth maps using the method described in Sec. 3.3.2, after inpainting algorithm [Bertalmio et al., 2000]. The final depth maps after applying guided filter [He et al., 2009] are shown in figs. 4.14b and 4.14d.

Based on the estimated depth, we are able to estimate the attenuation coefficient. Fig. 4.15 illustrates the proposed approaches to estimate the coefficient. Figs. 4.15a-4.15f shows the relation between the depth variation (Δd) and the logarithm of the intensities, as shown in Eq. 3.27. Figs. 4.15a-4.15c show the plot of the ground truth depth, *i.e.* Fig. 4.13, and the state-of-art Classic-NL [Sun et al., 2014] for the correspondence between pixels. Even in this case, one can easily see some outliers. The ground truth attenuation coefficient is also depicted using cyan color. Despite the presence of outliers, a stochastic method could recover the coefficient. Figs. 4.15d-4.15f show a more realistic case where both depth and optical flow are estimated using the proposed methodology. The number of outliers significantly increases being hardier to correctly estimate the coefficient.

A similar situation occurs in figs. 4.15g-4.15l. These figures illustrate the effect shown in Fig. 3.14 for our simulated images. We show the points, intensity × depth, for the two simulated images of Fig. 4.13 using light and dark color. One can see in the ground truth case, figs. 4.15g-4.15i, that the attenuation coefficient (curve in cyan) is well limited by the low intensities points, as previously reported by Roser et al.

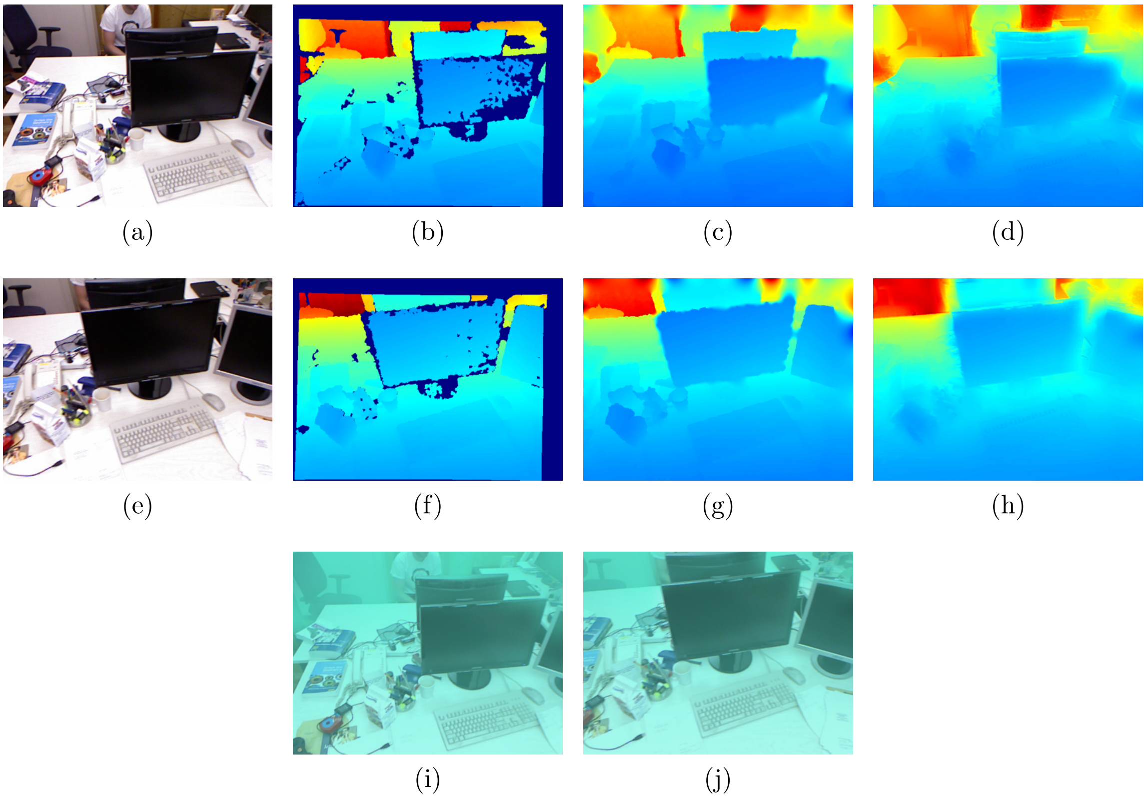
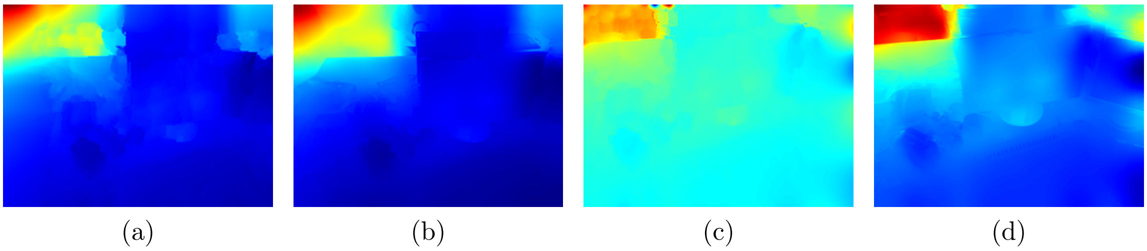
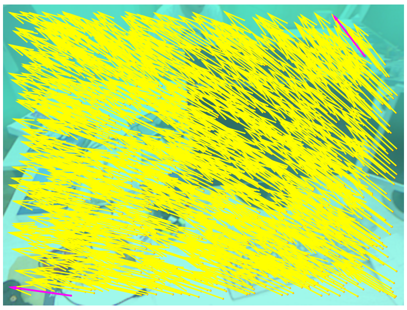


Figure 4.13. Simulation of underwater effects using images from RGB-D SLAM dataset [Sturm et al., 2012]: (a,e) images from the dataset, (b,f) depth maps provided by the dataset, (c,g) depth maps after inpainting algorithm [Bertalmio et al., 2000], (d,h) final depth maps after guided filter [He et al., 2010], and (i,j) simulated underwater effects with strong chlorophyll concentration of $C = 2.0\text{mg.m}^{-3}$ [Mobley, 1994] at 5m of water depth.

[2014]. This constraint, Eq. 3.30, is less sensible to the optical flow estimation due to the points on different image are not directly correlated. They are only indirectly correlated by the depth estimation. Figs. 4.15j-4.15l show the points using estimated values of depth. The fitting of the curve is not easily identifiable due to the outliers in the depth estimation. Thus, the estimation problem still remains challenging.

Based on both constraints, we have proposed two new methods to estimate the attenuation coefficient (Sec. 3.4). These method are called **MSAC-Single** and **MSAC-Multiple**. Two versions of the method **MSAC-Multiple** were proposed as described in Sec. 3.4. The first version sorts the depth of all points, and selecting the points with smallest intensity in each subset of points. This approach is called **MSAC-Multiple Sort**. The second approach uniformly selects the lowest intensity point in all depth depth bins, it is named **MSAC-Multiple Uniform**. Similarly to the MSAC-Multiple approach, we implemented a non-linear optimization [Coleman and Li, 1996], as pro-





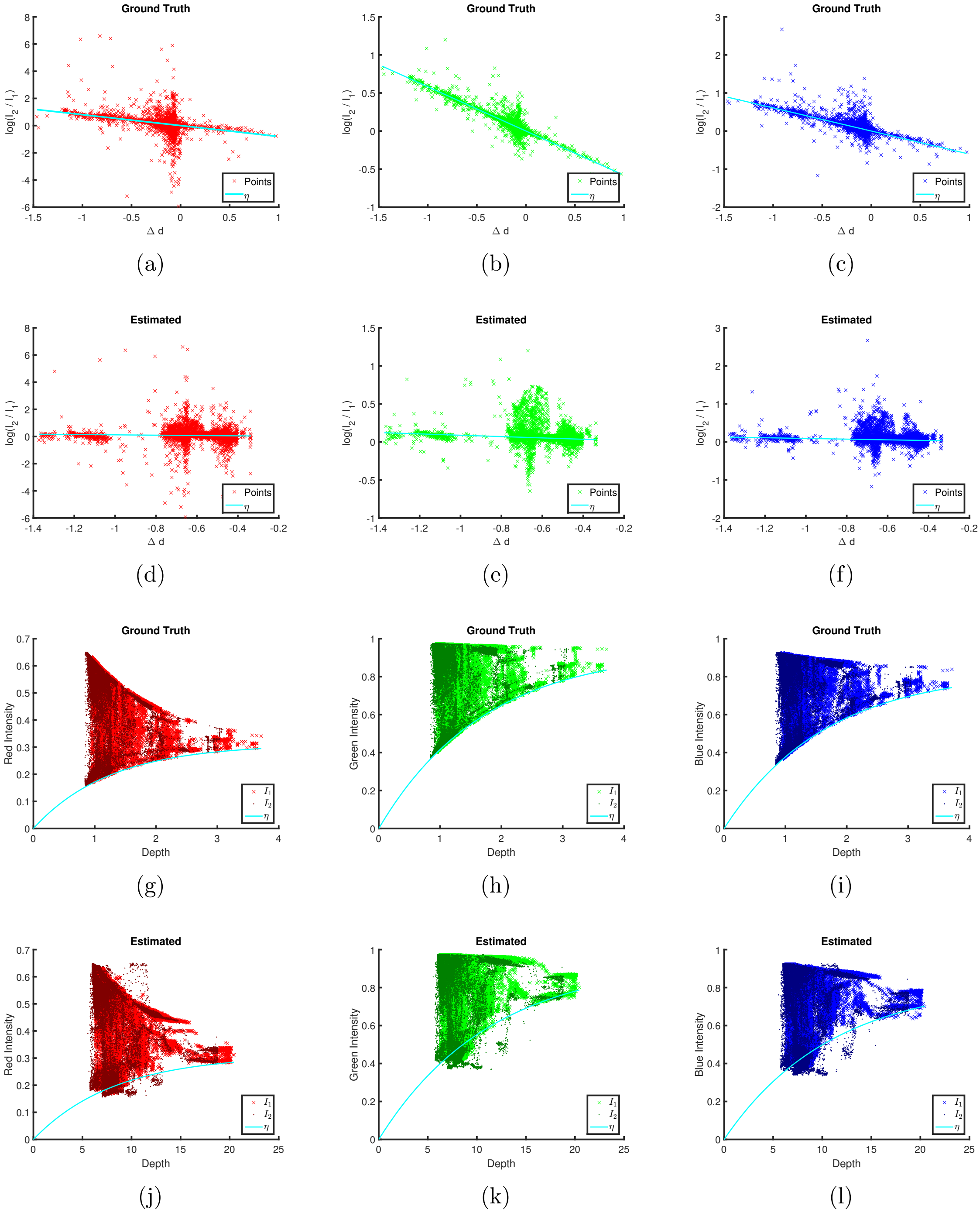
(e)

Figure 4.14. Results after the structure-from-motion procedure in the simulated images of Fig. 4.13: (a,c) estimated depth maps after inpainting algorithm [Bertalmio et al., 2000], (b,d) final depth maps after guided filter [He et al., 2010], and (e) estimated optical flow for the simulated images. The largest vector in horizontal and vertical axis are highlighted in pink.

posed in Roser et al. [2014], to solve Eq. 3.31. We named it as **OPT-Multiple**. Two versions are also proposed **OPT-Multiple Sort** and **OPT-Uniform**. They are compared with the state-of-art method proposed by Roser et al. [2014].

We evaluated the methods using our simulated images. We perform one thousand runs for each algorithm. The algorithms based on non-linear optimization were repeated due to the sensibility of the random initialization. We initialized these methods with values in the interval $\eta = [0; 1]$. The methods based on MSAC were repeated due to their stochastic nature and the difficult to adjust parameters that are typically unknown. All methods presented a small standard deviation ($< 1\%$). The average values are shown in Table 4.2.

The error obtained by the OPT-Multiple methods are the largest, mainly due to the difficulty of estimating the coefficient in the red channel. It is because of the sensibility of the approach to outliers, essentially in the third term in Eq. 3.31 that depends on the relation between pixels. The approach based on uniform selection obtains a smaller error than the OPT-Multiple Sort.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 4.15. Attenuation coefficient and its constraints: a-f) shows the first constraint (Eq. 3.27) using ground truth depth (a-c) and estimated depth (d-f); g-l) The second constraint (Eq. 3.30) using ground truth depth (g-i) and estimated depth (j-l). The depth varies between ground truth and estimated due to the estimation up to scale provided by the Structure-from-motion procedure. The ground truth coefficient is shown in cyan.

Table 4.2. Comparative study about the method to estimate the attenuation coefficient. The best results are highlighted in dark blue letters. Furthermore, the second best results are distinguished in light blue letters.

Method	η^{Red}	η^{Green}	η^{Blue}	Error
MSAC-Single	0.1323	0.0783	**0.0893**	0.0557
Roser et al. [2014]	0.1540	0.0959	0.1028	0.0352
OPT-Multiple Sort	1.5441	**0.0890**	0.0947	1.4225
OPT-Multiple Uniform	0.8348	0.0971	0.1073	0.7135
MSAC-Multiple Sort	**0.1295**	0.0844	**0.0915**	**0.0109**
MSAC-Multiple Uniform	**0.1193**	**0.0910**	0.1008	**0.0101**
Ground Truth Coef.	0.1215	0.0879	0.0918	-

A similar result is obtained for the MSAC-based methods, but with a small difference between them. Both methods are able to recover the coefficient for green and blue channel, with a very small advantage to the MSAC-Multiple Sort. However, the MSAC-Multiple Uniform obtains the best estimation for the red channel, the most challenging due to the higher absorption rate. This rate reduces the information carried and increases the noise. The MSAC-Single and the method of Roser et al. [2014] also present difficulties to compute the coefficient of red channel. The MSAC-Single is still competitive due to the second best result in terms of blue channel coefficient. However, the coefficient in the green channel is not well estimated. Fig. 4.15e explains this fact. One can see that the line in the estimated depth does not fit well in this color channel, however the curve fits better in the second relation (Fig. 4.15k).

Fig. 4.16 shows the restored images obtained by using the parameters shown in Table 4.2. The image after simulation of the underwater effects and the restored image using the ground truth coefficient are shown in figs. 4.16a and 4.16b, respectively. Our approach provides a significant improvement. However, some color distortion is still present since the estimated depth maps are prone to errors and gaps, mainly because of the noise, occlusion and the movement of the camera (Fig. 4.14). Results obtained by our approaches are visually similar, *i.e.* MSAC-Single in Fig. 4.16c, MSAC-Multiple Sort in Fig. 4.16g, and MSAC-Multiple Uniform 4.16h. Figs. 4.16e and 4.16f obtained by OPT-based methods present an important color distortion in terms of red channel because of the limited estimation of the attenuation coefficient (Table 4.2). Similar effect is observable in Fig. 4.16d which is obtained using the coefficient estimated by the method of Roser et al. [2014].

We compared the results of Fig. 4.16 using several metrics, *i.e.* IQA metrics [Zhang et al., 2011; Zhang and Li, 2012], the restoration score τ [Hautière et al., 2008a] and the keypoint detector SIFT [Lowe, 2004]. These results provide some insights about the capabilities of each metric.

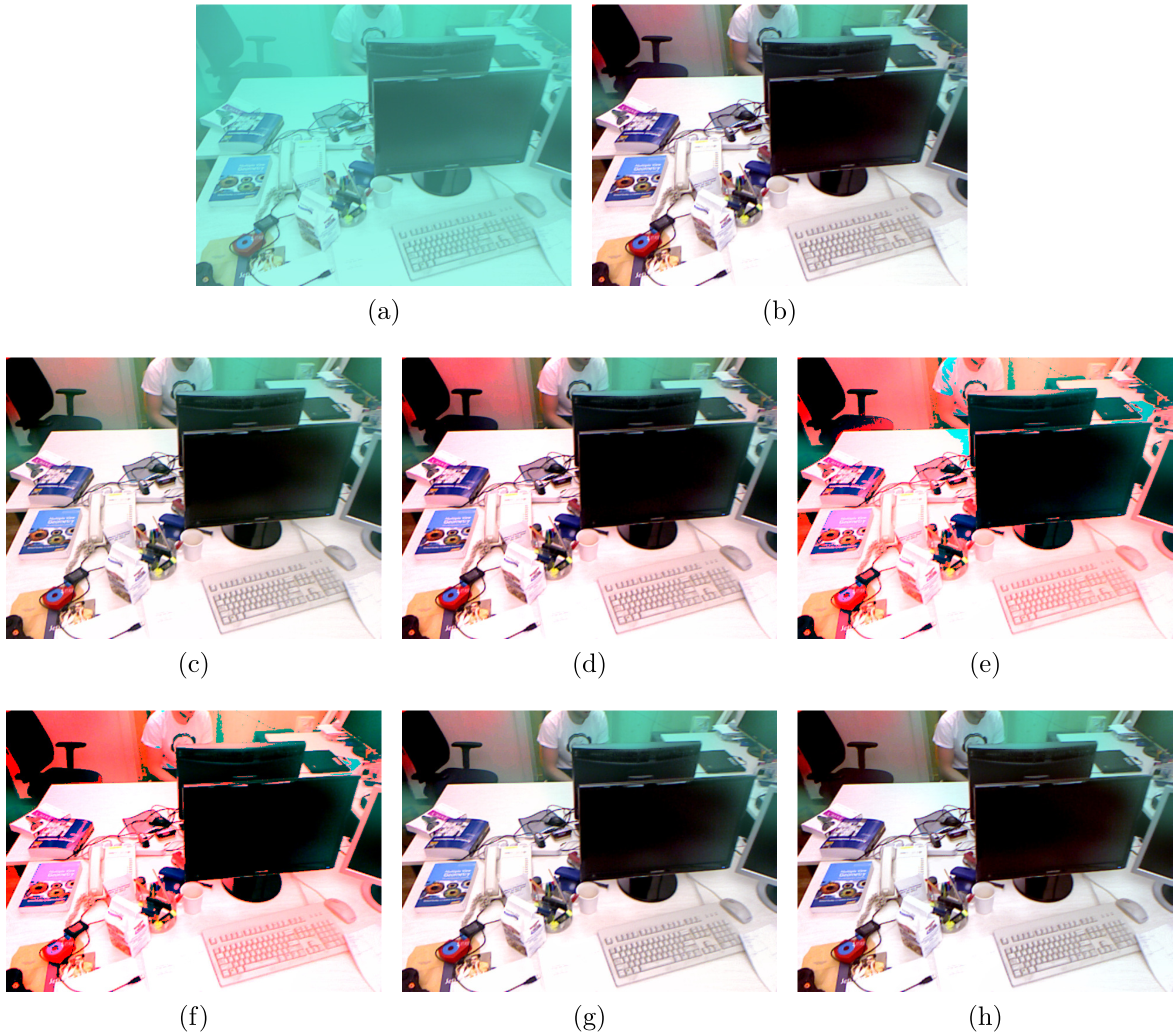

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4.16. Qualitative comparison of the estimation of the attenuation coefficient by using simulated effects: (a) sample image after simulated underwater effects with strong chlorophyll concentration; (b) restoration using our method with the ground truth attenuation coefficient. The results obtained by our method with attenuation coefficient estimated using the method: (c) MSAC-Single , (d)Roser et al. [2014], (e) OPT-Multiple Sort, (f) OPT-Multiple Uniform, (g) MSAC-Multiple Sort, and (h) MSAC-Multiple Sort.

Table 4.3 shows the results obtained using IQA metrics. The worst results for all metric is achieved using simulated underwater images, as expected. The OPT-Multiple methods obtain the smallest values. This fact is expected due to the error shown in Table 4.2. The restoration using the ground truth coefficient obtains the largest value in terms of FSIM and FSIMc metrics. However, the largest values for SR-SIM metric are obtained by the method of Roser et al. [2014] and MSAC-Multiple Sort. The difference to the ground truth coefficient is small, but the expected is the

Table 4.3. Comparative study using the IQA metrics for the simulated images, one of them shown in Fig. 4.16. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	FSIM I_1	FSIMc I_1	FSIM I_2	FSIMc I_2	SR-SIM I_1	SR-SIM I_2
Simulated Effects	0.8302	0.7332	0.8770	0.7736	0.8706	0.9243
MSAC-Single	0.9702	0.9473	0.9877	0.9742	0.9689	0.9890
Roser et al. [2014]	0.9674	0.9484	0.9789	0.9671	**0.9717**	0.9869
OPT-Multiple Sort	0.9218	0.8892	0.9442	0.9166	0.9461	0.9653
OPT-Multiple Uniform	0.9296	0.8997	0.9480	0.9221	0.9521	0.9687
MSAC-Multiple Sort	**0.9714**	0.9531	**0.9878**	0.9788	0.9705	**0.9892**
MSAC-Multiple Uniform	0.9712	**0.9538**	0.9873	**0.9791**	**0.9711**	0.9890
Ground Truth Coef.	**0.9718**	**0.9548**	**0.9882**	**0.9803**	0.9706	**0.9891**

method using ground truth coefficient obtains the largest values. The method FSIM is more coherent with the error in Table 4.2, mainly the FSIMc that considers the chrominance. The MSAC-Multiple approaches obtain the second best results, only worse than ground truth.

We also evaluated the proposed method to estimate the attenuation coefficient using the restoration score τ [Hautière et al., 2008a]. Table 4.4 shows the obtained results. This metric evaluates the improvement in terms of edges and contrast between degraded images and restored images. Although this technique is a important tool to evaluate the quality of the restoration, it presents limitations. The expected results for the original images, *i.e.* before simulation, are expected to be the best. Indeed, these image obtain the largest values for edges restoration and the τ metric. However, the contrast obtained by the restoration using OPT-Multiple are better. The distortion produced by the restoration with the wrong attenuation coefficient (see figs. 4.16e and 4.16f) generates saturation in red channel that artificially increases the contrast. Thus, these techniques obtain the largest values in terms of contrast, $\overline{r}$. The MSAC-Multiple methods obtain consistent improvement in terms of edge restoration, however it is expected to be smaller than the values obtained by the method using ground truth coefficient. Indeed, the MSAC-Multiple methods and ground truth coefficient obtain similar values of τ metric.

The restoration score τ should be adopted with caution, mainly due to the fact that distorted results could increase the contrast and even the number of edges. The color information is not take into account in this technique which also limits the performance. This explains the method using ground truth coefficient does not obtain larger values than the others.

The matching of keypoints obtained by the SIFT [Lowe, 2004] is also adopted to evaluate the results. Accordingly, the original images provide the largest number of

Table 4.4. Comparative study using the restoration score τ [Hautière et al., 2008a] to evaluate the restored images (Fig. 4.16). The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	e_1	$\bar{r}_1$	τ_1	e_2	$\bar{r}_2$	τ_2
Original	**1.6097**	2.3742	**4.984**	**1.9123**	1.9945	**4.9067**
MSAC-Single	1.1987	2.2945	4.4233	1.5352	2.0863	4.5354
Roser et al. [2014]	1.2787	2.4288	4.5821	1.5718	2.1399	4.5482
OPT-Multiple Sort	1.0	**2.7274**	**4.6095**	1.2162	**2.3698**	4.4419
OPT-Multiple Uniform	0.92782	**2.6471**	4.4123	1.1262	**2.2441**	4.1712
MSAC-Multiple Sort	1.245	2.3194	4.4845	1.6407	2.093	4.6352
MSAC-Multiple Uniform	**1.3428**	2.2722	4.5476	**1.7616**	2.0206	**4.7014**
Ground Truth Coef.	1.3095	2.2738	4.5133	1.7109	2.0446	4.6694

correct matches, while the images with simulated effects obtain the smallest. Again, the method using ground truth coefficient does not obtain the second best results. The result of the method MSAC-Multiple Sort is the second best, although the number of detected keypoints is slightly smaller than the MSAC-Multiple Uniform and Ground Truth Coefficient. The methods OPT-Multiple obtain smallest number of matching between all restoration techniques evaluated, as expected. Although the improvement in terms of contrast (Table 4.4), they do not provide stable features. The number of keypoints detected in the restoration using Roser et al. [2014] is the largest, considering only the restoration methods. This fact is also corroborated by the contrast improvement shown in Table 4.4. The color distortion generated by the Roser et al. [2014] method is not take into account by SIFT due to the use of grayscale images.

Matching between images using keypoint detector, like SIFT [Lowe, 2004], is a important tool of computer vision algorithms. As the method is partly invariant to illumination variation, it could deal with small color distortion. Moreover, it is also a important tool to evaluate the quality of the restoration, but as complementary information with other metrics like τ [Hautière et al., 2008a].

Finally, the MSAC-Multiple methods are similar, as shown in Table 4.2. The smaller error is obtained by MSAC-Multiple Uniform, but with a difference smaller than 10^{-3}. In terms of IQA, the MSAC-Multiple Uniform obtains the largest values in terms of FSIMc, while the MSAC-Multiple Sort obtains in terms of FSIM which does not take into account the color information. The MSAC-Multiple Uniform is better in terms of τ metric while the number of matching using SIFT is larger to MSAC-Multiple Sort. Theoretically, the uniform distribution of the points is more well distributed and with a larger number of points in the largest depths that, typically, present a smaller number of points. The sort distribution tends to following the point distribution in the scene and, usually, obtains more points in the smaller depths, *i.e.* following the

Table 4.5. Comparative analysis using the image descriptor SIFT [Lowe, 2004] to evaluate the restored images (Fig. 4.16). The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	Correct Matches	Keypoints I_1	Keypoints I_2
Original	**61**	2220	1582
Simulated Effects	34	1066	696
MSAC-Single	57	2086	1542
Roser et al. [2014]	55	2206	1618
OPT-Multiple Sort	34	1911	1376
OPT-Multiple Uniform	40	1900	1458
MSAC-Multiple Sort	**59**	2126	1560
MSAC-Multiple Uniform	56	2160	1582
Ground Truth Coef.	58	2139	1567

Table 4.6. Comparative study using the IQA metrics for the images shown in Fig. 4.17. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	FSIM I_1	FSIMc I_1	FSIM I_2	FSIMc I_2	SR-SIM I_1	SR-SIM I_2
Simul. Effects	0.8302	0.7332	0.8770	0.7736	0.8706	0.9243
UDCP	0.9267	0.8640	0.9167	0.8459	**0.9485**	0.9382
Hist. Eq.	0.9231	**0.8972**	0.8878	**0.8709**	0.9311	0.9149
CLAHE	**0.9324**	0.8593	**0.9384**	0.8702	0.9479	**0.9512**
Our Method	**0.9714**	**0.9531**	**0.9878**	**0.9788**	**0.9705**	**0.9892**

behavior of the attenuation curve. We adopted the **MSAC-Multiple Sort** in the present work due to its characteristics that tends to follow the attenuation curve.

4.3.1.1 Comparative Evaluation

Qualitative results for the restoration of the simulated images (Fig. 4.13) are presented in Fig. 4.17. We compared our approach with the UDCP, histogram equalization [Szeliski, 2010] and CLAHE [Zuiderveld, 1994].

The results of our method are presented in Fig 4.17e and 4.17f. The restored image is not perfect as argued in the previous section. However, we obtained a significant improvement, *e.g.* the chair in the top left.

Quantitative results are obtained using the same metrics adopted in previous section (Sec 4.3.1). Table 4.6 shows the results for the restored images (Fig. 4.17) using IQA. One can readily see that our methodology produces the largest values in those metrics.

Table 4.7 shows the results using the restoration score τ [Hautière et al., 2008a]. The limitation of this metric are possible to see again in the result obtained for the

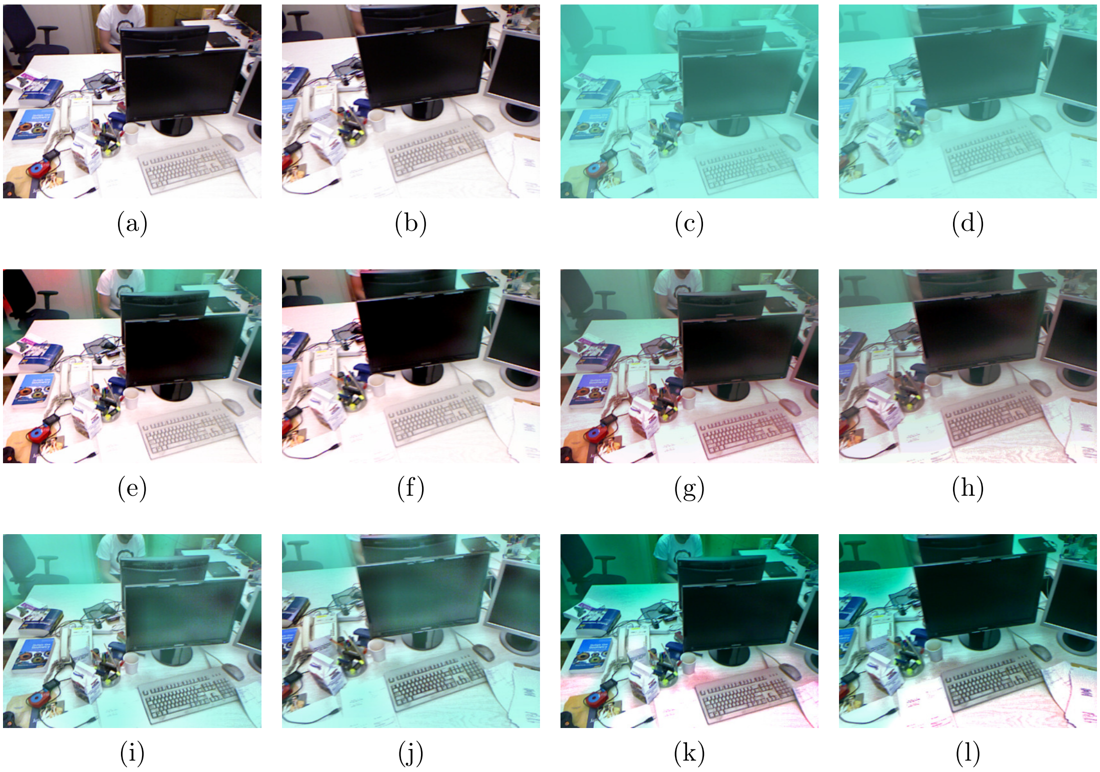

(a) (b) (c) (d)
(e) (f) (g) (h)
(i) (j) (k) (l)

Figure 4.17. Qualitative comparison using simulated effects in images from RGB-D SLAM dataset [Sturm et al., 2012] (a-b). (c-d) show simulated underwater effects with strong chlorophyll concentration of $C = 2.0\text{mg.m}^{-3}$ [Mobley, 1994] at 5m of water depth. Results obtained by our methodology (e-f), applying histogram equalization techniques (g-h), CLAHE (i-j) and UDCP (k-l).

Table 4.7. Comparative study using the restoration score τ [Hautière et al., 2008a] for the simulated images (Fig. 4.17). The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	e_1	$\overline{r}_1$	τ_1	e_2	$\overline{r}_2$	τ_2
Original	1.6097	**2.3742**	**4.984**	1.9123	1.9945	4.9067
UDCP	**1.6616**	2.153	4.8146	**2.2949**	2.1318	**5.4267**
Hist. Eq.	**1.8623**	2.0616	**4.9147**	**2.6685**	**2.265**	**5.9277**
CLAHE	1.1338	**2.3376**	4.4714	1.4471	**2.3361**	4.7832
Our Method	1.245	2.3194	4.4845	1.6407	2.093	4.6352

original image, *i.e.* in non-participating medium. We expected to see the best results for these images (as shown in Table 4.4), at least for τ metric. However, the results obtained for histogram equalization, and even for UDCP in the case of I_2, are better. The main function of the histogram equalization technique is increase the contrast, but can produce unrealistic effects like false color. The histogram equalization is not able to remove the "haze" effect, mainly in the large depth as the chair in the top left of figs.

Table 4.8. Comparative study using the image descriptor SIFT [Lowe, 2004] for the simulated images (Fig. 4.17). The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	Correct Matches	Keypoints I_1	Keypoints I_2
Original	**61**	2220	1582
Simulated Effects	34	1066	696
UDCP	49	2334	1848
Hist. Eq.	40	2188	1500
CLAHE	40	2373	1761
Our Method	**59**	2126	1560

4.17g and 4.17h. However, it increases the noise as one can see in the desk near the keyboard (bottom of figs. 4.17g and 4.17h). This fact explains the increases in edges and contrast in relation to the ground truth image. Despite the quality of the restored images, our method obtains intermediate values in these metrics.

Table 4.8 presents the obtained results using SIFT [Lowe, 2004] for the restored images shown in Fig. 4.17. Results show that our approach provides the largest number of correct matches, while the number of detected keypoints for UDCP and CLAHE are the largest. It is due to the increasing in the contrast and, mainly, in the noise. Despite the large number of detected points, the number of matched points does not increase in the same ratio. Notice that the number of keypoints estimated for UDCP and CLAHE approaches are greater than the obtained in the original image, *i.e.* in a non-participating medium.

4.3.2 Real Results

We obtained real results in three different scenarios. Firstly, we captured images in the Brazil's Southeast Coast with depth ranging from 12m to 20m. Secondly, we captured image in a river located at a Serra do Cipo National Park in Brazil with depth ranging from 1m to 5m. Finally, we obtained a sequence from a residential area in a foggy day.

4.3.2.1 Oceanic Water

Real underwater image sequences were obtained with an underwater vehicle, the Seabotix LBV300-5, from NAUTEC-FURG equipped with a color camera (Fig. 4.18). Our method was applied to a sequence of images acquired in Brazil's Southeast Coast, specifically in the *Parcel do Carpinteiro*, a reef with irregular bottom topography formed by rocks, gravel and sand. The reef is located approximately 17 nmi from the coast, and has depths ranging from 12m to 20m.

Figure 4.18. Experimental platform used to acquire sequence of images: an underwater vehicle Seabotix LBV300-5 equipped with a color camera.

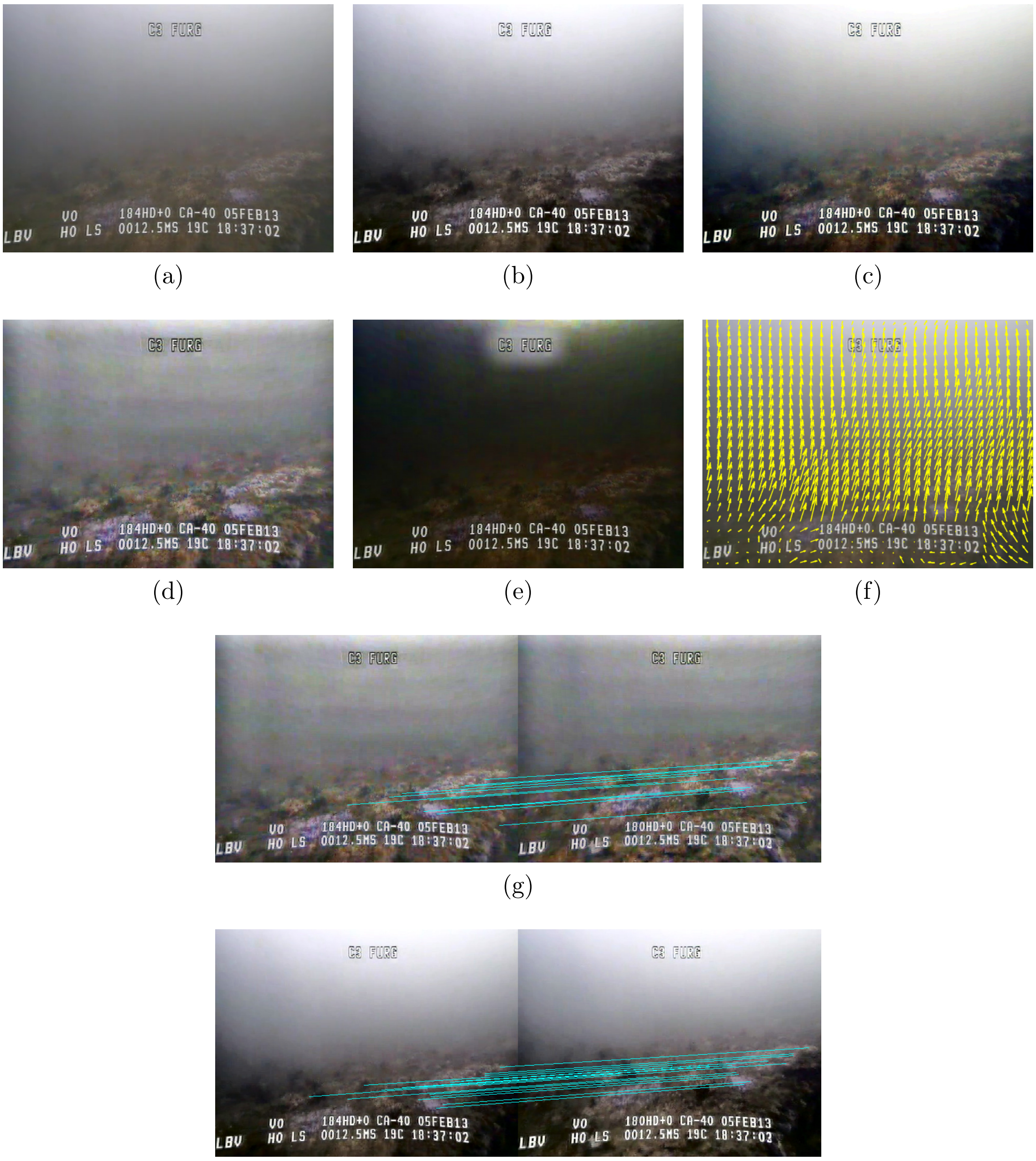
The sequence was acquired with On-Screen Display (OSD), and the resolution of the captured images is 640×480 pixels. All pixels in the OSD area were discarded. The radial distortion was corrected, and as a result the OSD information is distorted in the images.

The attenuation coefficient is estimated as $\eta = [0.1442, 0.1424, 0.0907]$, for each RGB channel respectively. The blue channel typically has a smaller attenuation value while the red channel has a larger value as shown in the results. It is worth noting that this coefficient is obtained up to a scale factor due to the depth map estimation.

Fig. 4.19 shows the qualitative comparison of our restoration method for a sample pair of images acquired in naturally lit shallow water. Fig. 4.19a is a sample image acquired by the ROV, with limited visibility and significant color distortion. Fig. 4.19b show the images restored by our methodology, where the quality is improved. The results obtained using adaptive histogram equalization are shown in Fig. 4.19d, where the contrast and the noise are increased but the colors are distorted. The restored images using histogram equalization are shown in Fig. 4.19c and using UDCP in Fig. 4.19e. Fig. 4.19f shows the optical flow estimated using the GOFM approach. The method is able to correctly estimate the flow mainly in textured areas in the center of the image, and in the bottom right corner.

Our method outperforms the others in term of the τ metric. It achieves a improvement in terms of new edges in the image, although it is limited in terms of contrast. CLAHE obtains the largest values of $\overline{r}$ because of the increase in contrast, including the seabed and the horizon area. The contrast is an interesting metric of the quality of the restored image but it is sensible to noise and color distortion as reported in Sec. 4.3.1.

Table 4.10 presents the results obtained for the restored images shown in Fig. 4.19. They show that our approach provides the largest number of correct matches, while the number of detected keypoints for the CLAHE method is larger. This is due to the increase in contrast and, mainly, to noise. This larger number of detection is



(a) (b) (c)

(d) (e) (f)

(g)

(h)

Figure 4.19. Qualitative comparison between our methodology, UDCP and histogram equalization techniques using a sequence of images acquired in naturally lit shallow oceanic water. (a) original image, (b) restored by our methodology, (c) restored using histogram equalization, (d) restored using CLAHE, and (e) using UDCP. (f) shows the optical flow estimated using GOFM, and (g, h) the SIFT matching in restored images using CLAHE and the proposed method, respectively.

expected since the restoration obtained by CLAHE (Table 4.9) presented the larger $\overline{r}$. However, this restoration is not stable, thus it does not increase the number of corrected matches. Figs.4.19g-4.19h show that the matches obtained by our approach

Table 4.9. Comparative study using the restoration score τ [Hautière et al., 2008a] for the images shown in Fig. 4.19. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	e_1	$\overline{r}_1$	τ_1	e_2	$\overline{r}_2$	τ_2
UDCP	1.7259	1.2682	3.9940	1.6918	1.2499	3.9416
Hist. Eq.	**2.7595**	2.3023	**6.0503**	**2.3265**	2.6065	**5.9266**
CLAHE	1.9001	**2.7862**	5.6863	1.8640	**2.9126**	5.7766
Our Method	**2.8621**	**2.4676**	**6.3292**	**2.6277**	**2.7266**	**6.3537**

Table 4.10. Comparative analysis using the image descriptor SIFT [Lowe, 2004] for the images shown in Fig. 4.19. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	Correct Matches	Keypoints I_1	Keypoints I_2
Underwater	0	4	4
UDCP	1	18	18
Hist. Eq.	10	151	243
CLAHE	**11**	359	422
Our Method	**17**	176	322

cover a larger area of the image than those obtained from CLAHE.

4.3.2.2 Fresh Water

Another real underwater image sequences were obtained with a standard OpenROV v2.7 [Stackpole et al., 2015] from ASL-CSIRO equipped with a Genius KYE F100 Ultra-wide angle full HD webcam (Fig. 4.20). Our method was applied to a sequence of images acquired in a natural pool in the Cipo River at the *Serra do Cipo National Park*. It is located around 100 km from Belo Horizonte, the capital of Minas Gerais state, Brazil. The natural pool presents a topography formed by rocks, gravel and sand, and the depths ranging from 1m to 5m. A GoPro Hero3+ Black Edition camera was attached to the vehicle, but it was not used due to its limited capabilities to provided reliable data for structure-from-motion because the rolling shutter scheme [Thoeni et al., 2014].

The evaluated sequences were acquired at 720p resolution. We needed to reduce the useful area of the image due to the process of correct the lens distortion, *e.g.* Fig. 4.21a. Thus, the final resolution adopted is 1030 × 557 pixels.

Fig. 4.21 shows the qualitative comparison of our restoration method for a sample image acquired in a fresh water. Fig.4.21a is a sample image after the correction of the radial distortion. Fig. 4.21b is the useful area of the image acquired by the ROV that presents limited visibility, and a significant presence of noise. Fig. 4.21c show the image restored by our methodology. The quality is improved in some areas of

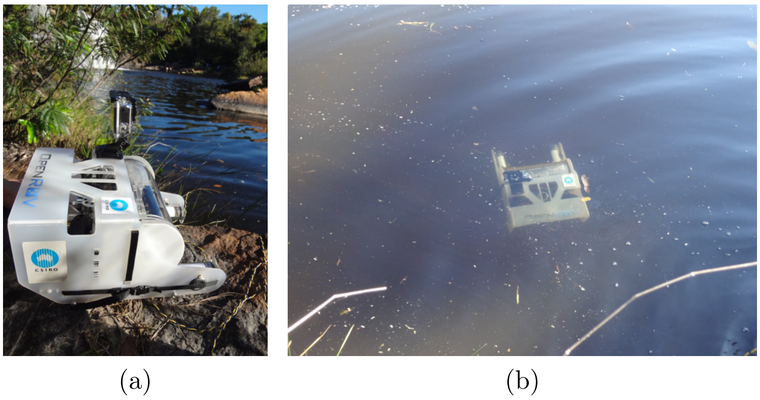

(a) (b)

Figure 4.20. Experimental setup of the experiment in fresh water: (a) an OpenROV platform near the natural pool where we conducted field tests; (b) the vehicle in the water.

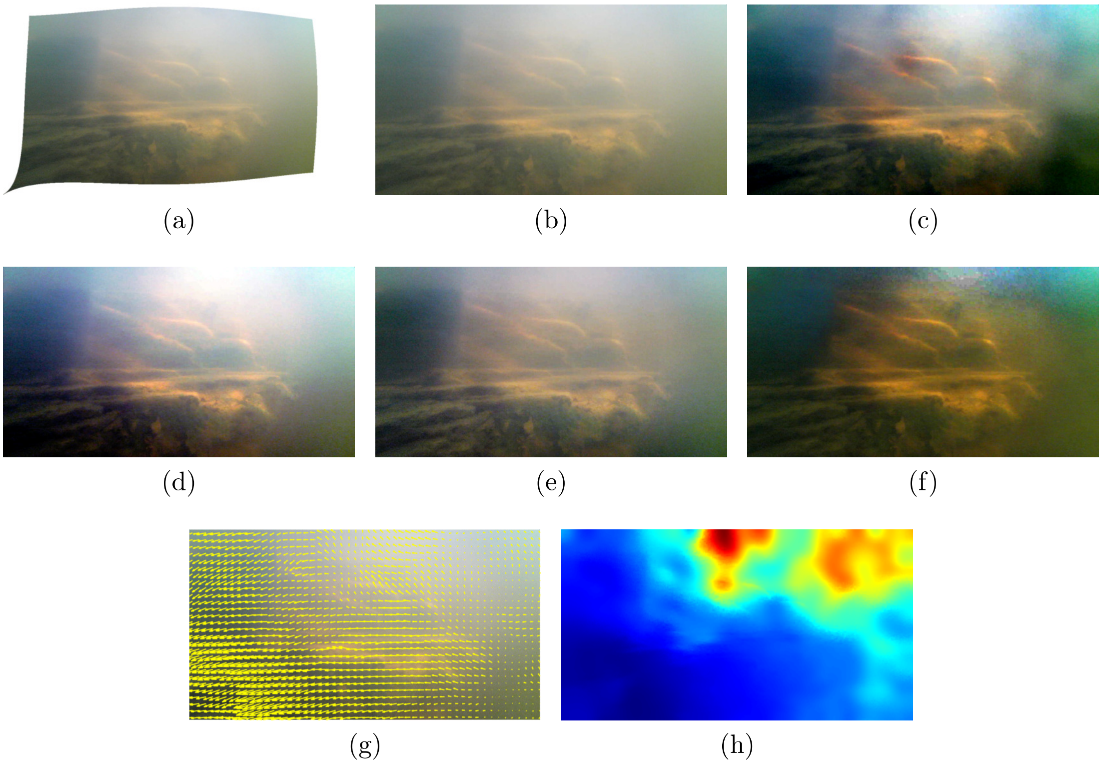

(a) (b) (c)
(d) (e) (f)
(g) (h)

Figure 4.21. Qualitative comparison between our methodology, UDCP and histogram equalization techniques using a sequence of images acquired in naturally lit shallow fresh water. (a) original image after correction of the radial distortion; (b) useful area of the image in (a); (c) restored by our methodology; (d) restored using histogram equalization; (e) restored using CLAHE, and (f) using UDCP. (g,h) show the optical flow estimated using GOFM and the estimated colorized depth map, respectively.

Table 4.11. Comparative study using the restoration score τ [Hautière et al., 2008a] for the images shown in Fig. 4.21. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	e_1	$\overline{r}_1$	τ_1	e_2	$\overline{r}_2$	τ_2
UDCP	**8.0147**	1.4888	**10.5035**	**7.6413**	1.5527	**10.1940**
Hist. Eq.	6.8679	1.5451	9.4003	6.2243	1.3906	8.6031
CLAHE	6.2802	**1.6918**	8.9720	5.835	**1.6711**	8.5061
Our Method	**9.338**	**1.604**	**11.9361**	**9.1599**	**1.5824**	**11.7172**

the image, mainly in the stone in the center of the image. Despite the improvement, the image presents some halos due to the difficult to estimate optical flow (Fig. 4.21g) and, consequently, the depth map (Fig. 4.21h) in this challenging scenario. The results obtained using histogram equalization and CLAHE are shown in figs. 4.21d and 4.21e, respectively. The results are similar and limited to deal with the "haze" effect. The restored images using UDCP are shown in Fig. 4.21f. The results are comparable to ours, but with some color distortion. However, they produce a reduced number of halos. Figs. 4.21g and 4.21h show the optical flow estimated using the GOFM approach and the colorized depth map, respectively. The method is able to correctly estimate the flow, and the depth, mainly in textured areas in the center and the bottom left of the image. Nonetheless, the estimation is prone to error in the top right area that produces halos in the restored image.

The attenuation coefficient is estimated as $\eta = [0.0439, 0.0406, 0.0309]$, for each RGB channel respectively. Similarly to Sec. 4.3.2.1, the blue channel is the smallest attenuation coefficient while the red channel has the largest value as shown in the results, as expected. It is worth noting that this coefficient is obtained up to a scale factor due to the depth map estimation.

Our method outperforms the others in term of the τ metric. As previously reported in Sec 4.3.2.1, our method achieves the largest improvement in terms of new edges in the image. However, it is limited in terms of contrast. CLAHE obtains the largest values of $\overline{r}$ because of the increase in contrast, mainly due to the increasing of the noise. This fact is corroborated by the reduce number of keypoints detected by the CLAHE using SIFT (Table 4.12).

Table 4.12 presents the results obtained for the restored images shown in Fig. 4.21. For each pair of images, we show the number of keypoints detected in both images and the number of correct matches. Despite the improvement in terms of edges and contrast, the methods are unable to matching the points. It is due to the quality of the image, but also to the natural ambiguities of the scene. The number of keypoints varies between approaches on which the histogram equalization obtaining the largest

Table 4.12. Comparative analysis using the image descriptor SIFT [Lowe, 2004] for the images shown in Fig. 4.21.

	Correct Matches	Keypoints I_1	Keypoints I_2
Original	0	2	1
UDCP	0	28	34
Hist. Eq.	0	33	21
CLAHE	0	12	5
Our Method	0	24	29

number in the image I_1 and the UDCP in the image I_2. The proposed method obtains a intermediary number of points between these approaches.

The quality of result obtained in this scenario is limited. All of the evaluated methods are not able to provide a good restoration. Our method obtain important improvement in terms of edges and contrast but not enough to use SIFT, for example. The limitation of the methods are related to the quality of the images acquired by the OpenROV's webcam, high turbidity, and characteristics of the scene. Fresh water is one of the most challenge participating medium due to the high turbidity, but the presence of sediments in the water, non-uniform illumination and the water currents impose additional hindrances.

4.3.2.3 Foggy Sequence

Another important participating medium is outdoor scenes under fog. We acquired a sequence of images in an urban foggy day of a static scene with smooth movements. The scene is composed by houses, buildings, trees, light poles, cars, etc. The camera adopted to acquire the sequence is a Sony Cyber-shot DSC-WX7. The images was acquired using 640×360 pixels.

Fig. 4.22 shows the qualitative comparison of our restoration method for a sample pair of images acquired in a foggy day. Fig.4.22a is a sample image that presents limited visibility, and color distortion due the presence of fog. Fig. 4.22b show the image restored by our methodology. The visibility and the color are improved, mainly in the houses (bottom left). The buildings in the center of image are under a strong "haze" layer, thus limiting the capability of restoration due to the optical depth [Treibitz and Schechner, 2009c]. However, the contours of the buildings are recovered by our approach.

The results obtained using histogram equalization method and CLAHE are shown in Fig. 4.22c and 4.22d, respectively. The result for histogram equalization is similar to our result. One important difference is in the largest tree that the color distortion produced by the fog remains in the restored by histogram equalization method. How-

Table 4.13. Comparative study using the restoration score τ [Hautière et al., 2008a] for the images shown in Fig. 4.22. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	e_1	$\overline{r}_1$	τ_1	e_2	$\overline{r}_2$	τ_2
DCP	1.5037	0.7937	3.2974	1.3333	0.78177	3.1151
UDCP	1.5475	0.79925	3.3467	**1.6929**	0.78697	3.4799
Hist. Eq.	**1.6288**	**1.9185**	**4.5340**	1.4791	**1.7349**	**4.1993**
CLAHE	0.51259	1.4918	3.0044	0.4866	1.5243	3.0109
Our Method	**1.666**	**2.1168**	**4.7703**	**1.4948**	**1.8209**	**4.3157**

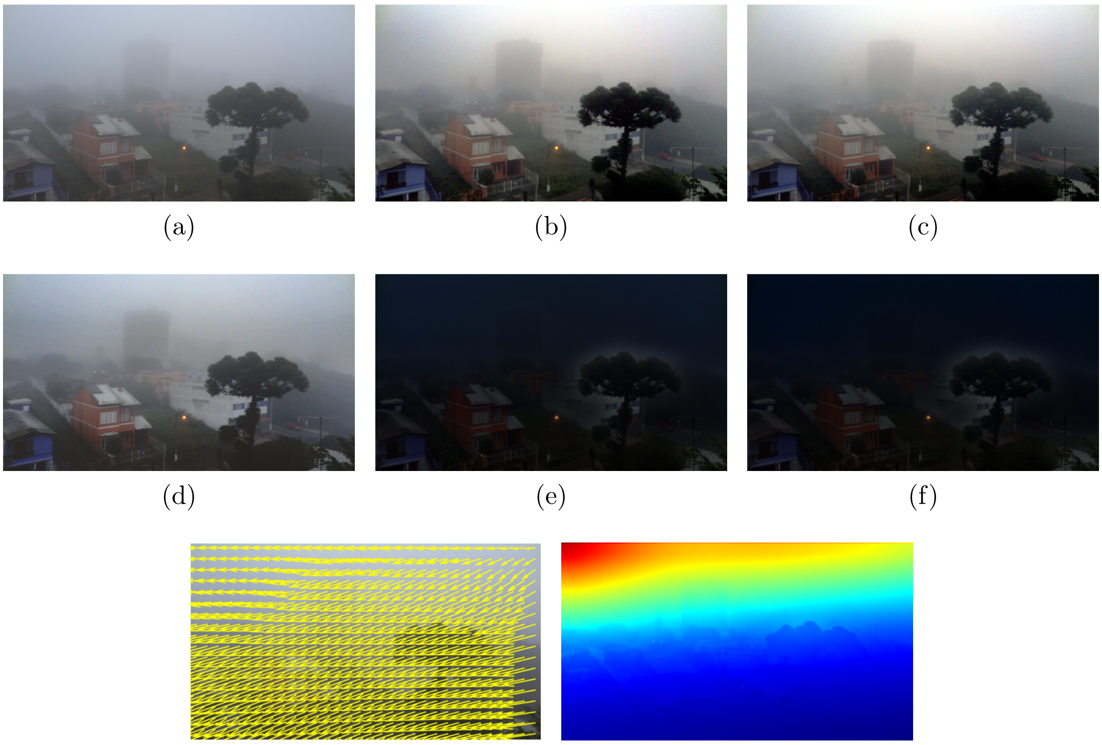
ever, our approach is able to improve the color and visibility, reducing the "haze effect". The improvement provided by CLAHE is imperceptible. Both DCP and UDCP are shown in figs. 4.22e and 4.22f. The estimation of the global light fails in both case, thus the image becomes darker. This fact limits their capabilities to restore the image.

The depth map estimated using our approach (Fig. 4.22h) is well estimated, but lacking in terms of details. The scene near the camera, *i.e.* composed of trees and houses, appear to be in the same plane. Conversely, the increasing in the depth is related with the height, as expected. The optical flow (Fig. 4.22g) is more uniform than expected, thus losing details about the depth of the scene.

The attenuation coefficient is estimated as $\eta = [0.1752, 0.2026, 0.1882]$, for each RGB channel respectively. Differently to Sec. 4.3.2.1 and 4.3.2.2, the red channel is the smallest attenuation coefficient while the green channel has the largest value. However, these three coefficients are relatively similar ($\approx 15\%$), as expected for the fog that is typically assumed as invariant to the wavelength.

Our method outperforms the others in term of the τ metric (Table 4.13). Our method obtains similar results to histogram equalization method, as expected. However, our method presents a small advantage in all metrics. The results of CLAHE technique is limited, presenting a small improvement in terms of visibility and color. Therefore, CLAHE obtains the smaller values in the number of edges, e, and the restoration score τ [Hautière et al., 2008a]. DCP and UDCP distort the colors, but this fact is not take into account by this metric. Conversely, the contrast metric ($\overline{r}$) indicates the bad restoration of these methods. This fact is corroborated by the reduce number of keypoints detected by the UDCP/DCP using SIFT (Table 4.14).

Table 4.14 presents the results using SIFT obtained by the restored images shown in Fig. 4.22. For each pair of images, we show the number of keypoints detected in both images and the number of correct matches. Our method obtains the largest number of matching points as well as keypoints detected. The histogram equalization also presents good results as described in Table 4.13. The DCP/UDCP are able to recover



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 4.22. Qualitative comparison between our methodology, UDCP and histogram equalization techniques using a sequence of images acquired in a foggy day. (a) original image; (b) restored by our methodology; (c) restored using histogram equalization; (d) restored using CLAHE, (e) DCP and (f) UDCP. (g,h) shows the optical flow estimated using GOFM and the estimated colorized depth map, respectively.

Table 4.14. Comparative analysis using the image descriptor SIFT [Lowe, 2004] for the images shown in Fig. 4.22. The best results are highlighted in blue letters. Furthermore, the second best results are distinguished in light blue letters.

	Correct Matches	Keypoints I_1	Keypoints I_2
Original	36	80	97
DCP	10	28	32
UDCP	8	28	32
Hist. Eq.	**70**	322	250
CLAHE	65	264	265
Our Method	**78**	410	310

a smaller number than the original image, *i.e.* obtained under fog.

Chapter 5

Conclusions

This dissertation proposed a new methodology to restore sequence of images acquired in participating media. We explore the temporal relation between the images that allow us to estimate scene structure and depth maps. This temporal relation is obtained by a new optical flow technique adapted for participating media. This approach depends on the knowledge of a rough medium transmission, which is computed by using statistical priors. At last, we developed a robust methodology to estimate the most critical parameter of the medium: the attenuation coefficient.

Although DCP is a very intuitive and fast method, we show its limitation in underwater conditions. The main important limitation is related to the red channel that suffers with a high absorption rate, generating a wrong estimation. UDCP presents an alternative option for underwater conditions. It shows good results even in situations where other single image methods can fail. Following the validation proposed for DCP, we performed experimental verification using several datasets which shows UDCP is able to deal with underwater and hazy images. We obtained experimental results to evaluate UDCP in terms of medium transmission estimation and image restoration. The method outperforms other statistical prior in the literature.

We proposed a novel methodology to estimate optical flow in participating media images, called GOFM. We evaluated it in simulated and real outdoor under fog and underwater image sequences. Results show an important gain in terms of endpoint and average angular errors, even though using restored images as input. Our method has shown better performance in comparison with classical methods like Horn-Schunk (HS) [Horn and Schunck, 1981] and Lucas-Kanade (LK) [Lucas and Kanade, 1981], with LK presenting better results for short-range motions and the HS for large range motion. State-of-the-art method proposed by Madjidi and Negahdaripour (MN) [Negahdaripour and Madjidi, 2003; Madjidi and Negahdaripour, 2006] is only competitive using restored images in fog situation that shows its limited application. The proposed implementation using modern methodologies presents better results in terms of endpoint and average angular errors when compared with state-of-the-art methods for non-participating medium as the modern implementation of Horn-Schunk [Sun et al., 2010] and the Classic-NL Sun et al. [2014]. This last method is one of the best ranked method in Middlebury Optical Flow Benchmark [Baker et al., 2011].

A structure-from-motion technique based on continuous essential matrix, triangulation and bundle adjustment was applied to obtain depth maps from the scene. Qualitative and quantitative results in simulated and real images show the quality and limitations of the estimated depth maps. Despite the limitation, the depth maps are enough to restore the images.

The present approach is sensible to the quality of the GOFM and CEM estimation that depends on the displacement between images. The GOFM uses a multiple scale approach that enables to estimate optical flow with large displacement, but GOFM is still sensible to it. Associated to this, the bundle adjustment technique only guarantees to achieve a local minimum. Thus, it solves just partially the limitation of GOFM and CEM. Moreover, the triangulation method is also sensible to the baseline between the captured images. Small baselines produces a noisy estimation. Therefore, our method is limited to a good trade-off between sensibility of the GOFM and the triangulation method to achieve a good depth estimation. We intend to overcome this using more robust techniques as a future work.

Finally, we proposed two approaches to estimate the attenuation coefficient called **MSAC-Single** and **MSAC-Multiple**. They are robust even in the presence of outliers. The proposed approaches outperforms the state-of-the-art method [Roser et al., 2014] in our simulated images. Our method enables us to restore the visibility, even though we obtain a bad estimation of the attenuation coefficient. However, the restored image suffers with color distortion.

5.1 Future Works

Future work will focus on investigating multiple view structure-from-motion approaches to improve the depth map estimation. There exist many techniques in the literature, all of them uses the photometric constancy as a standard constraint. The idea is improve the proposed method using multiple images but still relaxing the photometric constancy. Recently, some methods using dense reconstruction were proposed [Newcombe et al., 2011; Engel et al., 2014; Concha and Civera, 2015]. They produce impressive scene reconstruction by using dense information. This kind of technique was adopted by the author and colleagues [Concha et al., 2015] to underwater sequences, but removing “haze” areas from the images. We intend to extend this approach to deal with the effects of participating medium.

Other important issue to be addressed is the inclusion of artificial illumination in the scene, mainly in the underwater case. It will generate a non-uniform illumination and a non-homogeneous medium, thus a new algorithm needs to be proposed. Interestingly, the artificial illumination was the focus of the work of Treibitz and Schechner [2009a]. They deal with the artificial illumination adopting a simple calibration procedure. The global light becomes a function of the pixels, *i.e.* $A(x)$. The GOFM is not valid in this context, thus providing new opportunities of research.

Furthermore, the GOFM needs to be more explored. The effects of the classical “aperture” problem in the new formulation is one point to be investigated. Moreover, the sensibility of the GOFM to the quality of the depth estimation and the turbidity also needs to be investigated.

The UDCP/DCP seems to be a powerful tool to restore images acquired in participating media. However, the limited amount of information provided by a color image, *i.e.* the three RGB colors, constraints the applicability and robustness of the method. However, hyperspectral cameras are becoming more popular, mainly due to remote sensing application. Thus, the use of this kind of camera in ground and underwater applications could provide a large amount of data to improve the capabilities of the UDCP/DCP.

The problem of restoration using only images is still challenging. The fusion with other sensors as Sonar/Radar that provides depth information in participating medium is an interesting possibility. The fusion with inertial sensors [Lobo and Dias, 2007] also could improve the estimation of the movement of the camera and, consequently, the depth estimation.

Another important issue to be addressed is the definition of reliable metrics to the image restoration problem. We call attention that the adopted metrics present important limitation. For example, only the FSIMc [Zhang et al., 2011] deals, at least partially, with the problem of color distortion. Indeed, the most adopted metric, *i.e.* contrast, is sensible to noise and distortions produced by the restoration.

The field of image restoration remains divided into different research areas, *e.g.* computer vision, image processing, computer graphics, pattern recognition, robotics and optics. In order to unify the field for future development, a common set of validation benchmarks that provide ground-truth for a comparative analysis needs to be created. This unification will increase the impact and relevance of the field, as occurs in other fields like dense stereo [Scharstein and Szeliski, 2002] and optical flow [Baker et al., 2011].

Bibliography

Ahlen, J. (2003). *Colour Correction of Underwater Images Using Spectral Data*. Ph.D. dissertation, Uppsala University - Sweden, Germany.

Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). Enhancing underwater images and videos by fusion. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 81–88. ISSN 1063-6919.

Ancuti, C. O. and Ancuti, C. (2013). Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282. ISSN 1057-7149.

Ancuti, C. O., Ancuti, C., and Bekaert, P. (2010). Effective single image dehazing by fusion. In *IEEE International Conference on Image Processing (ICIP)*, pages 3541–3544. ISSN 1522-4880.

Ancuti, C. O., Ancuti, C., Haber, T., and Bekaert, P. (2011). Fusion-based restoration of the underwater images. In *IEEE International Conference on Image Processing (ICIP)*, pages 1557–1560. ISSN 1522-4880.

Anthoni, J. F. (2003). The plankton balance hypothesis - two opposing forces in the plankton soup nurture and kill. http://www.seafriends.org.nz/enviro/plankton/balance.htm.

Armangue, X., Pages, J., Salvi, J., and Batlle, J. (2001). Comparative survey on fundamental matrix estimation. In *IX Spanish Symposium on Pattern Recognition and Image Analysis*, pages 227–232.

Arnold-Bos, A., Malkasse, J. P., and Kerven, G. (2005). A pre-processing framework for automatic underwater images denoising. In *European Conference on Propagation and Systems*.

Arredondo, M. A., Lebart, K., and Lane, D. (2004). Optical flow using textures. *Pattern Recognition Letters (PRL)*, 25(4):449–457. ISSN 0031-9007.

Aydin, T., Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2008). Dynamic range independent image quality assessment. *ACM Transactions on Graphics (TOG)*, 27(3):69:1–69:10. ISSN 0730-0301.

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., and Szeliski, R. (2011). A database and evaluation methodology for optical flow. *International Journal of Computer Vision (IJCV)*, 92(1):1–31. ISSN 0920-5691.

Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision (IJCV)*, 12:43–77. ISSN 1051-4651.

Barros, W. (2010). *Uma Abordagem Automática para Restauração de Imagens em Cenas Subaquáticas (In Portuguese)*. Ph.D. dissertation, Computer Science Department - Universidade Federal de Minas Gerais(UFMG), Brazil.

Bazeille, S., Quidu, I., Jaulin, L., and Malkasse, J. P. (2006). Automatic underwater image pre-processing. In *Proceedings of the Caracterisation du Milieu Marin (CMM '06)*.

Beardsley, P., Zisserman, A., and Murray, D. (1994). Navigation using affine structure from motion. In Eklundh, J.-O., editor, *European Conference on Computer Vision (ECCV)*, volume 801 of *Lecture Notes in Computer Science*, pages 85–96. Springer-Verlag.

Bellingham, J. G. and Rajan, K. (2007). Robotics in remote and hostile environments. *Science*, 318(5853):1098–1102. ISSN 0036-8075.

Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C. (2000). Image inpainting. In *ACM Special Interest Group on Computer Graphics (SIGGRAPH)*, pages 417–424. ISSN 0097-8930.

Bieganski, W. and Kasinski, A. (2014). Image acquisition in an underwater vision system with NIR and VIS illumination. *Computer Science & Information Technology*, 4(1):215–224. ISSN 2074-9007.

Black, M. J. and Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding (CVIU)*, 63(1):75–104. ISSN 1077-3142.

Bonin, F., Burguera, A., and Oliver, G. (2011). Imaging systems for advanced underwater vehicles. *Journal of Marine Research*, VIII(1):65–86. ISSN 0022-2402.

Botelho, S., Drews-Jr, P., Oliveira, G., and Figueiredo, M. (2009). Visual odometry and mapping for underwater autonomous vehicles. In *IEEE Latin American Robotics Symposium (LARS)*, pages 1–6.

Bouguet, J.-Y. (2013). Camera calibration toolbox for matlab. Web Page - http://www.vision.caltech.edu/bouguetj/calib_doc/.

Brandou, V., Allais, A.-G., Perrier, M., Malis, E., Rives, P., Sarrazin, J., and Sarradin, P.-M. (2007). 3D reconstruction of natural underwater scenes using the stereovision system IRIS. In *IEEE/OES OCEANS*, pages 1–6.

Bryson, M., Johnson-Roberson, M.and Pizarro, O., and Williams, S. B. (2012). Colour-consistent structure-from-motion models using underwater imagery. In *Robotics: Science and Systems*, pages 5:1–5:8. ISSN 2330-765X.

Busch, C. and Debes, E. (1998). Wavelet transform for analyzing fog visibility. *IEEE Intel. Systems and their Applic.*, 13(6):66–71. ISSN 1094-7167.

Butler, D. J., Wulff, J., Stanley, G. B., and Black, M. J. (2012). A naturalistic open source movie for optical flow evaluation. In *European Conference on Computer Vision (ECCV)*, pages 611–625.

Caraffa, L. and Tarel, J.-P. (2012). Stereo reconstruction and contrast restoration in daytime fog. In *Asian Conference on Computer Vision (ACCV)*, volume IV, pages 13–25.

Caraffa, L. and Tarel, J.-P. (2013). Markov random field model for single image defogging. In *IEEE Intelligent Vehicles Symposium (IVS)*, pages 994–999. ISSN 1931-0587.

Caraffa, L. and Tarel, J.-P. (2014). Combining stereo and atmospheric veil depth cues for 3d reconstruction. *IPSJ Transactions on Computer Vision and Applications*, 6:1–11. ISSN 1882-6695.

Carlevaris-Bianco, N., Mohan, A., and Eustice, R. (2010). Initial results in underwater single image dehazing. In *MTS/IEEE OCEANS*, pages 1–8.

Carr, P. and Hartley, R. (2009). Improved single image dehazing using geometry. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 103–110.

Chao, L. and Wang, M. (2010). Removal of water scattering. In *International Conference on Computer Engineering and Technology (ICCET)*, volume 2, pages 35–39.

Chavez-Jr, P. (1996). Image-based atmospheric corrections-revisited and improved. *Photomgrammetric Engineering & Remote Sensing*, 62(9):1025–1036. ISSN 0099-1112.

Chen, Z., Wang, H., Shen, J., Li, X., and Xu, L. (2014). Region-specialized underwater image restoration in inhomogeneous optical environments. *Optik*, 125(9):2090–2098. ISSN 0030-4026.

Chiang, J. and Chen, Y. (2012). Underwater image enhancement by wavelength compensation and dehazing. *IEEE Transactions on Image Processing*, 21(4):1756–1769. ISSN 1057-7149.

Choset, H., Lynch, K. M., Hutchinson, S., Kantor, G. A., Burgard, W., Kavraki, L. E., and Thrun, S. (2005). *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press.

Codevilla, F., Botelho, S., Drews-Jr, P., Filho, N. D., and Gaya, J. (2014). Underwater single image restoration using dark channel prior. In *Symposium on Automation and Computation for Naval, Offshore and Subsea (NAVCOMP)*.

Coleman, T. and Li, Y. (1996). An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6:418–445. ISSN 1052-6234.

Concha, A. and Civera, J. (2015). DPPTAM: Dense piecewise planar tracking and mapping from a monocular sequence. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. ISSN 2153-0858.

Concha, A., Drews-Jr, P., Campos, M., and Civera, J. (2015). Real-time localization and dense mapping in underwater environments from a monocular sequence. In *MTS/IEEE OCEANS*, pages 1–5.

Cozman, F. and Krotkov, E. (1997). Depth from scattering. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 801–806. ISSN 1063-6919.

Darles, E., Crespin, B., Ghazanfarpour, D., and Gonzato, J.-C. (2011). A survey of ocean simulation and rendering techniques in computer graphics. *Computer Graphics Forum*, 30(1):43–60. ISSN 0167-7055.

Ding, M. and Tong, R. (2012). Efficient dark channel based image dehazing using quadtrees. *Science China Information Sciences*, pages 1–9. ISSN 1674-733X.

Drews-Jr, P., Kuhn, V., and Gomes, S. C. P. (2013). Controlling a system for underwater visual inspection. *Marine Systems & Ocean Technology*, 8:119–126. ISSN 1679-3962.

Duan, X., Sun, H., and Peng, L. (2013). Riemannian means on special euclidean group and unipotent matrices group. *The Scientific World Journal*, 2013:1–9. ISSN 2356-6140.

Engel, J., Schöps, T., and Cremers, D. (2014). LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision (ECCV)*, pages 834–849.

Fang, M., Takauj, H., Kaneko, S., and Watanabe, H. (2009). Robust optical flow estimation for underwater images. In *International Symposium on Optomechatronic Technologies*.

Fang, S., Zhan, J., Y.Cao, and Rao, R. (2010). Improved single image dehazing using segmentation. In *IEEE International Conference on Image Processing (ICIP)*, pages 3589–3592. ISSN 1522-4880.

Fattal, R. (2008). Single image dehazing. *ACM Transactions on Graphics (TOG)*, 27(3-72). ISSN 0730-0301.

Feng, C., Zhuo, S., Zhang, X., Shen, L., and Susstrunk, S. (2013). Near-infrared guided color image dehazing. In *IEEE International Conference on Image Processing (ICIP)*, pages 2363–2367. ISSN 1522-4880.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395. ISSN 0001-0782.

Fraser, R. and Kaufman, Y. (1985). The relative importance of aerosol scattering and absorption in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(5):625–633. ISSN 0196-2892.

Fuchs, C., Heinz, M., Levoy, M., Seidel, H.-P., and Lensch, H. (2008). Combining confocal imaging and descattering. *Computer Graphics Forum*, 27(4):1245–1253. ISSN 1467-8659.

Gander, W., Golub, G. H., and Strebel, R. (1994). Least-squares fitting of circles and ellipses. *BIT Numer. Math.*, 34(4):558–578. ISSN 0006-3835.

Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361. ISSN 1063-6919.

Gibbons, J. D. and Chakraborti, S. (2010). *Nonparametric Statistical Inference*. Statistics: Textbooks and Monographs. Chapman and Hall/CRC, 5th edition.

Gibson, K. and Nguyen, T. (2013). Fast single image fog removal using the adaptive wiener filter. In *IEEE International Conference on Image Processing (ICIP)*, pages 714–718. ISSN 1522-4880.

Gibson, K., Vo, D., and Nguyen, T. (2012). An investigation of dehazing effects on image and video coding. *IEEE Transactions on Image Processing*, 21(2):662–673. ISSN 1057-7149.

Gonzalez, R. and Woods, R. (2007). *Digital Image Processing*. Prentice Hall, 3rd edition.

Grabner, M. and Kvicera, V. (2014). Multiple scattering in rain and fog on free-space optical links. *Journal of Lightwave Technology*, 32(3):513–520. ISSN 0733-8724.

Grewe, L. and Brooks, R. (1998). Atmospheric attenuation reduction through multisensor fusion. In *Proc. SPIE*, volume 3376, pages 102–109. ISSN 0277-786X.

Groszek, M. and Allebach, J. (2005). Haze removal for image enhancement. In *Proc. SPIE*, volume 5678, pages 254–265. ISSN 0277-786X.

Gupta, M., Narasimhan, S., and Schechner, Y. (2008). On controlling light transport in poor visibility environments. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. ISSN 1063-6919.

Gutierrez, D., Seron, F. J., Munoz, A., and Anson, O. (2008). Visualizing underwater ocean optics. *Computer Graphics Forum*, 27:547–556. ISSN 1467-8659.

Hadjimitsis, D., Clayton, C., and Hope, V. (2004). An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs. *IJRS*, 25(18):3651–3674. ISSN 01431161.

Handa, A., Whelan, T., McDonald, J., and Davison, A. (2014). A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1524–1531. ISSN 1050-4729.

Hartley, R. I. and Sturm, P. (1997). Triangulation. *CVUI*, 68(2):146–157. ISSN 1077-3142.

Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition.

Haussecker, H. and Fleet, D. (2001). Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(6):661–673. ISSN 0162-8828.

Hautière, N., Bigorgne, E., Bossu, J., and Aubert, D. (2008a). Meteorological conditions processing for vision-based traffic monitoring. In *The Eighth International Workshop on Visual Surveillance*.

Hautière, N., Tarel, J. P., and Aubert, D. (2007). Towards fog-free in-vehicle vision systems through contrast restoration. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. ISSN 1063-6919.

Hautière, N., Tarel, J. P., and Aubert, D. (2010). Mitigation of visibility loss for advanced camera-based driver assistance. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):474–484. ISSN 1524-9050.

Hautière, N., Tarel, J.-P., Aubert, D., and Dumont, E. (2008b). Blind contrast enhancement assessment by gradient ratioing at visible edges. *Image Analysis & Stereology*, 27(2):87–95. ISSN 1854-5165.

He, K. (2011). *Single Image Haze Removal Using Dark Channel Prior*. Ph.D. dissertation, The Chinese University of Hong Kong, China.

He, K., Sun, J., and Tang, X. (2009). Single image haze removal using dark channel prior. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1956–1963. ISSN 1063-6919.

He, K., Sun, J., and Tang, X. (2010). Guided image filtering. In *European Conference on Computer Vision (ECCV)*, pages 1–14.

He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(12):2341–2353. ISSN 0162-8828.

He, R., Wang, Z., Fan, Y., and Feng, D. (2013). Multiple scattering model based single image dehazing. In *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 733–737.

Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203. ISSN 0004-3702.

Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics*, 13(2):435–475. ISSN 0090-5364.

Hudzietz, B. and Saripalli, S. (2011). An experimental evaluation of 3D terrain mapping with an autonomous helicopter. In *Conference on Unmanned Aerial Vehicle (UAVg)*, pages 1–6.

Iqbal, K., Odetayo, M., James, A., Salam, R., and Talib, A. (2010). Enhancing the low quality images using unsupervised colour correction method. In *IEEE Systems, Man, and Cybernetics Society*, pages 1703–1709. ISSN 1062-922X.

Jaffe, J. (1990). Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering (JOE)*, 15(2):101–111. ISSN 0364-9059.

Johnson-Roberson, M., Pizarro, O., Williams, S. B., and Mahon, I. (2010). Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *Journal of Field Robotics (JFR)*, 27(1):21–51. ISSN 1556-4967.

Joshi, N. and Cohen, M. (2010). Seeing mt. rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8.

Kaeli, J., Singh, H., Murphy, C., and Kunz, C. (2011). Improving color correction for underwater image surveys. In *MTS/IEEE OCEANS*, pages 1–6.

Kaufman, Y., Tanre, D., Dubovik, O., Karnieli, A., and Remer, L. (2001). Absorption of sunlight by dust as inferred from satellite and ground-based remote sensing. *Geophysical Research Letters*, 28(8):1479–1482. ISSN 1944-8007.

Kim, J., Jang, W., Park, Y., Lee, D., Sim, J., and Kim, C. (2012). Temporally coherent real-time video dehazing. In *IEEE International Conference on Image Processing (ICIP)*, pages 969–972. ISSN 1522-4880.

Kim, J., Jang, W., Sim, J., and Kim, C. (2013). Optimized contrast enhancement for real-time image and video dehazing. *Journal of Visual Communication and Image Representation*, 24(3):410–425. ISSN 1047-3203.

Kim, J., Sim, J., and Kim, C. (2011). Single image dehazing based on contrast enhancement. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1273–1276. ISSN 1520-6149.

Kocak, D. and Caimi, F. (2005). The current art of underwater imaging - with a glimpse of the past and vision of the future. *MTS Journal*, 39(3):5–26. ISSN 0025-3324.

Kocak, D., Dalgleish, F., Caimi, F., and Schechner, Y. (2008). A focus on recent developments and trends in underwater imaging. *MTS Journal*, 42(1):52–67. ISSN 0025-3324.

Koch, R., Pollefeys, M., and Van Gool, L. (2000). Realistic surface reconstruction of 3D scenes from uncalibrated image sequences. *The Journal of Visualization and Computer Animation*, 11(3):115–127. ISSN 1099-1778.

Kokhanovsky, A. and Zege, E. (2004). Scattering optics of snow. *Applied Optics*, 43(7):1589–1602. ISSN 1559-128X.

Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., Uyttendaele, M., and Lischinski, D. (2008). Deep photo: Model-based photograph enhancement and viewing. *ACM Transactions on Graphics (TOG)*, 27(5):116:1–116:10. ISSN 0730-0301.

Koschmieder, H. (1924). Theorie der horizontalen sichtweit. *Beitr. Phys. Freien Atm.*, 12:171–181.

Kratz, L. and Nishino, K. (2009). Factorizing scene albedo and depth from a single foggy image. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1701–1708. ISSN 1550-5499.

Kuhn, V. N., Drews-Jr, P., Gomes, S., Cunha, M., and Botelho, S. (2015). Automatic control of a ROV for inspection of underwater structures using a low-cost sensing. *Journal of the Brazilian Society of Mechanical Sciences and Engineering (JBSMSE)*, 37(1):361–374. ISSN 1678-5878.

Lavest, J., Rives, G., and Lapresté, J. (2000). Underwater camera calibration. In Vernon, D., editor, *European Conference on Computer Vision (ECCV)*, volume 1843 of *Lecture Notes in Computer Science*, pages 654–668. Springer Verlag.

Levin, A., Lischinski, D., and Weiss, Y. (2008). A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(2):228–242. ISSN 0162-8828.

Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., and Bolas, M. (2004). Synthetic aperture confocal imaging. *ACM Transactions on Graphics (TOG)*, 23(3):825–834. ISSN 0730-0301.

Lhuillier, M. and Quan, L. (2005). A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(3):418–433. ISSN 0162-8828.

Li, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, 3rd edition. ISBN: 1848002785.

Liang, J., Ren, L., Qu, E., Hu, B., and Wang, Y. (2014). Method for enhancing visibility of hazy images based on polarimetric imaging. *Photonics Research*, 2(1):38–44. ISSN 1559-128X.

Lobo, J. and Dias, J. (2007). Relative pose calibration between visual and inertial sensors. *International Journal of Robotics Research (IJRR)*, 26(6):561–575. ISSN 0278-3649.

Lourakis, M. and Argyros, A. (2009). SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM TOMS*, 36(1):1–30. ISSN 0098-3500.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110. ISSN 0920-5691.

Lu, H., Li, Y., and Serikawa, S. (2013). Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction. In *IEEE International Conference on Image Processing (ICIP)*, pages 3412–3416. ISSN 1522-4880.

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Imaging Understanding Workshop*, pages 121–130.

Lv, X., Chen, W., and Shen, I. (2010). Real-time dehazing for image and video. In *Pacific Conference on Computer Graphics and Applications*, pages 62–69.

Ma, Y., Soatto, S., Kosecka, J., and Sastry, S. S. (2006). *An Invitation to 3-D Vision: From Images to Geometric Models*, volume 26 of *Interdisciplinary Applied Mathematics*. Spring-Verlag.

Madjidi, H. and Negahdaripour, S. (2006). On robustness and localization accuracy of optical flow computation for underwater color images. *Computer Vision and Image Understanding (CVIU)*, 104(1):61–76. ISSN 1077-3142.

Mahiny, A. and Turner, B. (2007). A comparison of four common atmospheric correction methods. *Photogrammetric Engineering & Remote Sensing*, 73(4):361–368. ISSN 0099-1112.

Mahon, I., Williams, S., Pizarro, O., and Johnson-Roberson, M. (2008). Efficient view-based SLAM using visual loop closures. *IEEE Transactions on Robotics*, 24(5):1002–1014. ISSN 1552-3098.

Makarau, A., Richter, R., Muller, R., and Reinartz, P. (2014). Haze detection and removal in remotely sensed multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–11. ISSN 0196-2892.

Mantiuk, R., Kim, K., Rempel, A., and Heidrich, W. (2011). HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics (TOG)*, 30(4-40):1–14. ISSN 0730-0301.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441. ISSN 0036-1399.

McCartney, E. (1976). *Optics of the Atmosphere: Scattering by Molecules and Particles.* Pure and Applied Optics. John Wiley & Sons Inc.

McGlamery, B. (1980). A computer model for underwater camera systems. In *Proc. of SPIE*, volume 208-VI, pages 221–231.

Meinhardt-Llopis, E., Pérez, J. S., and Kondermann, D. (2013). Horn-Schunck Optical Flow with a Multi-Scale Strategy. *Image Processing On Line*, 3:151–172. ISSN 2105-1232.

Meng, G., Wang, Y., Duan, J., Xiang, S., and Pan, C. (2013). Efficient image dehazing with boundary constraint and contextual regularization. In *IEEE International Conference on Computer Vision (ICCV)*, pages 617–624. ISSN 1550-5499.

Middleton, W. (1952). *Vision through the atmosphere.* University of Toronto Press.

Moakher, M. (2002). Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16. ISSN 0036-1399.

Mobley, C. D. (1994). *Light and Water: Radiative Transfer in Natural Waters.* Academic Press.

Molnár, J., Chetverikov, D., and Fazekas, S. (2010). Illumination-robust variational optical flow using cross-correlation. *Computer Vision and Image Understanding (CVIU)*, 114(10):1104–1114. ISSN 1077-3142.

Mortazavi, H., Oakley, J., and Barkat, B. (2012). Mitigating the effect of optical back-scatter in multispectral underwater imaging. In *IEEE Imaging Systems and Techniques (IST)*, pages 633–636.

Mortazavi, H., Oakley, J., and Barkat, B. (2013). Mitigating the effect of optical back-scatter in multispectral underwater imaging. *Measurement Science and Technology*, 24(7). ISSN 0957-0233.

Murray, R. M., Sastry, S. S., and Zexiang, L. (1994). *A Mathematical Introduction to Robotic Manipulation.* CRC Press, Inc., Boca Raton, FL, USA, 1st edition. ISBN 0849379814.

Namer, E., Shwartz, S., and Schechner, Y. (2009). Skyless polarimetric calibration and visibility enhancement. *Optics Express*, 17(2):472–493. ISSN 1559-128X.

Narasimhan, S. and Nayar, S. (2000). Chromatic framework for vision in bad weather. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 598–605. ISSN 1063-6919.

Narasimhan, S. and Nayar, S. (2002). Vision and the atmosphere. *International Journal of Computer Vision (IJCV)*, 48(3):233–254. ISSN 0920-5691.

Narasimhan, S. and Nayar, S. (2003a). Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25(6):713–724. ISSN 0162-8828.

Narasimhan, S. and Nayar, S. (2003b). Interactive deweathering of an image using physical models. In *IEEE International Conference on Computer Vision (ICCV) - Workshop on Color and Photometric Methods in Computer Vision.*

Narasimhan, S. and Nayar, S. (2005). Structured light methods for underwater imaging: light stripe scanning and photometric stereo. In *MTS/IEEE OCEANS*, volume 3, pages 2610–2617.

Narasimhan, S., Nayar, S., Sun, B., and Koppal, S. (2005). Structured light in scattering media. In *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 420–427. ISSN 1550-5499.

Narasimhan, S. G., Gupta, M., Donner, C., Ramamoorthi, R., Nayar, S. K., and Jensen, H. W. (2006). Acquiring scattering properties of participating media by dilution. *ACM Transactions on Graphics (TOG)*, 25(3):1003–1012. ISSN 0730-0301.

Nascimento, E. (2008). Estimacao automatica de parametros de modelos para restauracao de imagens de cenas subaquaticas (In Portuguese). Master's thesis, Computer Science Department - Universidade Federal de Minas Gerais(UFMG), Brazil.

Nascimento, E., Campos, M., and Barros, W. (2009). Stereo based structure recovery of underwater scenes from automatically restored images. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 330–337. ISSN 1530-1834.

Nayar, S., Krishnan, G., Grossberg, M., and Raskar, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics (TOG)*, 25(3):935–944. ISSN 0730-0301.

Nayar, S. and Narasimhan, S. (1999). Vision in bad weather. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 820–827. ISSN 1550-5499.

Negahdaripour, S. (1998). Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 20(9):961–979. ISSN 0162-8828.

Negahdaripour, S. and Madjidi, H. (2003). Robust optical flow estimation using underwater color images. In *MTS/IEEE OCEANS*, volume 3, pages 2309–2316.

Negahdaripour, S., Zhang, H., and Han, X. (2002). Investigation of photometric stereo method for 3-d shape recovery from underwater imagery. In *MTS/IEEE OCEANS*, volume 2, pages 1010–1017.

Nery, M., Machado, A., Campos, M., Padua, F., Carceroni, R., and Queiroz-Neto, J. (2005). Determining the appropriate feature set for fish classification tasks. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 173–180. ISSN 1530-1834.

Newcombe, R. A., Lovegrove, S., and Davison, A. (2011). DTAM: Dense tracking and mapping in real-time. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2320–2327. ISSN 1550-5499.

Nishino, K., Kratz, L., and Lombardi, S. (2012). Bayesian defogging. *International Journal of Computer Vision (IJCV)*, 98(3):263–278. ISSN 0920-5691.

Oakley, J. and Bu, H. (2007). Correction of simple contrast loss in color images. *IEEE Transactions on Image Processing*, 16(2):511–522. ISSN 1057-7149.

Oakley, J. and Marceta, Z. (2011). An automatic hd defogging system. *The Best of IBC and IET*, 3:1–7. ISSN 2041-5923.

Oakley, J. and Satherley, B. (1998). Improving image quality in poor visibility conditions using a physical model for contrast degradation. *IEEE Transactions on Image Processing*, 7(2):167–179. ISSN 1057-7149.

Ohta, N. (1989). Optical flow detection by color images. In *IEEE International Conference on Image Processing (ICIP)*, pages 801–805. ISSN 1522-4880.

Pavlic, M., Rigoll, G., and Ilic, S. (2013). Classification of images in fog and fog-free scenes for use in vehicles. In *IEEE Intelligent Vehicles Symposium (IVS)*, pages 481–486. ISSN 1931-0587.

Persistence of Vision Pty. Ltd. (2004). Persistence of vision (tm) raytracer. http://www.povray.org/.

Pomerleau, D. (1997). Visibility estimation from a moving vehicle using the RALPH vision system. In *International Conference on Intelligent Transportation Systems (ITSC)*, pages 906–911.

Preisendorfer, R. (1976). *Hydrologic optics*. U.S. Dept. of Commerce - National Oceanic and Atmospheric Administration.

Queiroz-Neto, J., Carceroni, R., Barros, W., and Campos, M. (2004). Underwater stereo. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 170–177. ISSN 1530-1834.

Queiroz-Neto, J. P. (2005). *Modelagem automática de cenas com iluminação local a partir de imagens*. Ph.D. dissertation, Computer Science Department - Universidade Federal de Minas Gerais(UFMG), Brazil.

Ridao, P., Carreras, M. nd Ribas, D., and Garcia, R. (2010). Visual inspection of hydroelectric dams using an autonomous underwater vehicle. *Journal of Field Robotics (JFR)*, 27(6):759–778. ISSN 1556-4967.

Rodríguez-Telles, F., Pérez-Alcocer, R., Maldonado-Ramírez, A., Torres-Mendez, L., Dey, B., and Martínez-Garcia, E. (2014). Vision-based reactive autonomous navigation with obstacle avoidance: Towards a non-invasive and cautious exploration of marine habitat. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3813–3818. ISSN 1050-4729.

Roser, M., Dunbabin, M., and Geiger, A. (2014). Simultaneous underwater visibility assessment, enhancement and improved stereo. In *IEEE International Conference on Robotics and Automation (ICRA)*. ISSN 1050-4729.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880. ISSN 0162-1459.

Sarafraz, A., Negahdaripour, S., and Schechner, Y. (2009). Enhancing images in scattering media utilizing stereovision and polarization. In *Workshop on Applications of Computer Vision (WACV)*, pages 1–8. ISSN 1550-5790.

Sarafraz, A., Negahdaripour, S., and Schechner, Y. (2010). Performance assessment in solving the correspondence problem in underwater stereo imagery. In *MTS/IEEE OCEANS*, pages 1–7.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision (IJCV)*, 47(1/2/3):7–42. ISSN 0920-5691.

Schaul, L., Fredembach, C., and Susstrunk, S. (2009). Color image dehazing using the near-infrared. In *IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632. ISSN 1522-4880.

Schechner, Y. and Karpel, N. (2004). Clear underwater vision. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 536–543. ISSN 1063-6919.

Schechner, Y. and Karpel, N. (2005). Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering (JOE)*, 30(3):570–587. ISSN 0364-9059.

Schechner, Y., Narasimhan, S., and Nayar, S. (2001). Instant dehazing of images using polarization. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 325–332. ISSN 1063-6919.

Schechner, Y., Narasimhan, S., and Nayar, S. (2003). Polarization-Based Vision through Haze. *Applied Optics*, 42(3):511–525. ISSN 1559-128X.

Schilling, Z., Leary, R., and Stewart, C. (2010). Method for correction of underwater imagery with non-uniform illumination. In *Undergraduate Research Symposium - Rensselaer Polytechnic Institute*.

Sedlazeck, A. and Koch, R. (2011). Simulating deep sea underwater images using physical models for light attenuation, scattering, and refraction. In Eisert, P., Hornegger, J., and Polthier, K., editors, *Vision, Modeling, and Visualization*, pages 49–56. Eurographics Association.

Sedlazeck, A. and Koch, R. (2012). Perspective and non-perspective camera models in underwater imaging - overview and error analysis. In Dellaert, F., Frahm, J.-M., Pollefeys, M., Leal-Taixé, L., and Rosenhahn, B., editors, *Outdoor and Large-Scale Real-World Scene Analysis*, volume 7474 of *Lecture Notes in Computer Science*, pages 212–242. Spring-Verlag.

Serikawa, S. and Lu, H. (2014). Underwater image dehazing using joint trilateral filter. *Computers & Electrical Engineering*, 40(1):41–50. ISSN 0045-7906.

Shihavuddin, A., Gracias, N., Garcia, R., Gleason, A., and Gintert, B. (2013). Image-based coral reef classification and thematic mapping. *Remote Sensing*, 5(4):1809–1841. ISSN 2072-4292.

Shwartz, S., Namer, E., and Schechner, Y. (2006). Blind haze separation. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1984–1991. ISSN 1063-6919.

Soatto, S., Frezza, R., and Perona, P. (1996). Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3):393–413. ISSN 0018-9286.

Stackpole, E., Lang, D., and Borri, M. (2015). Openrov 2.7 mini observation class rov. Web Page - http://www.openrov.com/products/2-7.html.

Sturm, J., Engelhard, N., Endres, F., Burgard, W., and Cremers, D. (2012). A benchmark for the evaluation of RGB-D SLAM systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 573–580. ISSN 2153-0858.

Sun, B., Ramamoorthi, R., Narasimhan, S. G., and Nayar, S. K. (2005). A practical analytic single scattering model for real time rendering. *ACM Transactions on Graphics (TOG)*, 24(3):1040–1049. ISSN 0730-0301.

Sun, D. (2013). *From Pixels to Layers: Joint Motion Estimation and Segmentation.* Ph.D. dissertation, Department of Computer Science at Brown University, USA.

Sun, D., Roth, S., and Black, M. (2010). Secrets of optical flow estimation and their principles. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2432–2439. ISSN 1063-6919.

Sun, D., Roth, S., and Black, M. J. (2014). A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision (IJCV)*, 106(2):115–137. ISSN 0920-5691.

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Texts in Computer Science. Springer-Verlag.

Tan, K. and Oakley, J. (2000). Enhancement of color images in poor visibility conditions. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 788–791. ISSN 1522-4880.

Tan, K. and Oakley, J. (2001). Physics-based approach to color image enhancement in poor visibility conditions. *Journal of the Optical Society of America (JOSA)*, 18(10):2460–2467. ISSN 1084-7529.

Tan, R. (2008). Visibility in bad weather from a single image. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. ISSN 1063-6919.

Tarel, J. and Hautière, N. (2009). Fast visibility restoration from a single color or gray level image. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2201–2208. ISSN 1550-5499.

Tarel, J.-P., Hautière, N., Caraffa, L., Cord, A., Halmaoui, H., and Gruyer, D. (2012). Vision enhancement in homogeneous and heterogeneous fog. *IEEE ITS Magazine*, 4(2):6–20. ISSN 1939-1390.

Thoeni, K., Giacomini, A., Murtagh, R., and Kniest, E. (2014). A comparison of multi-view 3d reconstruction of a rock wall using several cameras and a laser scanner. In *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XL-5, pages 573–580. ISSN 1682-1750.

Tordoff, B. and Murray, D. (2005). Guided-MLESAC: faster image transform estimation by using matching priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(10):1523–1535. ISSN 0162-8828.

Torr, P. and Zisserman, A. (1998). Robust computation and parametrization of multiple view relations. In *IEEE International Conference on Computer Vision (ICCV)*, pages 727–732. ISSN 1550-5499.

Torr, P. and Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding (CVIU)*, 78(1):138–156. ISSN 1077-3142.

Torr, P. H. S. (2002). Structure and motion toolkit in matlab. Technical report, Microsoft Corporation, USA.

Treibitz, T. and Schechner, Y. (2006). Instant 3Descatter. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1861–1868. ISSN 1063-6919.

Treibitz, T. and Schechner, Y. (2009a). Active polarization descattering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(3):385–399. ISSN 0162-8828.

Treibitz, T. and Schechner, Y. (2009b). Polarization: Beneficial for visibility enhancement? In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 525–532. ISSN 1063-6919.

Treibitz, T. and Schechner, Y. (2009c). Recovery limits in pointwise degradation. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8.

Treibitz, T. and Schechner, Y. (2012). Turbid scene enhancement using multi-directional illumination fusion. *IEEE Transactions on Image Processing*, 21(11):4662–4667. ISSN 1057-7149.

Treibitz, T., Schechner, Y., and Singh, H. (2008). Flat refractive geometry. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. ISSN 1063-6919.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. (2000). Bundle adjustment - a modern synthesis. In Triggs, B., Zisserman, A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer Berlin Heidelberg.

Tripathi, A. and Mukhopadhyay, S. (2012). Single image fog removal using anisotropic diffusion. *IET Image Processing*, 6(7):966–975. ISSN 1751-9659.

Trucco, E. and Olmos-Antillon, A. (2006). Self-tuning underwater image restoration. *IEEE Journal of Oceanic Engineering (JOE)*, 31(2):511–519. ISSN 0364-9059.

Van der Berg, P. (2012). Costa Rica diving at Guanacaste Coast Pacific. Youtube - http://www.youtube.com/watch?v=vKduIggPd38.

Vasilescu, I., Detweiler, C., and Rus, D. (2010). Color-accurate underwater imaging using perceptual adaptive illumination. In *Robotics: Science and Systems*, pages 3:1–3:8. ISSN 2330-765X.

Vasilescu, I., Detweiler, C., and Rus, D. (2011). Color-accurate underwater imaging using perceptual adaptive illumination. *Autonomous Robots*, 31(2-3):285–296. ISSN 0929-5593.

Walther, D., Edgington, D., and Koch, C. (2004). Detection and tracking of objects in underwater video. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 544–549. ISSN 1063-6919.

Wang, Y., Fan, C., and Chang, C. (2013). Accurate depth estimation for image defogging using markov random field. In *Proc. SPIE*, volume 8768, pages 1–5. ISSN 0277-786X.

Wang, Y. and Wu, B. (2010). Fast clear single underwater image. In *International Conference on Computational Intelligence and Software Engineering (CiSE)*, pages 1–4.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612. ISSN 1057-7149.

Wang, Z., Sheikh, H., and Bovik, A. (2002). No-reference perceptual quality assessment of JPEG compressed images. In *IEE ICIP*, volume 1, pages 477–480. ISSN 1522-4880.

Werner, T. and Pajdla, T. (2001). Cheirality in epipolar geometry. In *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 548–553. ISSN 1550-5499.

Xiao, J., Hays, J., Ehinger, K., Oliva, A., and Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 3485–3492. ISSN 1063-6919.

Zhang, B. and Wang, L. (2011). Local color transfer based on dark-channel dehazing for visible/infrared image fusion. In *Proc. SPIE*, volume 8056, pages 1–7. ISSN 0277-786X.

Zhang, J., Li, L., Yang, G., Zhang, Y., and Sun, J. (2010). Local albedo-insensitive single image dehazing. *The Visual Computer*, 26(6-8):761–768. ISSN 0178-2789.

Zhang, L. and Li, H. (2012). SR-SIM: A fast and high performance IQA index based on spectral residual. In *IEEE International Conference on Image Processing (ICIP)*, pages 1473–1476. ISSN 1522-4880.

Zhang, L., Zhang, D., Mou, X., and Zhang, D. (2011). Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386. ISSN 1057-7149.

Zhang, S. and Negahdaripour, S. (2002). 3D shape recovery of planar and curved surfaces from shading cues in underwater images. *IEEE Journal of Oceanic Engineering (JOE)*, 27(1):100–116. ISSN 0364-9059.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(11):1330–1334. ISSN 0162-8828.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Heckbert, P. S., editor, *Graphics Gems IV*, pages 474–485. Academic Press Professional, Inc. ISBN: 0-12-336155-9.

Zuliani, M. (2008). RANSAC toolbox for Matlab. Web Page - http://www.mathworks.com/matlabcentral/fileexchange/18555.

Appendix A

Structure from Motion

Most of the works on the structure-from-motion literature deal with the problem of estimate the fundamental matrix because of its generality, *i.e.* can be recovered by uncalibrated camera. As described in Hartley and Zisserman [2003], the essential matrix can be computed using classical techniques developed for estimate the fundamental matrix since both satisfy the same epipolar constraint. The main difference between both estimation are related to the enforcement of the constraints. For, whereas the fundamental matrix satisfies $det(F) = 0$, the essential matrix satisfies the additional condition that its two singular value are equal [Hartley and Zisserman, 2003].

Both fundamental matrix and essential matrix are related with the estimation using the discrete features, however there is the case continuous. The continuous case is particularly important when the correspondence is obtained by optical flow. The continous epipolar constraint does not depend upon the position of the points in the space, but only their projection and the motion parameters.

Armangue et al. [2001] provide a comparative study between algorithms to estimate the fundamental matrix. Following this study, the most accurate methods for real image are the algorithms for robust fitting of models. It is due to their capability to deal with data outliers. We adopted MSAC approach proposed by Torr and Zisserman [2000] for both discrete and continuous essential matrix. It is an improvement in classic RANSAC method [Fischler and Bolles, 1981] that change the cost function to be minimized to a M-estimator [Huber, 1985]. This new function improves the accuracy with no additional computational burden [Torr and Zisserman, 2000].

Algorithm 1 MSAC-Based Discrete Essential Matrix Estimation - Input: $\hat{\boldsymbol{x}}_H$, $\hat{\boldsymbol{x}}'_H$, N, T

```
for $i = 1$ to $N$ do
    $\Pi \leftarrow Random\_Pair\_Selection(\hat{\boldsymbol{x}}_H, \hat{\boldsymbol{x}}'_H)$
    $E(i) \leftarrow Essential\_LS\_Estimation(\Pi)$
    $e^2 \leftarrow Error\_Fitting(E(i), \hat{\boldsymbol{x}}_H, \hat{\boldsymbol{x}}'_H)$
    $C(i) \leftarrow M\_Estimator(e^2, T)$
end for
$BestE \leftarrow Smallest\_Error(C)$
$[Inliers, NewE] \leftarrow Restimate\_E(BestE, \hat{\boldsymbol{x}}_H, \hat{\boldsymbol{x}}'_H)$
$[P', R, t] \leftarrow Camera\_From\_Essential(Inliers, NewE)$
return $NewE$,$Inliers$,$P$,$R$,$t$
```

A.1 MSAC-Based Discrete Essential Matrix

The discrete method is summarized by the Alg. 1 that receives as input the set of points correspondence in two different images ($\hat{\boldsymbol{x}}_H$ and $\hat{\boldsymbol{x}}'_H$), the number of samples N, and the threshold T.

In each iteration of the main loop, the algorithm selects a random sample of eight correspondences, computed by the function *Random_Pair_Selection*. Therefore, the set of pair of points Π is obtained. The function *Essential_LS_Estimation*(Π) compute the essential matrix based on the classic eight-points least square minimization [Hartley and Zisserman, 2003]. The algorithm solves the minimization problem described as:

$$\underset{E}{\operatorname{argmin}} \sum_j \hat{\boldsymbol{x}}(j)'^T_H E \hat{\boldsymbol{x}}(j)_H. \tag{A.1}$$

This minimization problem may be solved using SVD decomposition [Hartley and Zisserman, 2003]. This function still deals with the singularity constraint to estimate a valid essential matrix, *i.e.* $det(E) = 0$. The squared error, e^2, for all set of points correspondence is computed by function *Error_Fitting*. Based on this error metric, we estimate the cost function, as defined in Torr and Zisserman [1998] (function $M\_Estimator$):

$$C = \sum_j p(e_j^2), \;\; where \;\; p(e_j^2) = \begin{cases} e_j^2, & e_j^2 < T^2 \\ T^2, & e_j^2 \geq T^2 \end{cases}. \tag{A.2}$$

We select the best solution over all the samples by using the function *Smallest_Error*, *i.e* the sample with lowest C error. The essential matrix is re-estimated from all correspondences classified as inliers (function *Restimate_E*). Several techniques can be applied to solve the estimation problem as least square minimization or iterative optimization [Hartley and Zisserman, 2003]. We adopted the least square because of its simplicity.

Finally, the camera matrices are retrieved from the essential matrix up to scale and a four-fold ambiguity (function *Camera_From_Essential*). We assumed the first projection matrix is $P = [I|0]$. In order to compute the second projection matrix P', the rotation matrix R and the skew-symmetric matrix $[t]_\times$ needs to be estimated. The SVD decomposition of the essential matrix $E = Udiag(1,1,0)V^T$ allows to compute the four possible solutions, as described by Hartley and Zisserman [2003]:

$$[t]_\times = UZU^T \ \ and \ \ R = UWV^T \ \ or \ \ UW^TV^T, \tag{A.3}$$

where the matrices W and Z are defined as [Hartley and Zisserman, 2003]:

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ \ and \ \ Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{A.4}$$

The factorization (Eq. A.3) determines the t up to scale, from $[t]_\times$. Since $[t]_\times t = 0$, it follows that $t = U(0,0,1)^T = u_3$, *i.e* the last column of U. Thus, we can derive the four possible choices for the second projection matrix:

$$P' = [UWV^T| + u_3] \ or \ [UWV^T| - u_3] \ or \ [UW^TV^T| + u_3] \ or \ [UW^TV^T| - u_3]. \tag{A.5}$$

The geometric interpretation of the four solution is shown in details in Hartley and Zisserman [2003]. Using a single point, we can test whether the point is on the front of both cameras. This procedure is sufficient to determine the correct solution.

The essential matrix suffers from some shortcomings that are explained in details by Ma et al. [2006]. They are an important limitation on the applicability of the method. We discuss about the two most important limitations in the context of our work. Firstly, the points need to be in "general position" in order for the eight point algorithm to work properly, *i.e.* the algorithm fail if these points form certain degenerate configuration, called critical surfaces. A case of practical importance occurs when all the points happen to lie on the same 2D plane [Ma et al., 2006].

Other issue is related to the motion requirement. Because of the structure of the essential matrix, $E = 0 \Leftrightarrow t = 0$, thus the estimation algorithm requires that the translation be different of zero, $t \neq 0$ [Ma et al., 2006]. Therefore, the algorithm to be

Algorithm 2 MSAC-Based Continuous Essential Matrix Estimation - Input: $\boldsymbol{x}_H$, $\boldsymbol{u}$, N, T

```
for $i = 1$ to $N$ do
    $\Pi \leftarrow Random\_Pair\_Selection(\boldsymbol{x}_H, \boldsymbol{u})$
    $E'(i) \leftarrow Eight\_Point\_Linear\_Algorithm(\Pi)$
    $e^2 \leftarrow Error\_Fitting(E'(i), \boldsymbol{x}_H, \boldsymbol{u})$
    $C(i) \leftarrow M\_Estimator(e^2, T)$
end for
$BestE' \leftarrow Smallest\_Error(C)$
$[Inliers, NewE'] \leftarrow Restimate\_CEM(BestE', \boldsymbol{x}_H, \boldsymbol{u})$
$[P, R, t] \leftarrow Camera\_From\_CEM(Inliers, NewE')$
return $NewE'$,$Inliers$,$P$,$R$,$t$
```

well conditioned needs sufficient movement of translation.

A.2 MSAC-Based Continuous Essential Matrix

Similarly to the case discrete, the continuous case is summarized by the Alg. 2 that receives as input the set of points $\boldsymbol{x}_H$ and image motion field approximated by the optical flow $\boldsymbol{u}$, the number of samples N, and the threshold T. Both $\boldsymbol{x}_H$ and $\boldsymbol{u}$ are normalized by the intrinsic parameter.

In each iteration of the main loop, the algorithm selects a random sample of eight correspondences, computed by the function *Random_Pair_Selection*. Therefore, the set of $\boldsymbol{x}_H$ and $\boldsymbol{u}$ is defined as Π. The function *Eight_Point_Linear_Algorithm*(Π) compute the continuous essential matrix as described in details by Ma et al. [2006]. Assuming the the continuous essential matrix is defined as:

$$E' = \left\{ \begin{bmatrix} \hat{v} \\ 0.5(\hat{w}\hat{v} + \hat{v}\hat{w}) \end{bmatrix} \middle| w \in \mathbb{R}^3, v \in \mathbb{S}^2 \right\} \subset \mathbb{R}^{6\times 3}, \tag{A.6}$$

we can define the function *Error_Fitting*$(E'(i), \boldsymbol{x}_H, \boldsymbol{u})$ by using the continuous epipolar constraint, defined by:

$$e^2 = \sum_j ||\boldsymbol{u}(j)^T \hat{v} \boldsymbol{x}_H(j) + \boldsymbol{x}_H(j)^T \hat{w}\hat{v} \boldsymbol{x}_H(j)||^2. \tag{A.7}$$

The function $M\_Estimator$ is defined as proposed by Torr and Zisserman [1998], detailed in Sec. A.1. We select the best solution over all estimated CEM by using the function *Smallest_Error*, *i.e* the sample with lowest C error. The continuous essential matrix is re-estimated from all correspondences classified as inliers (function

Restimate_ E) using the eight point linear algorithm [Ma et al., 2006]. Finally, function $Camera\_From\_CEM(Inliers, NewE')$ determine rigid body transformation $\in SE(3)$ by using the body velocities v and $w \in \mathfrak{se}(3)$ estimated using the CEM. Details about the function is provided in Sec. 3.3.2.

A.3 Comparative Study

We performed a comparative study to evaluate the capabilities of both methods to obtain the rigid body transformation using essential matrix (EM). The ICL-NUIM dataset [Handa et al., 2014] is adopted, it aims at benchmarking RGB-D, Visual Odometry and SLAM algorithms. Two different scenes (the living room and the office room scene) are provided with ground truth. The synthetic dataset consists of images obtained from camera trajectories in raytraced 3D models in POVRay [Persistence of Vision Pty. Ltd., 2004]. They also proposed a synthetic sensor noise [Handa et al., 2014], where they assume that the noise occurrs in the RGB image and in the depth image.

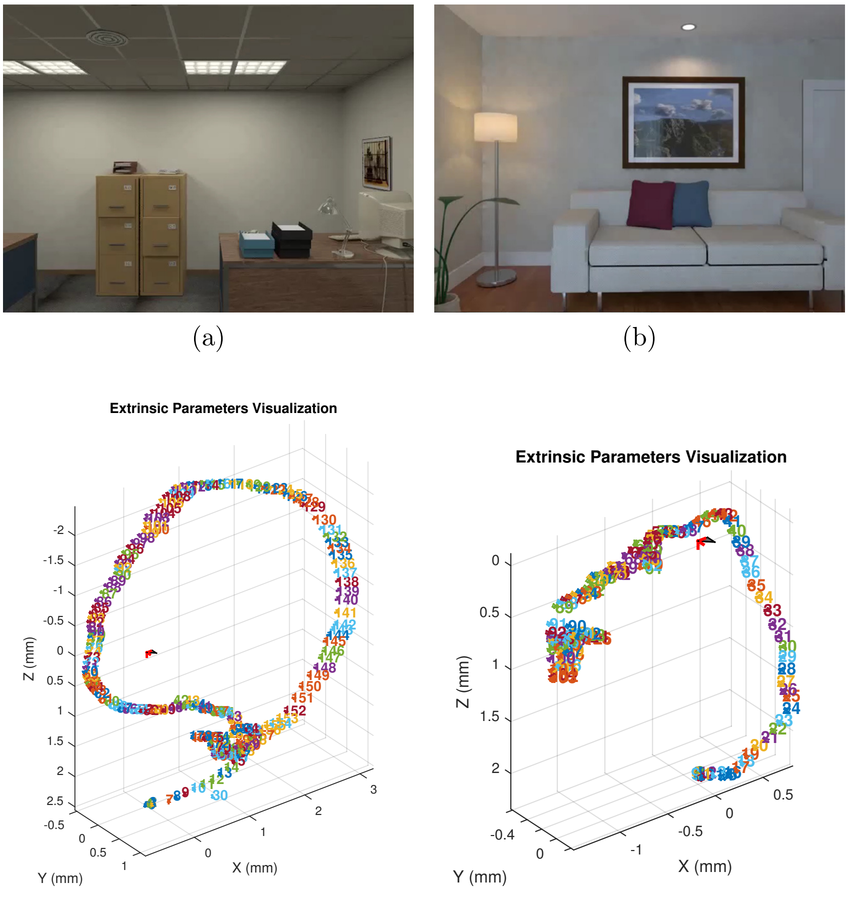
The sequence **living room** is composed of 882 frames at 30 Hz, while the sequence **office room** is composed of 1242 frames at 30 Hz. Fig. A.1 shows one sample frame and its trajectory from each sequence. We computed the optical flow between consecutive frames and, the first frame and the next frames on the sequences. The optical flow adopted is the Classic-NL [Sun et al., 2014], one of the state-of-art in terms of optical flow methods. One thousand points were randomly selected to be used in the estimation.

The comparative study evaluates the continuous essential matrix using the MSAC-approach detailed in Sec. A.2 using eight points, but also using nine and ten points. The idea is evaluate the sensibility of the method to an overdetermined estimation of the CEM. These methods were implemented by us. Moreover, we also evaluated the discrete case. We obtained results using the MSAC-based DEM algorithm, as described in Sec. A.1, based on the toolkit provided by Torr [2002]. Furthermore, RANSAC-based [Fischler and Bolles, 1981] and LMedS-based [Rousseeuw, 1984] DEM algorithms also provide results. We adopted the standard implementation provided by MATLAB for them.

We compared the estimated rigid body transformation with the ground truth. The error in translation E_t is obtained by using the Euclidean distance:

$$E_t = \sqrt{(t - t_g)^2}, \tag{A.8}$$

where $t \in \mathbb{R}^3$ is the estimated translation vector, and t_g is its ground truth. The error



(a) (b)

(c) (d)

Figure A.1. Sample image and the trajectory of each sequence from the ICL-NUIM dataset: a,c) office room, b,d) living room. We show the camera's pose in each seven frames to improve the trajectory visualization.

in rotation E_R is obtained using the Riemannian distance in $SO(3)$ [Moakher, 2002]:

$$E_R = \frac{1}{\sqrt{2}} \sqrt{(\log(R^T R_g)^2}, \tag{A.9}$$

where $R \in SO(3)$ is the estimated rotation matrix, and R_g is its ground truth. Furthemore, $\log(R)$ is defined as:

$$\log(R) = \begin{cases} 0 & \text{if } \theta = 0 \\ \frac{\theta}{2\sin\theta}(R - R^T) & \text{if } \theta \neq 0 \end{cases}, \tag{A.10}$$

where $\theta = acos\left(\frac{trace(R)-1}{2}\right)$ [Moakher, 2002]. Finally, the error E_P in the projective matrix $P \in SE(3)$ is defined as proposed in Duan et al. [2013]:

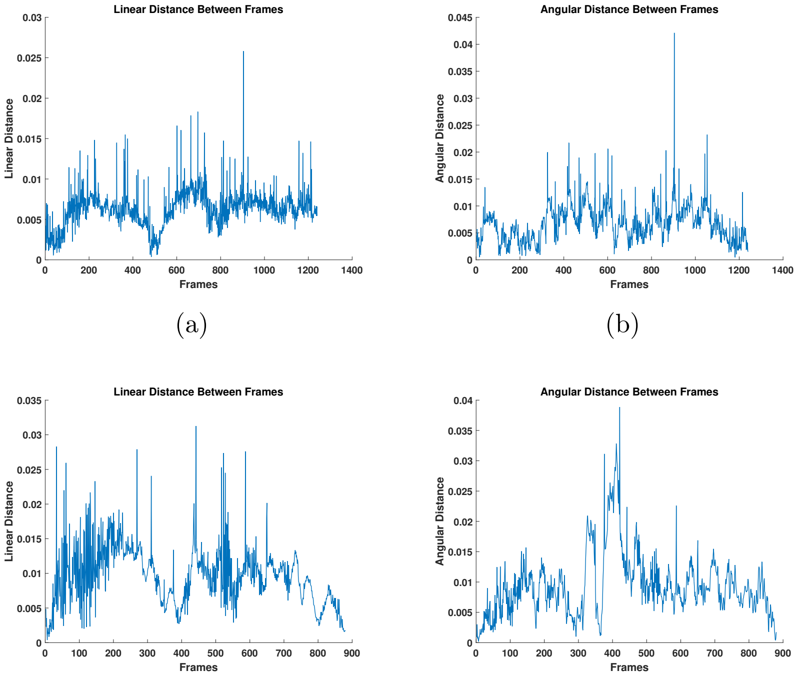
$$E_P = \sqrt{(E_t + E_R)}. \tag{A.11}$$

These three error metrics allow us to evaluate the quality of the estimation from CEM and DEM. Some additional procedures are adopted to guarantee a fair comparison. We repeated each method while it does not obtain more than 50% of inliers in terms of chirality constraint, *i.e.* the points needs to be in a positive depth. The parameters of the methods are hard to tune, thus we repeat at the maximum 25 times each execution of the methods.

Initially, we evaluate both datasets on consecutive frames, *i.e.* the transformation is estimated between consecutive frames. The linear and angular movements between frames are shown in Fig. A.2. It is possible to see that the movement is small between frames. Fig A.3 shows the obtained results in terms of error for both sequences, in each column. The algorithms based on CEM are able to provide a good estimation of rotation, however the quality of the translation vector is limited. The results obtained using DEM are also limited, both in rotation and translation. These facts impact in the transformation error, but the results for CEM algorithm are better due to the good estimation of the rotation matrix. Table A.1 shows the average and standard deviation of the errors. The results for CEM are consistent in terms of rotation, but the estimation for translation vector is not enough for the reconstruction. It is valid for all algorithms. The results obtained by DEM-RANSAC appears to be more smooth than the others, it is due to the small number of valid estimation in terms of chirality constraint. Therefore, the interpolation in the graphic generate this effect. Considering the algorithm using DEM, the MSAC obtains a better estimation in relation to RANSAC and LMedS.

Secondly, we evaluate the algorithms on both sequences, but computing the EM between the first frame and the next frames. It enable us to understand the limitation of both CEM and DEM with the increasing of the baseline. Fig. A.4 shows the linear and angular movements. One can see that the movements are almost linear on the living room sequence (last row of Fig. A.4. Fig. A.5 shows the error estimated in this experiment. Differently from Fig. A.3, we show the error in relation to the linear distance, *i.e.* the distance between optical center. It makes easier to understand the behavior of the estimated EM.

The sequence **office room** presents a strange behavior around $0.1m$ of linear distance, where the error decreases in terms of rotation and translation even though the increasing of the baseline. This fact is explained using Fig A.6a. The number of valid points using the chirality constraint and the ground truth transformation becomes



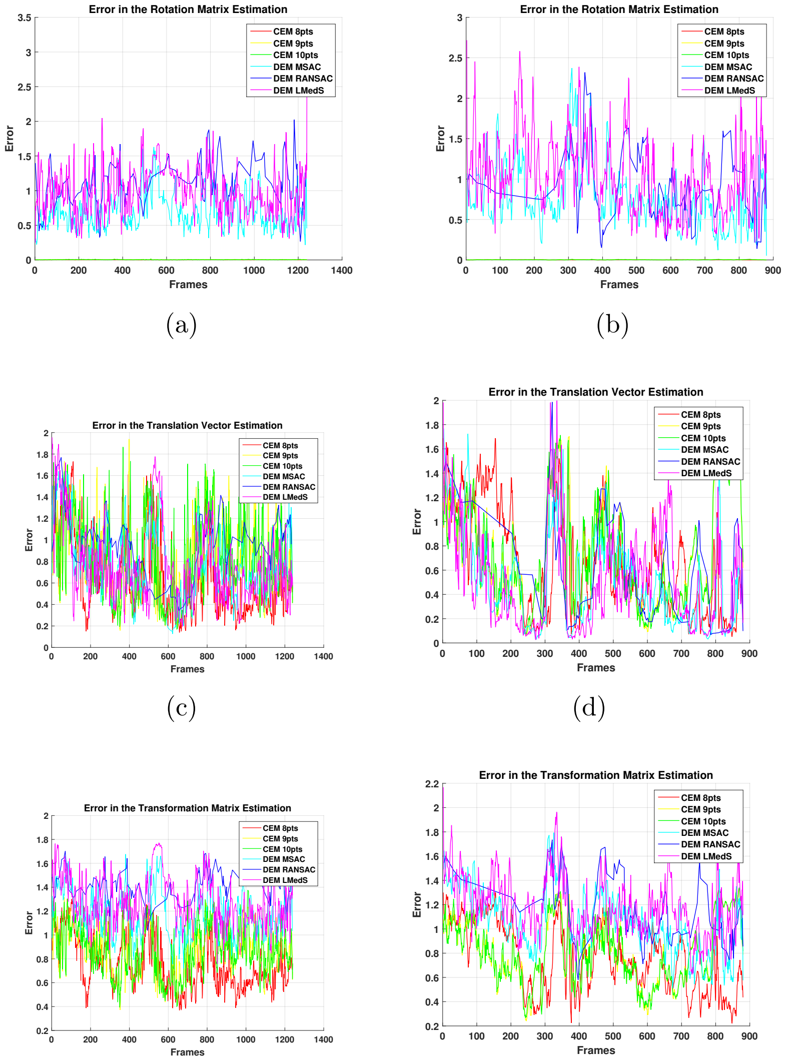
(a) (b)

(c) (d)

Figure A.2. Linear and Angular movements between consecutive frames on both sequences: a-b) office room, c-d) living room. All distances are in meters (linear) and radius (angular).

zero around $0.1m$ of linear distance (black line in the graphics). The optical flow misestimates the movement of the pixels for this sequence, thus both CEM and DEM are not able to recover a reliable rigid body transformation. The sequence **living room** presents a different behavior. The number of valid points by chirality remains the maximum for the reference, *i.e.* using the ground truth transformation. The rotation matrix error is small for the CEM-based algorithm but increases with the growing in the linear distance, mainly after 0.75 meters. Moreover, the error for the translation vector presents the inverse behavior of the rotation, *i.e.* the DEM-based algorithms obtain a better estimation for linear distance smaller than 0.75 meters. The transformation error is similar for DEM and CEM based algorithms. Table A.2 show the similarities between both approaches in terms of transformation error. The characteristics of the estimation of rotation and translation are also highlighted.

The comparative study shows that the CEM-based algorithm using eight points is more robust than the overdetermined solution using nine and ten points. The CEM-based algorithm are better to estimate the rotation, while the algorithm presents difficult to estimate the translation. Both techniques (DEM and CEM) cannot be a final solution to estimate the rigid body transformation between camera's pose. The quan-



(a) (b)

(c) (d)

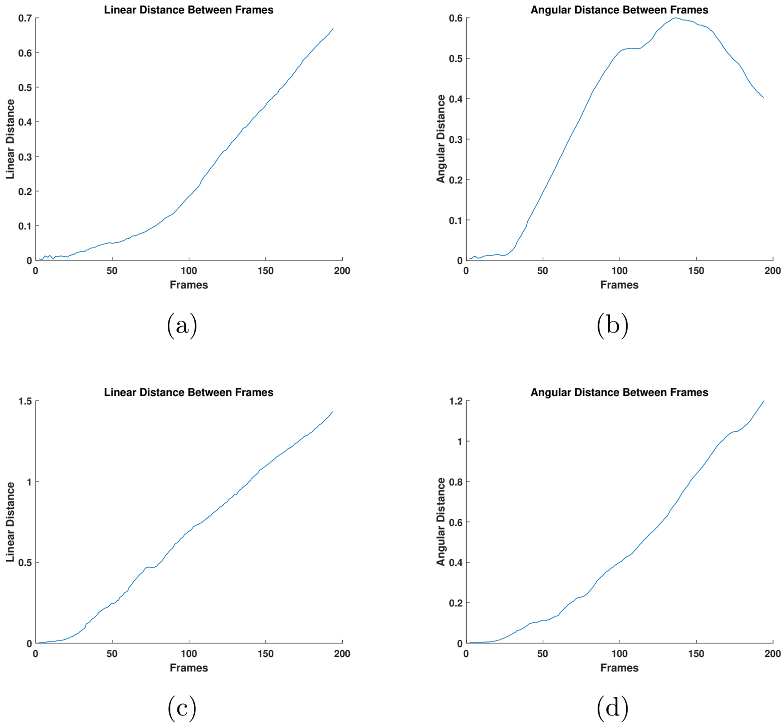
(e) (f)

Figure A.3. Error in the estimation of the rigid transformation using consecutive frames for both sequences: office room in the first column, and living room in the last column. The first row (a-b) presents the error of rotation matrix, the second row are presented the error of the translation vector (c-d), and the last row shows the error of the transformation matrix.

titative values show some advantage in terms of transformation by using CEM, thus we adopted the MSAC-based CEM using eight points in the present work as a initial guess for a bundle adjustment technique [Triggs et al., 2000; Hartley and Zisserman, 2003].

Table A.1. Average and standard deviation of the error in the rigid body transformation using consecutive frames for both sequences. The best averages are highlighted in bold letters.

		Office Room			**Living Room**		
Method		Rot.	Transl.	Transf.	Rot.	Transl.	Transf.
CEM 8pts	Avg.	**0.002330**	**0.657108**	**0.758753**	**0.002069**	0.667752	**0.753689**
	Std	0.001475	0.505243	0.289481	0.001428	0.498385	0.319202
CEM 9pts	Avg.	0.004258	0.852919	0.872409	0.002677	0.744838	0.806606
	Std	0.003099	0.582834	0.310093	0.001821	0.526670	0.311466
CEM 10pts	Avg.	0.004311	0.861890	0.877007	0.002663	0.742510	0.806467
	Std	0.00312	0.58803	0.31167	0.001868	0.524807	0.308045
DEM MSAC	Avg.	0.676872	0.816649	1.18150	0.736822	**0.475976**	1.038521
	Std	0.520095	0.491881	0.312596	0.641611	0.502258	0.366747
DEM RANSAC	Avg.	1.10128	1.010607	1.422841	0.903909	0.593603	1.145581
	Std	0.80129	0.42933	0.29665	0.854836	0.59409	0.432072
DEM LMedS	Avg.	0.98231	0.779188	1.281781	1.051306	0.524294	1.189755
	Std	0.659627	0.531614	0.34448	0.846314	0.582624	0.400370



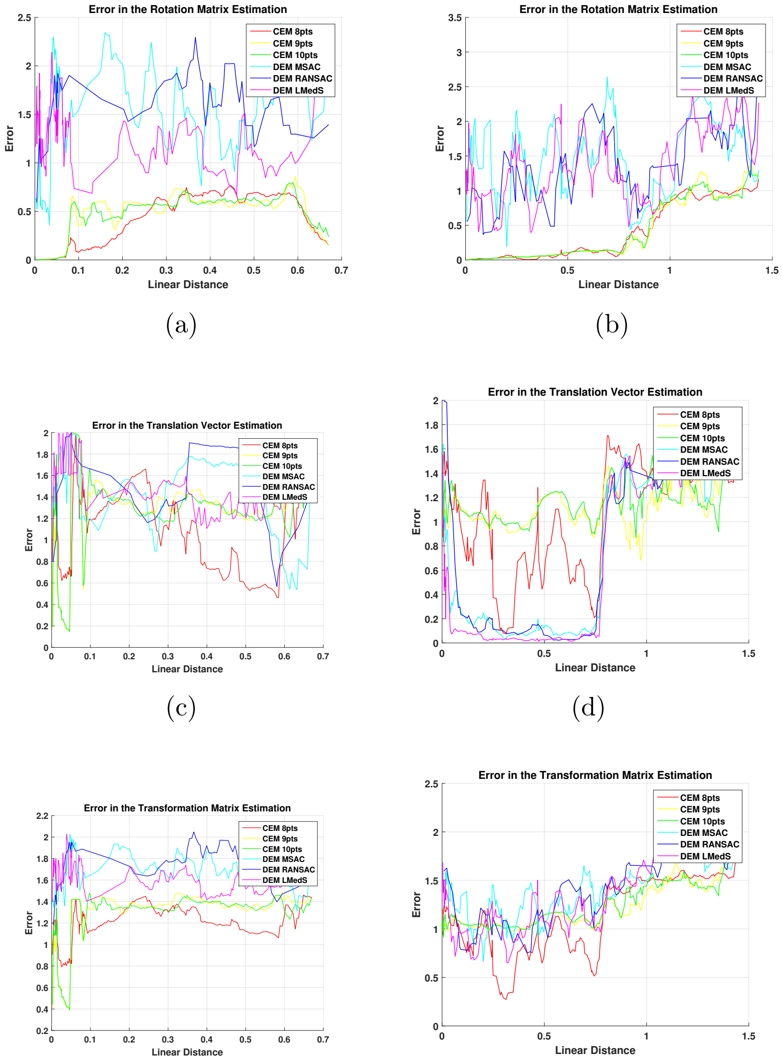
(a) (b)

(c) (d)

Figure A.4. Linear and Angular movements between the first frame and the next frames on both sequences: a-b) office room, c-d) living room. All distances are in meters (linear) and radius (angular).

A.4 Triangulation

Once the projection matrices P and P' are recovered, the 3D point $\boldsymbol{X}$ may be recovered by triangulation [Hartley and Sturm, 1997], where the true reprojection error is mini-



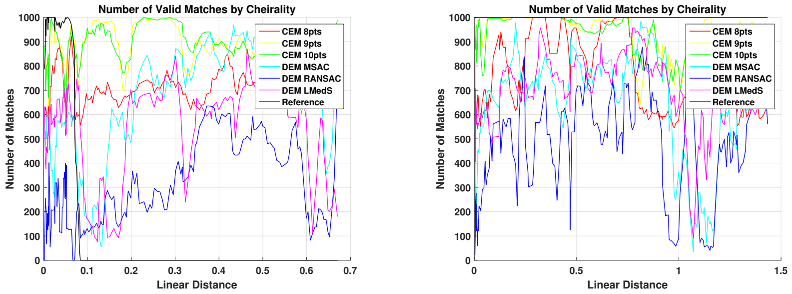
(a) (b)

(c) (d)

(e) (f)

Figure A.5. Error in the estimation of the rigid transformation using the first frame and the next frames for both sequences: office room in the first column, and living room in the last column. The first row (a-b) presents the error of rotation matrix, the second row are presented the error of the translation vector (c-d), and the last row shows the error of the transformation matrix. The linear distances is in meters.

mized by solving a sixth-order polynomial. Beardsley et al. [1994] suggest to compute the midpoint of the common perpendicular between the two rays. An overview of triangulation algorithms is given in Hartley and Zisserman [2003]. However, we adopted the approach proposed by Torr [2002] because of its simplicity. They estimate $\boldsymbol{X}$ by applying SVD decomposition from $\boldsymbol{x} = (x, y)^T$ and $\boldsymbol{x'} = (x', y')^T$. Let $\boldsymbol{p}^i$ be the i-th row of the projection matrix P and $\boldsymbol{p'}^i$ be the i-th row of the projection matrix P',



(a) (b)

Figure A.6. Number of valid points using the chirality constraint for each sequence: a) office room, and b) living room. The estimation is based on one thousand points estimated using optical flow. The linear distances is in meters.

Table A.2. Average and standard deviation of the error in the rigid body transformation using the first frame and the next frames for both sequences. The best averages are highlighted in bold letters.

		Office Room			**Living Room**		
Method		Rot.	Transl.	Transf.	Rot.	Transl.	Transf.
CEM 8pts	Avg.	**0.306121**	**1.105515**	**1.169817**	**0.378748**	1.098573	**1.146534**
	Std	0.290783	0.488267	0.208302	0.413188	0.527135	0.404511
CEM 9pts	Avg.	0.365388	1.257535	1.230013	0.387784	1.128535	1.208585
	Std	0.293676	0.528539	0.332513	0.456287	0.288679	0.236500
CEM 10pts	Avg.	0.361910	1.242662	1.225148	0.389719	1.160783	1.223141
	Std	0.288143	0.518709	0.322681	0.444695	0.291051	0.233903
DEM MSAC	Avg.	1.499262	1.513044	1.712151	1.442140	**0.663936**	1.387041
	Std	0.831767	0.454976	0.285233	0.945333	0.646120	0.428174
DEM RANSAC	Avg.	1.616332	1.603255	1.776378	1.361133	0.770464	1.385884
	Std	0.611496	0.484758	0.254921	0.941609	0.71062	0.461444
DEM LMedS	Avg.	1.246401	1.514946	1.621339	1.378038	0.669803	1.347090
	Std	0.765533	0.513975	0.365398	0.899111	0.723191	0.484253

some constraints can be obtained:

$$\begin{aligned} x\boldsymbol{p}^{3T}\boldsymbol{X} - \boldsymbol{p}^{1}\boldsymbol{X} &= 0, \\ y\boldsymbol{p}^{3T}\boldsymbol{X} - \boldsymbol{p}^{2}\boldsymbol{X} &= 0, \\ x\boldsymbol{p}^{2T}\boldsymbol{X} - y\boldsymbol{p}^{1}\boldsymbol{X} &= 0. \end{aligned} \tag{A.12}$$

Thus, we can derive an equation in the form $A\boldsymbol{X} = 0$ that may be solved using least square optimization [Torr, 2002]. The matrix A can be written as:

$$A = \begin{bmatrix} x\boldsymbol{p}^{3T} - \boldsymbol{p}^{1} \\ y\boldsymbol{p}^{3T} - \boldsymbol{p}^{2} \\ x'\boldsymbol{p}'^{3T} - \boldsymbol{p}'^{1} \\ y'\boldsymbol{p}'^{3T} - \boldsymbol{p}'^{2} \end{bmatrix}. \tag{A.13}$$

Finally, the chirality constraint is applied in order to remove outliers [Werner and Pajdla, 2001], *i.e.* the points in the 3D space need to lie in front of the cameras.

A.5 Bundle Adjustment

Bundle Adjusment (BA) is a technique that refines a visual reconstruction to produce jointly optimal three-dimensional structure and camera parameters. The bundle of light rays leaves each 3D point and converges on each camera center, crossing the image plane. The idea of the BA is minimize the distance between the crossing point and the measured 2D point, *i.e.* the reprojection error. One of the main drawbacks of BA is the possibility of converging to a local, rather than a global minimum of the cost function. Thus, a good initialization is a important issue. Details about BA can be found on Triggs et al. [2000] and Hartley and Zisserman [2003].

In the present work, we have adopted the Sparse Bundle Adjustment [Lourakis and Argyros, 2009]. It is based on the fact that the interaction among parameters for different three-dimensional points and cameras matrix results in the underlying normal equations exhibiting a sparse block structure. Thus, a specialized Levenberg-Marquardt [Marquardt, 1963] algorithm variant is used. It is tailored to fit the sparseness of the problem, obtaining a flexible and performance efficient approach. Details about the implementation is found on Lourakis and Argyros [2009].