



Amanda Cardoso Duarte

Geração de Conjuntos de Dados para Aplicações de Visão Computacional em Ambientes Subaquáticos

Rio Grande, 2017

Amanda Cardoso Duarte

Geração de Conjuntos de Dados para Aplicações de Visão Computacional em Ambientes Subaquáticos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande - FURG

Universidade Federal do Rio Grande - FURG

Centro de Ciências Computacionais - C3

Programa de Pós-Graduação em Engenharia de Computação

Orientadora: Prof^a. Dr^a. Silvia Silva da Costa Botelho

Coorientador: Prof. Dr. Paulo Lilles Jorge Drews Junior

Colaborador: MsC. Felipe Codevilla Moraes

Rio Grande,

2017

Ficha catalográfica

D812g Duarte, Amanda Cardoso.
Geração de conjunto de dados para aplicação de visão computacional em ambientes subaquáticos / Amanda Cardoso Duarte. – 2017.
81 p.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-graduação em Computação, Rio Grande/RS, 2017.

Orientadora: Dr^a. Silvia Silva da Costa Botelho.
Coorientador: Dr. Paulo Lilles Jorge Drews Junior.

1. Dataset 2. Restauração de imagem 3. Ambientes Subaquáticos
I. Botelho, Silvia Silva da Costa II. Drews Junior Paulo Lilles Jorge
II. Título.

CDU 004.92

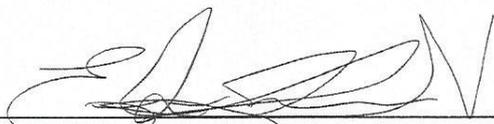
UNIVERSIDADE FEDERAL DO RIO GRANDE
Centro de Ciências Computacionais
Programa da Pós-Graduação em Computação
Curso de Mestrado em Engenharia de Computação

DISSERTAÇÃO DE MESTRADO

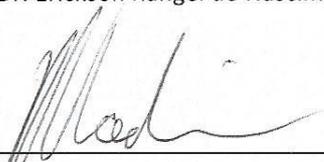
**GERAÇÃO DE CONJUNTOS DE DADOS PARA APLICAÇÕES DE VISÃO COMPUTACIONAL
EM AMBIENTES SUBAQUÁTICOS**

Amanda Cardoso Duarte

Banca examinadora:



Prof. Dr. Erickson Rangel do Nascimento



Prof. Dr. Ricardo Nagel Rodrigues



Profa. Dra. Silvia Silva da Costa Botelho
Orientadora



Prof. Dr. Paulo Lilles Jorge Drews Jr
Coorientador

Agradecimentos

Primeiramente, agradeço aos meus pais, por sempre terem colocado a minha educação e as dos meus irmãos em primeiro lugar e sempre terem se esforçado para que eu pudesse realizar os meus objetivos não medindo esforços para que isso acontecesse. Aos meus irmãos, cunhado e cunhada e as minhas sobrinhas pelo apoio, carinho, suporte e incentivo.

Aos meus amigos e amigas de longa data, que mesmo em dias de stress não deixaram de me chamar para um mate e me dar a força que eu precisava.

Aos colegas e amigos do Centro de Ciências Computacionais, por terem ajudado a colocar em prática as ideias e auxiliado de alguma forma na realização do presente trabalho, dedico o meu apreço. Em especial, agradeço ao companheirismo e colaboração dos colegas da sala B06 que estavam todos os dias presentes nos momentos bons e ruins, nas tardes de bolo ou nas manhãs de café, na hora de carregar aquário ou me ajudar em algo, a eles meu muito obrigada.

Aos amigos que fiz em Rio Grande e na FURG, um lugar que me acolheu de braços abertos e que com certeza levarei no meu coração daqui em diante.

À minha orientadora, Prof. Dr^a. Sílvia Silva da Costa Botelho e ao meu Co-orientador Prof. Dr. Paulo L. Drews Jr., pela confiança em mim depositada, pelos ensinamentos, pelo apoio e ajuda no desenvolvimento deste trabalho, pela amizade criada nesses últimos dois anos e acima de tudo pela oportunidade de crescimento não só profissional quanto pessoal que eu pude obter com a ajuda de vocês.

Agradeço à Petrobras, à Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), à Autoridade de Financiamento de Estudos e Projetos (FINEP) e ao Ministério da Ciência e Tecnologia pelo auxílio financeiro que possibilitou a realização deste trabalho através do Programa de Recursos Humanos da ANP para o Setor de Petróleo e Gás - PRH-ANP / MCT. Agradeço principalmente ao PRH27 e a sua coordenadora Maria Isabel por todo apoio e excelente trabalho.

Agradeço também aos professores Dr. Ricardo Nagel Rodrigues (FURG) e Dr. Erickson Rangel do Nascimento (UFMG), pela disponibilidade e participação na banca de avaliação

*“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda
pensou sobre aquilo que todo mundo vê.”
(Arthur Schopenhauer)*

Resumo

O presente estudo, desenvolvido em torno de aplicações para métodos de visão computacional e aprendizagem de máquina aplicados à imagens capturadas em ambientes subaquáticos, compreende na criação de conjuntos de dados, ou também chamados de *datasets*, para aplicações voltadas para restauração de imagens podendo ser estendida à outras áreas da visão computacional. De maneira mais detalhada o trabalho tem como objetivo a criação de dois tipos de conjuntos de dados, contendo: (i) imagens geradas a partir de um ambiente controlado onde as condições do meio e a degradação presente em cada imagem podem ser acessadas e monitoradas, (ii) imagens geradas através da síntese da degradação causada por ambientes participativos a fim de criar conjuntos de dados simulados de larga escala. Neste contexto, primeiramente foi proposto uma metodologia para a reprodução de ambientes subaquáticos em um ambiente controlado. Estes ambientes contam com a presença de características encontradas em cenas naturais onde as condições do meio podem ser controladas. A partir dessa metodologia é viável a criação de diferentes conjuntos de dados contendo imagens degradadas onde o acesso a imagem de referência, ou seja, a mesma cena sem degradação sob as mesmas condições, assim como a produção de tipos e intensidades de degradação distintos é possível. Visto como uma das contribuições do presente trabalho, o *dataset* TURBID 3D contendo todas as imagens geradas neste ambiente, bem como algumas das suas aplicações relacionadas aos métodos de restauração de imagens podem ser acessadas a partir do endereço: www.amandaduarte.com.br/turbid. Ao que se refere a segunda contribuição, foi criado um simulador, chamado de **TURBI**, capaz de simular em imagens não degradadas a degradação causada por meios participativos a partir de fragmentos de uma imagem capturada em um ambiente natural. A partir do desenvolvimento do simulador TURBI, é possível a criação de *datasets* simulados de larga escala contendo pares de imagens (imagens degradadas e suas respectivas imagens de referência). A criação de *datasets* de larga escala contribui para o avanço no desenvolvimento de novas abordagens utilizando métodos baseados em aprendizagem, como *Deep Learning*, que necessitam uma grande quantidade de imagens para serem treinados. O código-fonte do simulador bem como algumas das imagens utilizadas para a geração dos *datasets* de larga escala podem ser acessados pelo endereço: https://github.com/mandacduarte/Underwater_Simulator. A criação de ambos conjuntos de dados assim como algumas das aplicações onde as imagens geradas foram utilizadas são apresentados no decorrer deste trabalho.

Palavras-chaves: *Dataset*; Restauração de imagem; Ambientes Subaquáticos.

Abstract

The present study, developed around contributions to computer vision and machine learning methods applied to images captured in underwater environments, comprises the creation of datasets for applications focused on image restoration and more areas of computer vision. Specifically, this work is focused on the creation of two types of datasets containing: (i) images generated from a controlled environment where the media conditions and the degradation present in each image can be accessed and monitored, ii) images generated through the synthesis of degradation caused by participating media in order to create large-scale datasets. In this context, it was first created an controlled underwater environment with characteristics found in natural environments. Through this environment it is possible to create different datasets containing degraded images where to the reference image, (the same scene without degradation under the same conditions) as well as different types and levels of degradation is available . Seen as one of the contributions of the present work, the dataset TURBID 3D containing all the images generated in this environment, as well as some of its applications related to image restoration methods can be accessed by the address: www.amandaduarte.com.br/turbid. Referring to the second contribution, a simulator called TURBI was created. With this simulator it is possible to synthesize in non-degraded images the degradation caused by participating media from patches of an image captured in a natural environment. With the TURBI simulator, it is possible to create large-scale datasets containing pairs of images (degraded images and their respective reference images). Large scale datasets contribute to the development of new approaches using learning-based methods such as deep learning, which require a large amount of images to be trained. The simulator code as well as some of the images used for large-scale datasets generation can be accessed at: https://github.com/mandacduarte/Underwater_Simulator. The creation of both datasets as well as some of applications made using the generated images are presented in the course of this work.

Key-words: Dataset; Image restoration; Underwater Environments;

Lista de ilustrações

Figura 1 – Exemplos de imagens capturadas em meios participativos	19
Figura 2 – Exemplos de imagens turvas	20
Figura 3 – Modelo de formação da imagem em meios participativos	21
Figura 4 – Estrutura principal do experimento proposto	40
Figura 5 – Cena 1	41
Figura 6 – Cena 2	41
Figura 7 – Cena 3	42
Figura 8 – Metodologia	43
Figura 9 – Estrutura principal do experimento proposto com água	43
Figura 10 – Conjunto de imagens <i>Milk</i>	49
Figura 11 – Conjunto de imagens <i>Chlorophyll</i>	49
Figura 12 – Conjunto de imagens <i>Chlorophyll</i> com vista lateral	50
Figura 13 – Conjunto de imagens <i>DeepBlue</i>	50
Figura 14 – Resultado visual da restauração da imagens do conjunto de imagens <i>Milk</i>	59
Figura 15 – Erro quadrático médio entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	60
Figura 16 – Relação Sinal-Ruído de Pico entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	60
Figura 17 – Índice de similaridade estrutural entre as imagens restauradas utili- zando os métodos de restauração e a imagem de referência	60
Figura 18 – Índice de similaridade de características entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	61
Figura 19 – Índice de similaridade SR-SIM entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	61
Figura 20 – Índice de similaridade estrutural em várias escalas entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	61
Figura 21 – Erro quadrático médio entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	62
Figura 22 – Relação Sinal-Ruído de Pico entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	62
Figura 23 – Índice de similaridade estrutural entre as imagens restauradas utili- zando os métodos de restauração e a imagem de referência	63
Figura 24 – Índice de similaridade de características entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	63

Figura 25	– Índice de similaridade SR-SIM entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	63
Figura 26	– Índice de similaridade estrutural em várias escalas entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	64
Figura 27	– Erro quadrático médio entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	64
Figura 28	– Relação Sinal-Ruído de Pico entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	64
Figura 29	– Índice de similaridade estrutural entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	65
Figura 30	– Índice de similaridade de características entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	65
Figura 31	– Índice de similaridade SR-SIM entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	65
Figura 32	– Índice de similaridade estrutural em várias escalas entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência	66
Figura 33	– Restauração de imagens com cenários urbanos com névoa	68
Figura 34	– Restauração de imagens com névoa	69

Lista de tabelas

Tabela 1 – Quantidade de leite adicionado no experimento	45
Tabela 2 – Quantidade de Chá Verde adicionada no experimento	46
Tabela 3 – Quantidade de Clorofila adicionada no experimento	47
Tabela 4 – Conjuntos de imagens disponíveis e suas respectivas informações	49
Tabela 5 – Propriedades dos <i>datasets</i> relacionados presentes no estado da arte em comparação com o TURBID 3D.	50

Lista de Abreviaturas e Siglas

NOAA *National Oceanic and Atmospheric Administration*

ROV *Remotely Operated Vehicle* - Veículo Operado Remotamente

AUV *Autonomous Underwater Vehicle* - Veículo Subaquático Autônomo

VOC *Visual Object Classes Recognition*

ILSVRC *ImageNet Large Scale Visual Recognition Challenge*

LED *Light Emitting Diode* - Diodo Emissor de Luz

CNN *Convolutional Neural Network* - Redes Neurais Convolucionais

MSE *Mean Squared Error* - Erro Quadrático Médio

PSNR *Peak-Signal-to-Noise Ratio* - Relação Sinal-Ruído de Pico

SSIM *Structural Similarity Index* - Índice de Similaridade Estrutural

MSSSIM *Multi-Scale Structural Similarity* - Índice de Similaridade Estrutural de Várias Escalas

FSIM *Feature Similarity Index* - Índice de Similaridade de Características

SRSIM *Spectral Residual Based Similarity* - Similaridade Baseada no Espectro Residual

IQA *Image Quality Assessment* - Avaliação da Qualidade de Imagem

FR-IQA *Full-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Completa

RR-IQA *Reduced-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Reduzida

NR-IQA *No-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem sem Imagem de Referência

Sumário

1	Introdução	12
1.1	Motivação e Justificativa	13
1.2	Definição do Problema	15
1.3	Objetivos	15
1.4	Contribuições	16
1.5	Estrutura da Dissertação	16
2	Fundamentação Teórica	18
2.1	Propriedades das Imagens Capturadas em Meios Participativos	18
2.2	Métodos de Restauração de Imagens	24
2.2.1	Avaliação da qualidade da imagem: Métricas	27
2.2.1.1	Mean Squared Error (MSE) e Peak-Signal-to-Noise Ratio (PSNR)	28
2.2.1.2	Structural Similarity Index (SSIM)	29
2.2.1.3	Multi-Scale Structural Similarity (MS-SSIM)	29
2.2.1.4	Feature Similarity Index (FSIM)	30
2.2.1.5	Spectral Residual Based Similarity (SR-SIM)	30
3	Datasets na Visão Computacional	31
3.1	Trabalhos Relacionados	36
4	Criação dos Datasets Propostos	38
4.1	Aquisição de Imagens Turvas em um Espaço Controlado	38
4.1.1	Criação do <i>dataset</i> TURBID 3D: Descrição do Experimento	38
4.1.1.1	Estrutura Desenvolvida	39
4.1.1.2	Cenas Fotografadas	40
4.1.1.3	Metodologia adotada	42
4.2	Simulação de Meios Participativos em Imagens Ópticas	51
4.2.1	Desenvolvimento do Simulador TURBI	51
4.2.2	Usando mapa de profundidade para sintetizar imagens turvas	54
5	Aplicações dos Datasets Propostos	55
5.1	Avaliação de Métodos de Restauração de Imagens Aplicados à Ambientes Subaquáticos	55
5.1.1	Métodos Avaliados	56
5.1.2	Avaliação Quantitativa	57

5.1.3	Avaliação da Qualidade das Imagens Restauradas	57
5.1.3.1	Resultados	58
5.2	Treinamento de Redes Neurais Convolucionais para Restauração de Imagens capturadas em Ambientes Participativos	67
6	Considerações Finais	70
	Referências	72

1 Introdução

Nas últimas décadas, os conjuntos de dados padronizados ou também chamados de *datasets*, vem sendo uma alavanca para o crescimento de várias áreas principalmente na visão computacional. Grandes áreas como a classificação e reconhecimento de objetos, fluxo óptico, dentre outras, se beneficiam do progresso e desafios trazidos pelo constante desenvolvimento de novos *datasets*.

Na área de visão computacional, esses tipos de dados são importantes na avaliação e acompanhamento do progresso científico, bem como para despertar o interesse em novas áreas de pesquisa. Com eles é possível avaliar o funcionamento do método durante seu desenvolvimento, assim como a sua comparação objetiva com as abordagens conhecidas como estado da arte na área. *Datasets*, dependendo da área de estudo, também podem ser elaborados trazendo novos desafios e incentivando o avanço no desenvolvimento de novas abordagens.

No entanto, a criação de um *dataset* que possa ser válido para o crescimento de uma área e que tenha uma vida útil razoável é um trabalho árduo. Alguns pontos chaves independente da área em que os dados serão utilizados devem ser levados em consideração, tais como: a maneira e o local onde os dados serão coletados, em alguns casos os sensores utilizados, a qualidade e a quantidade dos dados capturados, a seleção e avaliação desses dados e o quanto esses dados abordam o problema a ser tratado. Também, na maioria das vezes estes processos demandam muito tempo e trabalho humano, o que hoje em dia gera um custo elevado.

Por conta disso, diversas áreas sofrem com problemática de aquisição de dados e por consequência trazem desafios para a criação de *datasets*. Como é o caso da robótica, da mineração de dados, do processamento de linguagem natural, da visão computacional, dentre outras. Dentro da visão computacional uma área afetada pela falta de *datasets* acessíveis é a restauração de imagens (KATSAGGELOS, 2012). Embora haja *datasets* disponíveis para a restauração de imagens que tratam de problemas genéricos como ruídos e borrões causados por movimento ou imagens sem foco, para a comunidade de pesquisa voltada para a restauração de imagens subaquáticas, por exemplo, esse número é bastante reduzido.

Isso ocorre principalmente pelo alto preço de equipamentos necessários, pela demanda de pessoal especializado e pelas dificuldades práticas de captura de imagens de alta qualidade em ambientes subaquáticos naturais. Além disso, quando se trata de *datasets* com imagens subaquáticas, há um desafio ainda maior: obter imagens com diferentes níveis de degradação decorrentes da turbidez, incluindo o *ground-truth* das imagens captu-

radas. Nesses ambientes a captura da mesma cena com e sem degradação é praticamente impossível em ambientes naturais.

A falta desses dados prejudica principalmente a avaliação e comparação dos algoritmos desenvolvidos para lidar com esse tipo de problema. Com isso, esse processo torna-se subjetivo, não tendo padrões definidos e dificultando assim o progresso no desenvolvimento de métodos mais robustos.

Adicionado a isso, muitos métodos na sua concepção requerem a utilização de pares de imagens, (a imagem degradada e o seu respectivo *ground-truth*), como por exemplo técnicas de aprendizado de máquina em geral ou mais recentemente nas abordagens chamadas de *Deep Neural Networks* ou também conhecidas como *Deep Learning* (LECUN; BENGIO; HINTON, 2015) (XIE; XU; CHEN, 2012) (XU et al., 2014). Essas redes são capazes de aprender automaticamente complexas relações de entrada-saída com base na observação de dados. Abordagens envolvendo *Deep Learning* vem ganhando uma grande atenção nos últimos anos devido ao seus resultados relevantes na resolução de problemas como classificação de imagens e detecção de objetos, alcançando hoje resultados que são considerados estado da arte.

Entretanto, esses métodos são orientados a dados, ou seja, precisam de uma grande quantidade de informação para serem treinados, o que dificulta ainda mais a criação de *datasets* para esse tipo de aplicação.

1.1 Motivação e Justificativa

Pesquisadores têm explorado a Terra desde os tempos antigos investigando desertos, montanhas, florestas e praticamente tudo que está a frente dos seus olhos. Entretanto, um grande sistema ecológico ainda não foi bem explorado: os oceanos. Mesmo os ambientes hidrográficos sendo considerados fundamentais para o nosso planeta e indispensáveis para o sustento, transporte, comércio e crescimento da humanidade, ainda assim segundo o *National Oceanic and Atmospheric Administrations* (NOAAs) (NOAA, 2014) menos de 5% dos oceanos da Terra foram explorados por olhos humanos ou por robôs controlados por humanos.

Recentemente, novas tecnologias robóticas estão permitindo avanços significativos associados à exploração e realizações de missões subaquáticas. Desde o advento dos *Remotely Operated Vehicle* - Veículo Operado Remotamente (ROVs) (CENTENO, 2007) e dos *Autonomous Underwater Vehicle* - Veículo Subaquático Autônomo (AUVs) (WYNN et al., 2014), a exploração dos ambientes subaquáticos ganhou um estímulo maior devido aos recursos que esses modelos de robôs oferecem. Uma das vantagens é a possibilidade de captura de imagens, principalmente em ambientes perigosos ou de difícil acesso como em águas profundas, na inspeção e manipulação de estruturas subaquáticas e tubulações de

óleo, dentre outros. Uma das grandes áreas beneficiadas com esse progresso é a indústria petrolífera. Com muitas das reservas situadas em zonas oceânicas ultra profundas, e sendo o petróleo ainda o principal componente da matriz energética mundial, atividades de exploração e extração de petróleo vem demandando o estudo, desenvolvimento e aplicação de tecnologias de imageamento para percepção de ambientes subaquáticos.

Além da aplicação no importante mercado de petróleo e gás, a obtenção e análise de imagens subaquáticas também são importantes em diversas outras aplicações, tais como: Reconstrução de cenas em três dimensões (BRANDOU et al., 2007; CONCHA et al., 2015), classificação de corais por imagens (SHIHAVUDDIN et al., 2013; STOKES; DEANE, 2009; CODEVILLA et al., 2015a), navegação de robôs subaquáticos (AULINAS et al., 2011; EUSTICE; PIZARRO; SINGH, 2008; SILVEIRA et al., 2015), mapeamento (BOTELHO et al., 2009) dentre outros.

Entretanto, devido às propriedades fotométricas presentes nas imagens capturadas nesses ambientes, na maioria das vezes as imagens originais adquiridas não trazem consigo as informações necessárias para certas aplicações. A degradação presente em imagens subaquáticas são causadas pela interação dos raios de luz com as partículas presentes no meio, nas quais podem causar efeitos característicos que degradam a imagem (DUNTLEY, 1963).

Para tratar a degradação causada pela interação dos raios com as partículas suspensas no ambiente, a utilização de algoritmos de processamento de imagens que melhoram a qualidade geral da imagem são amplamente utilizados. Esses métodos são chamados de métodos de restauração de imagens. Entretanto, muitas vezes a utilização de um só método não é o suficiente para recuperar todas as informações contidas na imagem. Um dos grandes problemas enfrentados para a criação de métodos robustos de restauração de imagens subaquáticas, se dá pela dificuldade de estimativa dos parâmetros do ambiente. Diferentes meios possuem diferentes parâmetros a serem estimados, além de cada tipo de ambiente possuir propriedades específicas, como a diferença de cor, diferentes partículas presentes no meio, dentre outros.

Além disso, para algumas abordagens a falta de *datasets* que contenham informações relevantes desses ambientes, além das imagens de referência, é outro grande problema enfrentado. Nesses ambientes a captura da mesma cena com e sem degradação é praticamente impossível em ambientes naturais.

Apesar da existência de uma comunidade de pesquisa voltada para estudos em ambientes subaquáticos, os recursos financeiros e humanos destinados a criação de *datasets* úteis nessa área ainda é muito limitado. Diferentemente de outras áreas onde os esforços voltados para a criação e disponibilização de dados recebem uma grande atenção, a de pesquisa subaquática ainda sofre com a falta de dados públicos e disponíveis. Essa falta de dados ocorre principalmente pelas dificuldades de captura de imagens de alta

resolução e pelo alto custo dos equipamentos necessários para esse tipo de captura.

1.2 Definição do Problema

Esta dissertação trata o problema da criação de conjuntos de dados para métodos de aprendizado de máquina baseados em aprendizagem supervisionada e métodos de visão computacional aplicados a ambientes subaquáticos. Mais especificamente a criação de dados úteis para métodos direcionados a restauração de imagens que apresentam degradação causada pela interação da luz com o ambiente.

1.3 Objetivos

O presente trabalho tem como objetivo tratar o problema da criação de conjuntos de dados para a utilização em métodos de visão computacional e aprendizado de máquina quando aplicados à imagens capturadas em ambientes subaquáticos.

De maneira mais específica o trabalho tem como objetivo a criação de dois tipos de conjuntos de dados contendo: (i) imagens geradas a partir de um ambiente controlado onde as condições do meio e a degradação presente em cada imagem pudessem ser acessada e monitoradas, (ii) imagens geradas por meio da síntese da degradação causada por ambientes participativos a fim de criar conjuntos de dados simulados de larga escala. Para esse fim o trabalho conta com os seguintes objetivos específicos:

- Propor uma metodologia para a reprodução de ambientes subaquático que contam com a presença de características encontradas em cenas naturais onde as condições do meio podem ser controladas;
- Gerar diferentes conjuntos de dados contendo imagens degradadas com sua respectiva imagem de referência utilizando a metodologia proposta e que possam ser utilizadas por métodos de visão computacional aplicados a ambientes subaquáticos;
- Propor um sistema computacional para estimar parâmetros de um modelo de propagação da luz na água e gerar imagens degradadas a partir de imagens limpas com base nos parâmetros estimados tanto de maneira uniforme ou a partir do seu respectivo mapa de profundidade;
- Utilizar o sistema proposto para a geração de *datasets* simulados de larga escala que contenha imagens degradadas com seus respectivos *ground-truth* que possam ser utilizados no treinamento de métodos de restauração de imagens baseados em aprendizagem;

- Disponibilizar os conjuntos de dados gerados durante o presente trabalho a fim de contribuir com a comunidade científica.

1.4 Contribuições

O trabalho desenvolvido nesta dissertação busca apresentar à comunidade científica uma análise dos principais métodos propostos para melhorar a qualidade de imagens subaquáticas e contribuir com:

1. Uma nova metodologia para a criação de conjuntos de dados gerados em um ambiente subaquático controlado contendo as informações do meio e diferentes níveis de degradação que permite avaliar quantitativamente os métodos de restauração de imagens existentes.

2. Um sistema computacional capaz de simular em imagens não degradadas a degradação causada por meios participativos tanto de maneira uniforme ou a partir do seu respectivo mapa de profundidade com o auxílio de um modelo físico.

3. Utilizar o sistema proposto para a geração de *datasets* simulados de larga escala que contenha imagens degradadas com seus respectivos *ground-truth* que possam ser utilizados no treinamento de métodos de restauração de imagens baseados em aprendizagem supervisionada.

Durante o desenvolvimento deste trabalho destaca-se a publicação de dois artigos e a apresentação em forma de *poster* em um *workshop* em conferência internacional:

- (DUARTE et al., 2016a) **A dataset to evaluate underwater image restoration methods**. OCEANS 2016-Shanghai, 2016.

- (GAYA et al., 2016a) **Vision-based Obstacle Avoidance Using DeepLearning**. 13th Latin-America Robotics Symposium, 2016.

- (DUARTE et al., 2016b) **TURBID: An Underwater Turbid Image Dataset**. Presented in European Conference on Computer Vision (ECCV) Workshop on Datasets and Analysis Performance in Early Vision, 2016.

1.5 Estrutura da Dissertação

Além do presente capítulo, este trabalho está dividido da seguinte maneira:

- O **Capítulo 2** introduz a fundamentação teórica acerca dos principais tópicos estudados para o desenvolvimento desta dissertação;
- No **Capítulo 3** é apresentado a descrição de alguns dos *datasets* mais conhecidos na área de visão computacional assim como os trabalhos relacionados;

-
- O [Capítulo 4](#) apresenta a descrição dos *datasets* produzidos durante o desenvolvimento deste trabalho.
 - O [Capítulo 5](#) apresenta as duas principais aplicações dos *datasets* apresentados neste trabalho.
 - O [Capítulo 6](#) mostra as considerações finais e contribuições alcançadas juntamente com as sugestões de trabalhos futuros propostos para a continuação deste trabalho.

2 Fundamentação Teórica

O presente capítulo apresenta a fundamentação teórica dos principais tópicos pesquisados durante o desenvolvimento deste trabalho.

Inicialmente, é abordada uma visão geral acerca das propriedades das imagens capturadas em meios participativos. Tendo em vista o objetivo geral deste trabalho como a criação de *datasets* que contenham imagens subaquáticas degradadas e seus respectivos *ground-truth*, um dos principais desafios é a reprodução dessa degradação em imagens não degradadas. Para isso, é importante a compreensão dos fatores associados ao fenômeno de degradação nas imagens. Com isso, sendo o meio subaquático um meio participativo, um estudo sobre o processo de formação de imagens nesses meios bem como as características que eles possuem foi realizado e é apresentado neste capítulo.

Visto que a restauração de imagens capturadas em meios participativos é uma das principais aplicações para os dados criados durante o presente trabalho, ainda neste capítulo serão abordados alguns métodos de restauração de imagens presentes no estado da arte tanto baseados no modelo físico de formação de imagem, quanto baseados em aprendizagem supervisionada.

2.1 Propriedades das Imagens Capturadas em Meios Participativos

Em óptica, meios participativos são aqueles onde o meio participa das interações de iluminação à medida que a luz viaja entre as superfícies. Ou seja, nestes meios, em algum grau, as imagens capturadas são degradadas por fenômenos físicos que envolvem a propagação da luz. Esses meios normalmente possuem um grande número de partículas suspensas as quais interagem com os fótons que estão sendo propagados, reduzindo a quantidade de luz refletida por um objeto que chega até a câmera durante o processo de formação da imagem. Ambientes subaquáticos, assim como ambientes com névoa, fumaça e tempestade de areia são considerados meios participativos. A [Figura 1](#) apresenta alguns exemplos de imagens capturadas em ambientes participativos.

No presente trabalho dá-se uma atenção especial aos ambientes subaquáticos, porém os problemas causados assim como o modelo físico de formação de imagem podem ser estendidos para outros meios participativos. Esta seção busca apresentar as propriedades da propagação da luz em meios participativos, bem como isso interfere no processo de formação das imagens capturadas nesses ambientes.

A atenuação da luz em meios participativos é a perda gradual da intensidade dos raios durante a propagação da luz. Essa perda está relacionada com a concentração e com

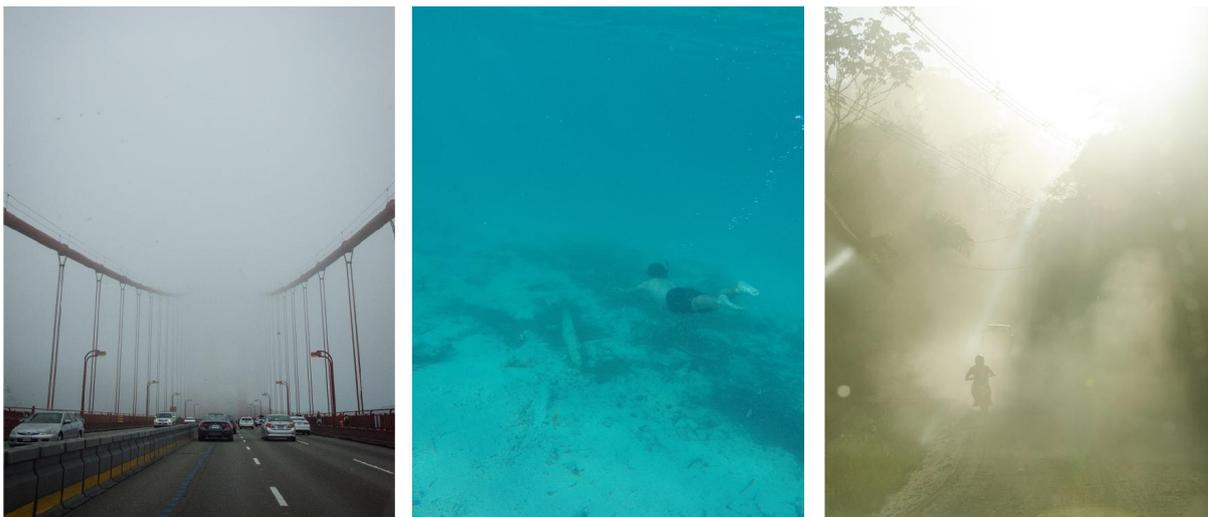


Figura 1 – Exemplos de imagens capturadas em ambientes participativos. A primeira imagem exemplifica parte da cena oclusa por névoa. Na segunda imagem pode ser observado o efeito da turbidez em imagens capturadas em ambientes subaquáticos. Na terceira imagem pode ser observado a degradação causada por ambientes com tempestade de areia. Fotos de autoria própria.

os diferentes tipos de refração encontrados nas partículas presentes no ambiente. Esse fenômeno pode acontecer por meio de dois processos, absorção e espalhamento (KIRK, 1994).

A absorção pode ser entendida como a conversão da energia da luz em outras formas, normalmente calor enquanto o espalhamento muda a direção da propagação dos raios luminosos. Sendo assim, imagens óticas capturadas em meios participativos normalmente sofrem com a falta de qualidade. Essa perda de informação resulta em uma versão degradada da imagem. Baixo contraste, ruído, distorção das cores e efeitos de borramento são problemas comuns encontrados por exemplo em imagens subaquáticas (SAHU; GUPTA; SHARMA, 2014).

Em imagens subaquáticas, um termo comumente utilizado relacionado a esses fenômenos é *turbidez*. Turbidez é uma propriedade comum no meio aquático que está relacionada com a quantidade de luz que é absorvida ou espalhada ao invés de ser transmitida em uma linha reta (OMAR; MATJAFRI, 2009). Esta absorção e espalhamento causam a perda de transparência e claridade da água, acarretando na presença de efeitos característicos nestas imagens. São consideradas imagens turvas imagens na qual a visibilidade da cena capturada é degradada pela turbidez do ambiente. A Figura 2 apresenta alguns exemplos de imagens turvas capturadas em ambientes subaquáticos reais.

Considerando um raio de luz como um número inicial de fótons n_0 que se propaga em um meio homogêneo, onde o coeficiente de espalhamento somado ao coeficiente de atenuação é representado por c , a perda de informação luminosa pode ser descrita pela



Figura 2 – Exemplos de imagens turvas capturadas em ambientes subaquáticos naturais. Nessas imagens podem ser observadas a degradação normalmente encontrada em imagens capturadas nesse tipo de ambiente. Fonte: www.flickr.com

lei de Atenuação de *Lambert*, dada pela [Equação 2.1](#):

$$n = n_0 e^{-cd}, \quad (2.1)$$

sendo n o número de fótons que alcança o sensor e d a distância percorrida pelo feixe de luz.

Desta forma, o modelo de formação de imagens busca descrever os caminhos por onde esses fótons se propagam, desde a fonte até a sua captura onde ocorre a formação da imagem. Este processo pode ser modelado como uma complexa interação entre a luz propagada no ambiente, o meio e a estrutura a ser imageada ([DREWS-JR, 2016](#)).

Um dos modelos mais populares na literatura foi proposto por [McGlamery \(1980\)](#) e [Jaffe \(1990\)](#). Segundo o modelo, a irradiância total (E_T), ou seja, a densidade de potência que incide sobre um pixel pode ser obtida pelo somatório de três componentes: A componente direta (E_d), o *Forward Scattering* (E_f) e o *Backscattering* (E_{bs}), como mostra a [Equação 2.2](#)

$$E_T = E_d + E_f + E_{bs}. \quad (2.2)$$

Este processo de propagação pode ser observado na [Figura 3](#). Sendo a luz direta sem espalhamento vinda do objeto capturado modelada pela **Componente Direta**. Quando os raios de luz vindos da cena imageada são espalhados em pequenos ângulos criando um efeito de borrramento na imagem, esse fenômeno é chamado de espalhamento dianteiro ou *forward scattering*, representado pela componente **Forward Scattering**.

Por outro lado, a componente **Backscattering** está associada a luz que incide no plano da imagem vindo de um ponto que não faz parte da cena observada. Isso acontece devido ao volume de partículas presentes entre a cena e a câmera. Essas partículas podem desviar a trajetória da luz para o plano da imagem, comportando-se muitas vezes como um ruído aditivo. Essa componente é alterada normalmente pela profundidade, no caso subaquático, e influenciada pelo meio em que a cena se encontra. Pode-se entender que quanto maior a distância entre o objeto e a câmera maior será a influência dessa componente na imagem.

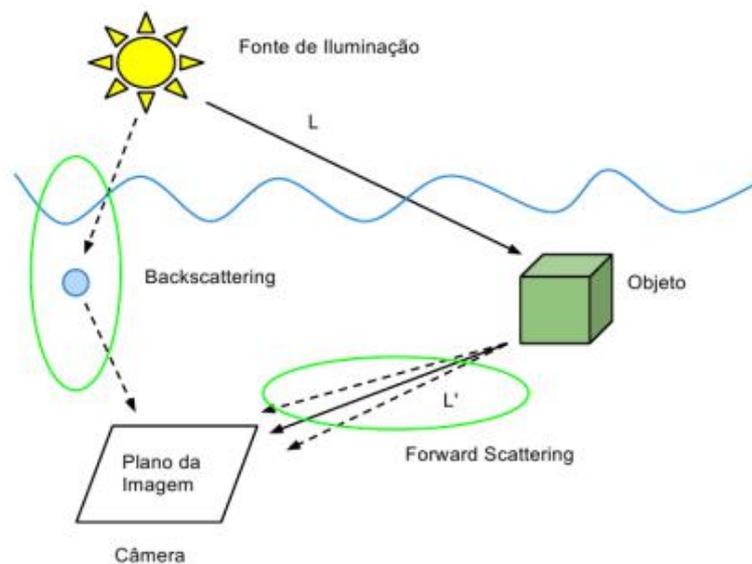


Figura 3 – Os raios luminosos que incidem no plano da imagem da câmera podem ter três trajetórias/componentes diferentes. A **componente direta**, contendo a fração da luz que chega até a câmera vinda diretamente da cena, o **forward-scattering** contendo a luz refletida da cena que sofre espalhamento no caminho até a câmera, e a componente **backscattering** que contém informações vindas de fora da cena resultantes da interação da fonte de iluminação e as partículas suspensas no ambiente.

Para calcular cada componente algumas simplificações são consideradas afim de tornar o modelo mais simples e tratável computacionalmente. Com isso, assume-se que a iluminação que incide sobre cena assim como a refletividade do objeto fotografado como sendo uniformes e descarta-se os parâmetros da câmera. Na componente direta, a luz é produzida pela fonte de iluminação e segue em todas as direções em torno dela. Como citado anteriormente, parte da luz que irradia do objeto imageado é perdida por meio de dois processos, absorção e espalhamento. A componente direta captura esses efeitos

modelando a fração de luz que alcança a câmera como:

$$E_d = J e^{-cd}, \quad (2.3)$$

onde J é a irradiância da cena, que é atenuada por e^{-cd} , onde d é a distância entre a cena e a câmera e c é o coeficiente de atenuação. O termo e^{-cd} é conhecido também como a transmissão. A transmissão pode ser entendida como a porção de degradação em cada parte da imagem em função da distância do objeto até a câmera.

Segundo [Jaffe \(1990\)](#), a absorção é uma propriedade associada aos ambientes subaquáticos, o que faz com que o coeficiente de atenuação c seja a soma dos coeficientes de absorção a e espalhamento total β , sendo ambos dependentes do comprimento de onda λ , sendo assim $a(\lambda)$ e $\beta(\lambda)$. Além disso, os valores de distância d podem ser estimados utilizando-se da transmissão, porém a constante c (coeficiente de atenuação) utilizada para calculá-la é normalmente desconhecida.

A componente *forward scattering* compreende a parcela dos raios de luz capturada pela câmera na qual sofreram um espalhamento em pequenos ângulos, causando um efeito de borramento na imagem. Para descrever essa componente um modelo normalmente empregado pode ser visto através da [Equação 2.4](#) ([MCGLAMERY, 1980](#); [NEGAHDARIPOUR](#); [ZHANG](#); [HAN, 2002](#); [TRUCCO](#); [OLMOS-ANTILLON, 2006](#))

$$\begin{aligned} E_{fs} &= I_d * (x, y | d, G, c, B) \\ &= I_d [e^{-Gd} - e^{-cd}] \mathcal{F}^{-1} \{ e^{-Bdw} \} \\ &= I_d \mathcal{F}^{-1} \{ K e^{-Bdw} \}, \end{aligned} \quad (2.4)$$

onde B e G são parâmetros do meio. Assumindo o fato de que G e c são similares, um modelo simplificado com uma constante K é empregado ([TRUCCO](#); [OLMOS-ANTILLON, 2006](#)). \mathcal{F}^{-1} indica o inverso da transformada de *Fourier* e w indica a frequência radial. O fator de amortecimento B é um parâmetro empírico que está atrelado as propriedades de espalhamento inerentes ao meio. Normalmente este parâmetro é similar ao coeficiente de espalhamento total β ([NEGAHDARIPOUR](#); [ZHANG](#); [HAN, 2002](#)).

Por último, o efeito *backscattering* é resultante da interação entre a fonte de iluminação do ambiente, tanto natural ou artificial, e as partículas suspensas no meio. Um modelo simplificado desta interação é definido pela [Equação 2.5](#)

$$I_{bs} = A(1 - e^{-cd}), \quad (2.5)$$

onde A é a luz global da cena ou também chamado de luz do ambiente no infinito, ou seja, no horizonte não ocluso. Esse termo é uma função dependente do comprimento de onda e da profundidade do ambiente, no caso de meios subaquáticos, além de ser afetado pela iluminação e as condições do meio onde ela se encontra. Com essa equação pode-se

concluir que quanto mais distante estiver o objeto da câmera, maior será a componente *backscattering*, com isso, menos da cena real irá existir na imagem.

A constante A pode ser estimada de diferentes maneiras: com o processo de calibração (QUEIROZ-NETO et al., 2004; GIBSON; VO; NGUYEN, 2012), encontrando o pixel mais distante na cena (NASCIMENTO; CAMPOS; BARROS, 2009), através da intensidade máxima em imagens na qual métodos de balanceamento de branco foram aplicados (ZHANG et al., 2010; TAREL; HAUTIERE, 2009), encontrando o pixel mais intenso na imagem (NARASIMHAN; NAYAR, 2002; TAN, 2008; FATTAL, 2008), fazendo a média de 1% dos *pixels* mais intensos na imagem (CARR; HARTLEY, 2009), ou estimando o pixel com maior *Dark Channel* (HE; SUN; TANG, 2011).

Segundo Schechner e Karpel (2005), em meios participativos o *backscattering* é o fator responsável pela maior parte da perda de contraste e visibilidade na formação da imagem. Tendo assim a componente *forward scattering* uma pequena contribuição na degradação total podendo ser normalmente desprezado dependendo da aplicação. Entretanto, essa componente foi levada em consideração em alguns trabalhos, tais como em (NEGAHDARIPOUR; ZHANG; HAN, 2002; TRUCCO; OLMOS-ANTILLON, 2006; BARROS, 2010; HE et al., 2013).

Com essas definições pode ser estabelecido um modelo simplificado que descreve a formação de imagens em meios participativos como:

$$I(x, y) = J(x, y)e^{-cd(x,y)} + A(1 - e^{-cd(x,y)}), \quad (2.6)$$

onde x, y são as coordenadas do pixel e $I(x, y)$ é a imagem obtida no meio participativo. Essa versão simplificada é similar ao modelo proposto por Koschmieder (1924) bastante utilizada em imagens capturadas em ambientes com névoa.

Além disso, esse modelo pode ser estendido de forma a incorporar diferentes componentes de onda. Para tal, adotando-se o padrão RGB de representação, pode-se descrever o fenômeno através dos seus três canais conforme a Equação 4.1

$$I^\lambda(x, y) = J^\lambda(x, y)e^{-c^\lambda d(x,y)} + A^\lambda(1 - e^{-c^\lambda d(x,y)}), \lambda \in \{R, G, B\}. \quad (2.7)$$

Com essa definição, métodos de restauração de imagens são desenvolvidos baseando-se no inverso desse modelo. A irradiância pode ser recuperada usando a Equação 2.8 após a estimativa dos parâmetros $e^{-cd(x,y)}$ e A :

$$J(x, y) = \frac{I(x, y) - A}{e^{-cd(x,y)}} + A. \quad (2.8)$$

Esse modelo permite a estimativa da imagem restaurada baseando-se na imagem degradada $I(x)$, no mapa de profundidade $d(x, y)$ e nos parâmetros do meio A e c . Sendo

assim, o problema de restauração de imagens pode ser entendido como um problema de estimativa dos parâmetros do meio em que a imagem foi capturada e do mapa de profundidade referente a ela.

2.2 Métodos de Restauração de Imagens

Como mencionado anteriormente, certas imagens podem apresentar uma baixa qualidade visual devido à degradação causada por diferentes fenômenos presentes no seu processo de formação. Para lidar com esse tipo de problema, diferentes técnicas de visão computacional podem ser empregadas. Métodos de restauração de imagens buscam resgatar a qualidade da imagem e tratar problemas específicos gerados pelo ambiente em que a imagem foi capturada, como por exemplo imagens capturadas em meios participativos. Nessa seção serão abordados os principais métodos computacionais utilizados para a melhoria das imagens capturadas em meios participativos, dando atenção especial as imagens subaquáticas.

Dependendo do tipo de imagem e da degradação que ela apresenta, alguns métodos gerais também chamados de *enhancement* são utilizados para recuperar a visibilidade das imagens, tais como: Equalização de histograma adaptativo (HUMMEL, 1977), Balanceamento de Branco (FINLAYSON; TREZZI, 2004), Filtro Bilateral (TOMASI; MANDUCHI, 1998), *Color Constancy* (WEIJER; GEVERS; GIJSENIJ, 2007), etc. Além dessas técnicas, *Unsharp masking* e filtros passa-alta são algumas técnicas comumente utilizadas para remover ruídos ou aumentar o contraste em imagens.

Em Eustice et al. (2002) foi apresentado, além de técnicas para a geração de mosaicos de imagens subaquáticas, uma modificação da Equalização de Histograma Adaptativa com Limitação de Contraste (ZUIDERVELD, 1994) (*Contrast Limited Adaptive Histogram Equalization* ou também conhecido como CLAHE) chamado de *Contrast Limited Adaptive Histogram Specification* (CLAHS) como uma extensão do *toolbox* de processamento de imagens do MATLAB. Essa adaptação visa reparar o baixo contraste e a iluminação não uniforme encontrados em imagens capturadas em ambientes subaquáticos.

A utilização de transformação de histogramas para imagens subaquáticas é apresentado também por (CHAMBAH D. SEMANI, 2004). O método proposto é baseado no modelo de Equalização Automática de Cor (*Automatic Color Equalisation*) o qual é inspirado nos mecanismos de adaptação do sistema visual humano que considera a distribuição espacial da informação de cor da imagem. Este método propõe recuperar a qualidade da cor das imagens capturadas em ambientes subaquáticos com a utilização deste modelo.

Especificamente associada a correção do problema da iluminação não uniforme em ambientes subaquáticos, uma análise foi feita por Garcia, Nicosevici e Cuff (2002) utilizando quatro técnicas diferentes. A primeira delas leva em consideração um modelo

na qual a imagem é considerada como um produto da iluminação e propriedades de refletância da cena. A segunda técnica é uma adaptação da Equalização de histograma para o caso de iluminação não uniforme na qual é utilizada uma Equalização de histograma local para tentar solucionar este problema. A terceira técnica consiste na utilização de uma filtragem das altas frequências através de um filtro homomórfico e a quarta técnica realiza uma subtração do campo de iluminação extraído de curvas polinomiais.

Considerando a estratégia de combinação de técnicas para tratar os diferentes problemas encontrados nas imagens, Bazeille S. e Malkasse (2006) propôs uma aplicação de filtro de pré processamento para imagens subaquáticas. Este método automático propõe melhorar a qualidade de imagens segmentadas e reduzir os problemas causados pelos ambientes subaquáticos. Esse método é composto por sucessivos passos de processamento independentes que buscam corrigir a iluminação não uniforme, reduzir o ruído, melhorar o contraste e ajustar a cor das imagens capturadas em ambientes subaquáticos.

Alguns problemas podem ocorrer quando vários filtros são aplicados sucessivamente em uma imagem. Primeiro, as contribuições de cada filtro não são utilizadas perfeitamente. Além disso, quando aumenta-se o contraste de uma imagem isso pode acarretar no aumento também de ruído. Com o intuito de evitar esse tipo de problemas, Ancuti et al. (2012) propôs um método de fusão considerando imagens com diferentes filtros. Baseado nisso, o aprimoramento da imagem é baseado da medida dos pesos considerando somente a versão degradada da imagem.

No entanto, métodos de *enhancement* de imagem não investem em recuperar as propriedades da imagem original. A grande maioria dos métodos de restauração de imagens concebidos para recuperar a imagem deteriorada removendo a degradação, e não só melhorando a visibilidade geral da imagem. Para isso, uma relação entre o modelo físico de formação de imagens, apresentado na seção 2.1 é levado em consideração.

Um dos principais métodos de restauração de imagens baseados nesta abordagem foi proposto por He, Sun e Tang (2011). Esse método foi chamado de *Dark Channel Prior*, onde os autores consideram que em uma cena natural a maioria dos *patches* locais contém alguns *pixels* na qual ao menos um canal de cor a sua intensidade é aproximadamente zero. Com isso, o método é capaz de estimar o quanto de degradação contém aquele fragmento e recuperar a imagem sem degradação. Esse método foi adaptado várias vezes para imagens subaquáticas (BIANCO; MOHAN; EUSTICE, 2010) (CHIANG; CHEN, 2012) (LU et al., 2015) (DREWS et al., 2016).

Também, algumas abordagens consideram a grande variedade de cores que existem em ambientes subaquáticos assumindo alguma condição específica como a absorção do canal vermelho, como em (GALDRAN et al., 2015) por exemplo. Este método propõe a recuperação das cores dos baixos comprimentos de onda lidando com a recuperação do contraste perdido.

Outro método proposto por [Fattal \(2008\)](#), assume que não existe covariância entre a refletância da cena e a iluminação. Com isso, a transmissão pode ser definida como a fonte de covariância. Essa suposição apresentou bom desempenho principalmente em condições onde a imagem está pouco degradada ([FATTAL, 2014](#)).

Nesse mesmo trabalho, [Fattal \(2014\)](#) propôs também um método para estimar a transmissão utilizando o pressuposto da linha de cor ([OMER; WERMAN, 2004](#)). Essa suposição permite inferir a transmissão utilizando o ponto de intersecção entre a linha de cor e o vetor com a orientação da luz do ambiente. Apesar dos bons resultados obtidos, o método é dependente da obtenção de partes da imagem onde algumas propriedades do modelo existem.

Além de métodos específicos, uma abordagem geral para restauração de imagens capturadas em meios participativos foi proposta por [Gaya et al. \(2016b\)](#). Nesse trabalho, os autores propuseram a combinação de duas diferentes *priors*, o contraste local e a cor, como uma eficiente abordagem para restauração de imagem.

Embora esses métodos apresentem relevantes contribuições e excelentes resultados para algumas imagens. Entretanto muitos deles sendo limitados por recursos projetados a mão. Ou seja, geralmente nem todas as condições da cena são conhecidas para acessar os parâmetros do modelo de formação de imagem. Tradicionalmente, esses métodos foram desenvolvidos com base no modelo físico descrito na [Equação 2.6](#).

Esses métodos tentam estimar o mapa de transmissão usando heurísticas e *priors* desenvolvidos com base na observação de imagens sem degradação. Embora esses métodos possam obter resultados muito bons em alguns casos, eles falham em situações em que as *priors* utilizadas não são verdadeiras.

Aliado a isso, e ao alto desempenho em soluções de vários problemas relacionados a Visão Computacional, os métodos baseados em aprendizado de máquina, tais como as Redes Neurais Convolucionais (CNNs), tem se demonstrado uma abordagem promissora também na área de restauração de imagem. Embora um dos grandes problemas para essa abordagem seja a quantidade de dados para o seu treinamento, algumas redes foram desenvolvidas para o tratamento de remoção de névoa em imagens.

Por outro lado, um dos principais problemas enfrentados pelos métodos de restauração de imagem para realizar a remoção da degradação, é estimar o mapa de transmissão do meio para uma dada imagem degradada. Com isso, ambos os trabalhos ([CAI et al., 2016](#)), ([REN et al., 2016](#)) e ([GAYA et al., 2016a](#)) utilizam-se de CNNs para estimar o mapa de transmissão da imagem, e assim recupera-lá utilizando a modelo tradicional.

Essas redes são capazes de aprender automaticamente complexas relações de entrada-saída com base na observação de dados. Esses métodos são capazes de aprender heurísticas mais complexas, que os humanos não conseguiam perceber, podendo levar a melhores re-

sultados de restauração em uma gama mais ampla de situações.

Embora em outras abordagens esse tipo de rede é capaz de solucionar todo o ciclo de vários problemas relacionados a Visão Computacional, até o presente momento nenhuma abordagem levando em consideração a utilização de CNNs foi utilizada para resolver o problema de restauração de imagens capturadas em meios participativos. Acredita-se que isso se deve ao fato da falta de uma grande quantidade dados relacionados à esses ambientes. Com isso, um dos objetivos desse trabalho é contribuir com essa área desenvolvendo e disponibilizando *datasets* de grande escala que contenham as informações necessárias para o treinamento desse tipo de rede.

Na [seção 5.1](#) alguns dos principais métodos apresentados nesta seção serão avaliados quanto a sua robustez quando aplicados a diferentes níveis de degradação. Essa avaliação será feita com a utilização do *dataset* produzido neste trabalho apresentado no [Capítulo 4](#) utilizando algumas das métricas apresentadas no [Capítulo 5](#).

2.2.1 Avaliação da qualidade da imagem: Métricas

Com o grande progresso no desenvolvimento de métodos de restauração de imagens, uma forma de acessar e comparar a performance desses métodos se tornou uma tarefa necessária. O objetivo principal das *Image Quality Assessment* - Avaliação da Qualidade de Imagens (IQAs) é desenvolver modelos matemáticos capazes de prever a qualidade de uma imagem com precisão e automaticamente. Nos últimos, diferentes métricas foram desenvolvidas com essa finalidade.

Tendo a disponibilidade de uma imagem de referência, ou seja, uma imagem livre de degradação e que possua uma qualidade perfeita, as métricas objetivas de acesso da qualidade de imagens podem ser divididas em três categorias.(1) A primeira categoria é chamada de *Full-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Completa (FR-IQA) onde a imagem de referência com qualidade perfeita está totalmente disponível. Assim, a imagem restaurada (ou distorcida) pode ser avaliada quanto à sua qualidade por meio de comparação direta com a imagem referência, aumentando assim a exatidão da comparação.(2) A segunda categoria é chamada de *Reduced-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Reduzida (RR-IQA), onde a imagem de referência não está totalmente disponível. Ao invés vez disso, algumas características da imagem de referência são extraídas e utilizadas como informação adicionais para avaliar a qualidade da imagem de teste.(3) A terceira categoria é *No-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem sem Imagem de Referência (NR-IQA) onde não temos acesso à imagem de referência.

Para a avaliação quantitativa dos métodos de restauração escolhidos foram utiliza-

das sete métricas FR-IQA que serão descritas nas subseções a seguir. Dentre as métricas escolhidas estão: *Mean Squared Error* - Erro Quadrático Médio (MSE) e *Peak-Signal-to-Noise Ratio* - Relação Sinal-Ruído de Pico (PSNR), o *Structural Similarity Index* - Índice de Similaridade Estrutural (SSIM) (WANG et al., 2004), o *Multi-Scale Structural Similarity* - Índice de Similaridade Estrutural de Várias Escalas (MSSSIM) (WANG; SIMONCELLI; BOVIK, 2003), o *Feature Similarity Index* - Índice de Similaridade de Características (FSIM) (ZHANG et al., 2011), e o *Spectral Residual Based Similarity* - Similaridade Baseada no Espectro Residual (SRSIM) (ZHANG; LI, 2012).

Em todas as subseções seguintes, I_{ref} e I_{tst} denotam as imagens de referência e de teste, respectivamente, e W e H representam a largura e a altura da imagem respectivamente.

2.2.1.1 Mean Squared Error (MSE) e Peak-Signal-to-Noise Ratio (PSNR)

As medidas mais simples de erro são, o Erro Quadrático Médio (MSE) e a Relação Sinal-Ruído de Pico (PSNR). No entanto estes são conhecidos por se correlacionar fracamente com a qualidade visual para a maioria das aplicações.

O MSE indica a potência da distorção, ou seja, a diferença entre as imagens de referência e de teste. O valor de MSE pode ser calculado usando a seguinte equação:

$$\text{MSE} = \frac{1}{WH} \sum_{j=1}^H \sum_{i=1}^W (I_{ref}(i, j) - I_{tst}(i, j))^2.$$

O MSE é frequentemente convertido na Relação Sinal-Ruído de Pico (PSNR). PSNR é a razão da potência máxima possível de um sinal e potência de distorção, e pode ser calculado por:

$$\text{PSNR} = 10 \log \frac{D^2}{\text{MSE}},$$

Onde D indica o intervalo dinâmico de intensidade de pixel. Por exemplo, para uma imagem de 8bits/pixel $D= 255$.

Mesmo tendo uma relação pequena com a qualidade visual da imagem, o MSE possui algumas características que o tornam uma medida de desempenho amplamente utilizada no campo do processamento de sinais (WANG; BOVIK, 2009). Dentre elas estão:

- É um método simples e computacionalmente barato;
- Tem um significado fisicamente claro, isto é, é uma maneira natural de definir o erro da energia de um sinal;

- Uma vez que MSE satisfaz propriedades como convexidade, simetria e diferenciabilidade, o MSE é considerada uma medida excelente na otimização de aplicações.
- É considerado como uma convenção, isto é, é amplamente utilizado para otimização e avaliação em uma ampla extensão de aplicações de processamento de sinais.

Apesar das características interessantes acima citadas do MSE, quando se trata da estimativa da qualidade da imagem pela percepção humana, o MSE mostra desempenho ruim. Isto é devido ao fato de que algumas das importantes características fisiológicas e psicofísicas do sistema visual humano (HVS) não são contabilizadas por esta medida.

2.2.1.2 Structural Similarity Index (SSIM)

O índice de similaridade estrutural (SSIM) (WANG et al., 2004) supõe que HVS é altamente adaptado para extrair informações estruturais de uma cena. Portanto, este algoritmo tenta modelar a informação estrutural de uma imagem. O algoritmo SSIM baseia-se no fato de que os *pixels* de uma imagem natural demonstram fortes dependências e essas dependências carregam informações úteis sobre a estrutura de uma cena. Portanto, um método que é capaz de medir a mudança de informação estrutural pode fornecer uma boa aproximação da distorção da imagem percebida. O algoritmo SSIM define a degradação da imagem como mudança percebida na informação estrutural.

Em (WANG et al., 2004), os autores afirmam que a estrutura dos objetos em uma cena é independente da luminância local e contraste. Portanto, para extrair a informação estrutural, devemos separar o efeito da iluminação. Neste algoritmo, a informação estrutural de uma imagem é definida como aqueles traços que representam a estrutura de objetos nessa imagem, independentemente da luminância local e do contraste.

O algoritmo SSIM realiza medição de similaridade em três etapas: comparação de luminância, comparação de contraste e comparação de estrutura.

2.2.1.3 Multi-Scale Structural Similarity (MS-SSIM)

O algoritmo SSIM descrito anteriormente é considerado uma abordagem de escala única que atinge seu melhor desempenho quando aplicado em uma escala apropriada. Além disso, a escolha da escala certa depende das condições de visualização, por exemplo, da distância de visualização e da resolução do visor. Sabendo que este algoritmo não tem a capacidade de se adaptar a essas condições, foi proposto o índice de similaridade estrutural em várias escalas (MS-SSIM) (WANG; SIMONCELLI; BOVIK, 2003).

A vantagem dos métodos multi-escala, como o MS-SSIM, em relação aos métodos de uma única escala, como o SSIM, é que nos métodos em escala múltipla os detalhes da

imagem em diferentes resoluções e condições de visualização são incorporados ao algoritmo de avaliação da qualidade.

Depois de ter as imagens de referência e de teste como entrada, este algoritmo aplica um filtro passa-baixa e reamostra (por um fator de 2) de forma iterativa a imagem filtrada (*downsampling*). Em cada escala, a função de comparação de contraste e estrutura são calculados. No entanto, a função de comparação de luminância é calculada apenas em uma das escalas.

2.2.1.4 Feature Similarity Index (FSIM)

O índice de similaridade de características (FSIM) (ZHANG et al., 2011) é baseado no fato que o sistema visual humano compreende uma imagem principalmente devido às suas características de baixo nível como por exemplo detecção de bordas (MARR, 2010)(MORRONE; BURR, 1988). Para avaliar a qualidade de uma imagem, o algoritmo FSIM utiliza dois tipos de características. Experimentos fisiológicos e psicofísicos têm demonstrado que em pontos com alta congruência de fase (PC) o HVS pode extrair características altamente informativas (MORRONE; BURR, 1988)(HENRIKSSON; HYVÄRI-NEN; VANNI, 2009).

No entanto, o PC é invariante ao contraste e nossa percepção da qualidade de uma imagem também é afetada pelo contraste local dessa imagem. Como resultado desta dependência, a magnitude do gradiente da imagem (GM) é usada como uma característica secundária no algoritmo FSIM. O cálculo da medida FSIM consiste em duas etapas: computar a congruência de fase e a magnitude do gradiente da imagem, e a medida de similaridade entre as imagens de referência e de teste.

2.2.1.5 Spectral Residual Based Similarity (SR-SIM)

O índice de similaridade SR-SIM (ZHANG; LI, 2012) é baseado em um modelo específico de saliência visual de uma imagem chamado "*spectral residual visual saliency (SRVS)*". Este índice foi desenvolvido com base na hipótese de que o mapa de saliência visual de uma imagem está intimamente relacionado à sua qualidade percebida. Uma das vantagens trazidas por essa métrica é a baixa complexidade computacional e consequentemente a sua performance em tempo real.

3 *Datasets* na Visão Computacional

Muitas áreas relacionadas à visão computacional investem na criação e disponibilização de *datasets*, não só dados para o uso interno de um grupo de pesquisa, mas também na criação de *datasets* de referência ou também conhecido como *benchmarks* para o auxílio e avaliação dos métodos criados em diferentes áreas. Embora exista um grande número de *datasets* hoje em dia, poucos deles estão disponíveis publicamente para download. Alguns desses podem ser encontrados em *websites* online para download como por exemplo nos *websites* de comunidades de pesquisa voltadas para área de visão computacional [Computer Vision Datasets \(2017\)](#) [CV Datasets on the web \(2017\)](#) [Computer Vision online \(2017\)](#) [List of RGBD datasets \(2017\)](#). Esse tipo de *datasets* são indispensáveis para comprovar e acompanhar o progresso científico, bem como para despertar o interesse em novas áreas de pesquisa. Além disso, esses conjuntos de dados podem muitas vezes determinar quais aspectos em um campo de pesquisa devem ser analisados a seguir.

Existem alguns *datasets* pequenos porém bem rotulados que abrangem objetos genéricos que servem como dados de referência de treinamento e avaliação para muitos dos algoritmos de visão computacional existentes até hoje. Alguns exemplos de *datasets* desse tipo são: Caltech 101/256 ([FEI-FEI; FERGUS; PERONA, 2006](#)) ([GRIFFIN; HOLUB; PERONA, 2007](#)) e MSRC ([SHOTTON et al., 2006](#)). Ou ainda, *datasets* com propósito específico como o FERET ([PHILLIPS et al., 1998](#)) e o *Labeled faces in the Wild* ([HUANG et al., 2008](#)) para identificação de faces ou ainda o *dataset* chamado de *From aardvark to zorro* ([FINK; ULLMAN, 2008](#)) utilizado para classificação de imagens de mamíferos.

Embora a quantidade de imagens, em particular as com dados anotados, tenha registado um enorme crescimento ao longo dos últimos anos, o levantamento, a utilização e reutilização de dados ainda é limitada e frequentemente focada em conjuntos de *dataset* de referência bem conhecidos que muitas vezes não suprem as necessidades de todas as áreas de pesquisa dentro da visão computacional.

Com os avanços nas pesquisas voltadas para a visão computacional, *datasets* maiores e com mais desafios são cada vez mais necessários para a nova geração de algoritmos. Com isso, alguns *datasets* de larga escala podem ser citados. No *dataset* chamado TinyImage apresentado por [Torralba, Fergus e Freeman \(2008\)](#) são apresentadas 80 milhões de imagens com baixa resolução de 32x32. Essas imagens foram coletadas da internet por meio de consulta em vários motores de busca a partir das palavras presentes no WordNet ([FELLBAUM, 1998](#)). Cada conceito significativo no WordNet possivelmente descrito por várias palavras ou frases de palavras, é chamado de "conjunto de sinônimo" ou "synset". Cada *synset* no *dataset* TinyImage contém uma média de 1000 imagens, dentre elas de

10% a 25% são possivelmente imagens limpas. Embora esse conjunto de imagens tenha tido sucesso em certas aplicações, o alto nível de ruído e a baixa resolução das imagens fez com que ele se tornasse menos adequado para o desenvolvimento, treinamento e avaliação de algoritmos com propósito geral.

O *dataset* proposto por Ahn e Dabbish (2004) chamado EPS foi adquirido utilizando um jogo online. Dois jogadores propõem de forma independente anotações para uma imagem com o objetivo de combinar o maior número possível de palavras em um determinado limite de tempo. Milhões de imagens são rotuladas através deste jogo, mas sua natureza acelerada também representa uma grande desvantagem. Outra grande desvantagem desse *dataset* é o fato de que a maioria das imagens contidas nele não são publicamente disponíveis. Somente 60 mil imagens e suas respectivas anotações podem ser acessadas¹.

Os *datasets* LabelMe(RUSSELL et al., 2008) e Lotus Hill(YAO; YANG; ZHU, 2007) provém 30 e 50 mil imagens anotadas e segmentadas respectivamente. Esses dois *datasets* possuem em torno de 200 categorias com a vantagem de que além das anotações convencionais os contornos e locais dos objetos são fornecidos. As imagens contidas nesses dois *datasets* carregadas ou fornecidas pelos usuários ou pesquisadores responsáveis pelo projeto. O *dataset* Lotus Hill não é publicamente disponível.

Além dos dados disponíveis por esses *datasets*, alguns projetos lançam para a comunidade científica *datasets* de referências a cada ano contendo desafios capazes de instigar o desenvolvimento de métodos mais robustos e acompanhar o progresso da área como por exemplo os *datasets* ligados aos desafios anuais de reconhecimento visual de classes PASCAL: *Visual Object Classes Recognition* (VOC) (EVERINGHAM et al., 2010) e o de reconhecimento visual de larga escala ImageNet: *Large Scale Visual Recognition Challenge* (ILSVRC) (RUSSAKOVSKY et al., 2015).

O projeto apresentado por Everingham et al. (2010) chamado de PASCAL VOC fornece: (i) conjuntos de imagens padronizados para reconhecimento de classe de objetos, (ii) um conjunto de ferramentas para acessar os conjuntos de dados e anotações, (iii) permite a avaliação e comparação de diferentes métodos. Esse projeto foi difundido entre os anos de 2005 a 2015 gerando desafios de avaliação de performance em reconhecimento de classes de objetos. A competição consiste em cinco desafios: classificação, detecção, segmentação, classificação de ações e layout de pessoa. As imagens utilizadas para o desenvolvimento do conjunto de dados disponível por esse projeto foram obtidas através do *website* "flickr"². Mais detalhes e dados ainda podem ser encontrados no *website* ³ oficial do projeto.

¹ <http://www.hunch.net/~jl/>

² www.flickr.com

³ <http://host.robots.ox.ac.uk/pascal/VOC/>

O trabalho apresentado por [Deng et al. \(2009\)](#) chamado de ImageNet é um conjunto de imagens organizado de acordo com a hierarquia do WordNet ([FELLBAUM, 1998](#)) considerando os substantivos. Há mais de 100.000 *synsets* no WordNet, a maioria deles são substantivos (80.000+). No ImageNet, é fornecido, em média, 1000 imagens para ilustrar cada *synset*. As imagens de cada conceito são controladas pela qualidade e anotadas por humanos. Atualmente, esse *dataset* é composto por uma média de mais de quinhentas imagens por nodo em um total de 14.197.122 imagens. Este projeto apresenta também um desafio anual desde o ano de 2010 chamado de Desafio de Reconhecimento Visual de Grande Escala ImageNet (ILSVRC) ([RUSSAKOVSKY et al., 2015](#)). Este desafio é proposto para avaliação de algoritmos para detecção de objetos e classificação de imagens em grande escala. Esse desafio tem o intuito de permitir que pesquisadores comparem o progresso dos algoritmos de detecção em uma variedade mais ampla de objetos. Outra motivação é medir o progresso da visão computacional para indexação de imagens em grande escala para recuperação e anotação de imagens. As imagens contidas nesse *dataset* foram coletadas da internet por meio de consulta em vários motores de busca. Mais detalhes e dados sobre o ImageNet podem ser encontrados no *website* ⁴ oficial do projeto.

Em outra vertente da visão computacional, o trabalho proposto por [Mavridaki e Mezaris \(2014\)](#) fornece um *dataset* abrangente de imagens consistindo de imagens não desfocadas, naturalmente desfocadas e artificialmente desfocadas para fins de avaliação de qualidade de imagem. Esse conjunto de imagens chamado de CERTH consiste em 2450 imagens digitais, das quais 1850 são fotografias capturadas por vários modelos de câmara em diferentes condições de disparo que não foram alteradas de qualquer forma após a sua captura. Os 600 restantes são imagens artificialmente desfocadas. Para sua criação, 60 imagens não distorcidas foram selecionadas aleatoriamente e, em seguida, foram aplicados vários tipos de filtros gaussianos, de movimento e de média circular. Mais informações sobre os dados podem ser encontrados no *website* ⁵ oficial do projeto. As imagens desse *dataset* não estão publicamente disponíveis.

No *dataset* apresentado por [Raskar, Agrawal e Tumblin \(2006\)](#) é disponibilizado quatro conjuntos de dados para avaliar diferentes algoritmos de borrão de movimentos e procedimentos de captura. Esses conjuntos de dados incluem: (a) vídeos de alta velocidade, (b) imagens de exposição codificadas e (c) imagens de exposição variadas. Todas as imagens presentes nesse *dataset* foram capturados usando uma câmara estática. Para download das imagens e vídeos e mais informações sobre os dados podem ser encontrados no *website* ⁶ oficial do projeto.

No trabalho apresentado por [Shi, Xu e Jia \(2014\)](#) chamado de "Blur Detection Dataset" é disponibilizado um conjunto de dados para detecção de borrão que contém

⁴ <http://www.image-net.org/challenges/LSVRC/>

⁵ <http://mklab.itl.gr/project/imageblur>

⁶ <http://www.amitkagrawal.com/MotionBlur/>

1000 imagens com regiões de borrão com seus respectivas versões de *ground-truth* rotuladas por humanos. Esses dados fornecem recursos úteis para entender o desfoque em relação à diversidade de estruturas em imagens naturais. As imagens contidas nesse *dataset* foram coletadas da internet e as versões de *ground-truth* foram produzidas por 10 pessoas devidamente treinadas. O conjunto de dados consiste em imagens com borrões causados por desfocagem e imagens que contém borrões parciais. Mais informações e as imagens para download podem ser encontradas no *website* ⁷ oficial do projeto.

Para área de segmentação de imagem e aplicações que necessitam de pares de imagens RGB-Mapa de profundidade, alguns *datasets* são muito difundidos (SILBERMAN; FERGUS, 2011) (SILBERMAN DEREK HOIEM; FERGUS, 2012). Esses dois *datasets* chamados NYU-Depth Versão 1 e 2 são composto por uma sequência de vídeo de uma variedade de cenas internas gravadas por ambas as câmeras RGB e de profundidade do Microsoft Kinect V1. A primeira versão desse *dataset* (SILBERMAN; FERGUS, 2011) é composto por: (1) Um subconjunto dos dados de vídeo acompanhados por rótulos densos de várias classes. Estes dados também foram pré-processados para preencher os rótulos de profundidade em falta, (2) dados brutos de RGB, profundidade e acelerômetro fornecidos pelo Kinect, (3) funções úteis para manipular os dados e as etiquetas, (4) conjunto de treinamento e teste usado para avaliação. A segunda versão desse *dataset* (SILBERMAN DEREK HOIEM; FERGUS, 2012) é composto por: 1449 pares de imagens RGB e de profundidade alinhadas rotuladas, 464 cenas novas tiradas de 3 cidades e 407.024 novos quadros não rotulados. Além disso, a segunda versão conta com outros componentes como: Um subconjunto dos dados de vídeo acompanhados por rótulos densos de várias classes. Estes dados também foram pré-processados para preencher os rótulos de profundidade em falta, os dados brutos de RGB, profundidade e acelerômetro fornecidos pelo Kinect e funções úteis para manipular os dados e as etiquetas.

Mais recentemente na área de fluxo óptico e carros autônomos alguns *datasets* são amplamente conhecidos, como no caso dos *datasets* propostos por Baker et al. (2011) Menze e Geiger (2015) Geiger et al. (2013). No trabalho proposto por Baker et al. (2011) é proposto além de um conjunto de dados uma metodologia para avaliação de algoritmos de fluxo óptico. Nesse trabalho é disponível quatro tipos de dados para testar diferentes aspectos de algoritmos de fluxo óptico: (1) Sequências com movimento não rígido em que a versão de *ground-truth* do fluxo é determinado pelo rastreamento de texturas fluorescentes ocultas, (2) sequências sintéticas realistas, (3) vídeo de alta taxa de quadros usado para estudar o erro de interpolação e (4) sequências estéreo modificadas de cenas estáticas. Na configuração desse *dataset* foi utilizado uma câmera Canon EOS 20D configurada para capturar fotos com resolução de 3504 x 2336. O conjunto de dados é dividido em 12 sequências de treinamento com seu respectivo *ground-truth*, que pode ser usado para

⁷ <http://www.cse.cuhk.edu.hk/leojia/projects/dblurdetect/dataset.html>

estimação de parâmetros ou 12 sequências de teste, onde a versão de *ground-truth* é retido. Mais informações e download dos dados podem ser obtidos no *website*⁸ oficial do projeto.

O trabalho apresentado por Menze e Geiger (2015) consiste em 200 cenas de treino e 200 cenas de teste (4 imagens coloridas por cena, salvo em formato png o que permite menos perda de qualidade). Esse conjunto de dados apresenta cenas dinâmicas para as quais a versão de *ground-truth* foi estabelecida em um processo semi-automático. Esse projeto apresenta também uma metodologia de avaliação de métodos de fluxo óptico que calcula a porcentagem de pixels ruins em média sobre todos os pixels da versão de *ground-truth* de todas as 200 imagens de teste. Os detalhes sobre o formato de dados, bem como funções de utilitário MATLAB / C++ para leitura e escrita de mapas de disparidade e campos de fluxo podem ser encontrados no *website*⁹ oficial do projeto.

O trabalho apresentado por Geiger et al. (2013) chamado de KIITI *dataset* foi capturado utilizando uma plataforma de condução autônoma chamada AnnieWAY¹⁰ para desenvolver novos e desafiadores benchmarks de visão computacional para serem utilizados na avaliação de algoritmos aplicados aos problemas do mundo real. As tarefas de interesse desse conjunto de dados são: estéreo, fluxo óptico, odometria visual, detecção de objetos 3D e rastreamento 3D. Para isso, um carro equipado com duas câmeras de alta resolução de vídeo coloridas e em tons de cinza foi utilizado. Um *ground-truth* preciso é fornecido por um *scanner* de laser Velodyne e um sistema de localização GPS. Os conjuntos de imagens são capturados por condução em torno da cidade de Karlsruhe, em áreas rurais e em rodovias. Até 15 carros e 30 pedestres são visíveis por imagem. Além de fornecer todos os dados em formato raw, o *benchmark* é extraído para cada tarefa. Para cada um dos *benchmarks*, também é fornecido uma métrica e um sistema de avaliação. Mais informações sobre o *dataset*, além do sistema de avaliação e o conjunto dados para *download* pode ser encontrado no *website*¹¹ do projeto.

Para a identificação e estimativa de estradas e pistas para área de carros autônomos, foi criado um *dataset* composto por 289 imagens de treinamento e 290 imagens de teste (FRITSCH; KUEHNL; GEIGER, 2013). Esse conjunto de dados contém três categorias diferentes de cenas de estrada: (1) imagens urbanas não rotuladas, (2) imagens urbanas rotuladas, (3) múltiplas vias rotuladas, (4) uma combinação das três anteriores. O *ground-truth* foi gerado por anotação manual das imagens e disponível para dois tipos de terreno diferentes: estrada - a área da estrada, ou seja, a composição de todas as pistas - a via, ou seja, a pista do veículo. A versão de *ground-truth* é fornecida apenas para imagens de treinamento. Mais informações e download dos dados podem ser acessadas no *website*¹²

⁸ <http://vision.middlebury.edu/flow/>

⁹ http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=flow

¹⁰ <http://www.mrt.kit.edu/annieway/>

¹¹ <http://www.cvlibs.net/datasets/kitti/index.php>

¹² http://www.cvlibs.net/datasets/kitti/eval_road.php

oficial do projeto.

Existem ainda vários outros *datasets* publicamente ou não publicamente disponíveis dependendo da área de aplicação. As informações de alguns *datasets* mais recentemente publicados podem ser acessados tanto nos sites descritos nessa seção ou em análises e levantamentos como a apresentada por [Firman \(2016\)](#) que descreve os *datasets* no passado, no presente e as perspectivas de criação no futuro.

3.1 Trabalhos Relacionados

Alguns *datasets* são desenvolvidos com intuito de isolar propriedades importantes para a avaliação de algoritmos em diferentes áreas. Porém, são poucos os trabalhos na literatura a disponibilizarem *datasets* específicos para uso e análise de algoritmos aplicáveis a ambientes subaquáticos. A simulação da degradação causadas em imagens capturada em ambientes subaquáticos é uma tarefa complexa tendo em vista as propriedades fotométricas decorrentes das interações entre o meio fluídico, a iluminação e os objetos presentes na cena. Esses efeitos ocasionados pela propagação da luz são dependentes normalmente do espalhamento natural da luz na água e do tamanho, forma e densidade das partículas presentes no ambiente([HULST; HULST, 1957](#)).

Outro desafio está relacionado com a dificuldade de reprodução em menor escala em ambiente controlado da diversidade e da especificidade associadas à paisagens subaquáticas. Este é um fator limitante à construção de *datasets* de imagens subaquáticas que possam ser utilizadas na avaliação de métodos de restauração dessas imagens, por exemplo.

Até o presente momento, poucos são os trabalhos que reproduzem em ambientes controlados os aspectos de degradação desse tipo de imagem. No trabalho exposto por [Garcia e Gracias \(2011\)](#) é apresentado um ambiente controlado composto por um sistema de vídeo formado por três câmeras Rolera-XR, uma fonte de luz semelhante a um ponto e um tanque de 400 litros de água revestido internamente com material preto anti-reflexivo. A fonte de luz utilizada foi uma lâmpada halógena de 12 V, colocada em um invólucro subaquático e conectada a uma fonte de alimentação estável para ter intensidade de luz constante durante todo o experimento. A fonte de luz foi fixada na água, a 5 cm acima do plano dos centros da câmara, à distância de 20, 25 e 45 cm de cada câmaras, respectivamente. Para a composição da cena fotografada foi utilizado somente um conjunto pequeno de estruturas naturais como areia, rochas e conchas que representam o solo oceânico mas que não provém características suficientes desses ambientes nem uma variedade de estruturas e cores que contenham uma diversidade maior de informações. Este trabalho foi desenvolvido com uma abordagem diferente, onde o principal objetivo é a comparação de performance dos detectores de pontos de interesse mais populares na literatura quando

aplicados em imagens turvas. A geração de 10 diferentes níveis turbidez foi gerada a partir da adição de leite desnatado na água.

No trabalho apresentado por Nascimento, Campos e Barros (2009) foi utilizado objetos simples com formas conhecidas onde a geração de dois níveis de turbidez foi obtida através da utilização de diferentes quantidades de chá verde dissolvidos em água. Esse procedimento é utilizado pela comunidade de pesquisas hidráulicas para calibração de instrumentos como turbidímetros. Nesse trabalho foi utilizado um sistema de visão estéreo composto por duas câmeras do tipo *Dragon Fly* calibradas geometricamente ($f=8\text{mm}$, resolução de 640×480 a 30 frames por segundo), com uma linha de base de 33 cm. As duas câmeras foram posicionadas na posição fronto- paralelo na parte externa de um aquário de vidro medindo $168 \text{ cm} \times 47 \text{ Cm} \times 45 \text{ cm}$, que acomodava 320 litros de água. O sistema de iluminação utilizado nesse experimento foram quadro pares de lâmpadas fluorescentes posicionados à 2 metros acima do aquário.

Posteriormente, no trabalho proposto por Codevilla et al. (2015b), foi apresentado o *dataset* **TURBID**. Mais características oceânicas e informações estruturais foram apresentadas em imagens reais previamente fotografadas nas Bahamas. Porém, essas imagens foram impressas e colocadas em uma lâmina de vidro, resultando em ruídos indesejados providos pelo processo de impressão. Além disso, esse trabalho não considerara diferentes distâncias dos objetos presentes na cena até a câmera. Com isso, a validação de algoritmos que dependem da variação de distância se tornam inviáveis. Esse experimento apresenta 19 diferentes níveis de turbidez gerados a partir da adição de leite integral em 1000 litros de água. O sistema de iluminação é constituído de dois pares de lâmpadas fluorecentes posicionadas a 1 metro acima do tanque de plástico utilizado. O *setup* utilizado neste trabalho serviu de base para a criação do *dataset* apresentado na presente dissertação. A configuração do experimento bem como a metodologia adotada foram utilizadas e serão melhor descritas no Capítulo 4.

4 Criação dos *Datasets* Propostos

Este capítulo apresenta a descrição dos *datasets* produzidos durante o desenvolvimento do presente trabalho.

Inicialmente, é abordado o desenvolvimento do *dataset* aqui chamado de **TURBID 3D**. Para a criação deste *dataset* diversas cenas foram preparadas além um ambiente onde a iluminação e a quantidade de degradação presente em cada imagem capturada pode ser observada e controlada.

O processo de criação de cada conjunto de imagens produzido é chamado aqui de experimento. Cada experimento conta com a mesma estrutura e organização porém a combinação entre cenas e substâncias utilizadas são diferentes. Para diferenciar essa combinação, cada experimento é apresentado dentro do *dataset* como um conjunto de imagens nomeado diferentemente. Os cinco conjuntos de imagens apresentados são: *Milk*, *GreenTea*, *Chlorophyll*, *DeepBlue*, *Blue*. Cada conjunto de imagens conta com 19 diferentes níveis de turbidez além da imagem de referência.

A descrição do *dataset*, a estrutura utilizada, a metodologia adotada bem como a apresentação dos 5 conjuntos de imagens serão abordadas na [seção 4.1](#).

Em seguida, a criação do *dataset* contendo imagens com degradação sintética produzido a partir do simulador aqui chamado de **TURBI** será apresentada. O desenvolvimento do simulador bem como a geração das imagens serão descritos na [seção 4.2](#).

4.1 Aquisição de Imagens Turvas em um Espaço Controlado

A configuração e a metodologia adotada para a criação do *dataset TURBID 3D* tem como base o trabalho apresentado por [Codevilla et al. \(2015b\)](#). Nas subseções a seguir serão descritas as configurações criadas, as cenas fotografadas bem como a metodologia adotada para produção das imagens que constituem o *dataset*.

4.1.1 Criação do *dataset* TURBID 3D: Descrição do Experimento

No desenvolvimento do *dataset TURBID 3D* diversas cenas foram preparadas bem como um ambiente onde a iluminação e a quantidade de degradação presente em cada imagem obtida pôde ser monitorada. Essa configuração foi desenvolvida com o objetivo de manter o ambiente controlável.

Para isso, a câmera e o cenário utilizados permaneceram estáticos, a iluminação pôde ser controlada e se manteve o mais uniforme possível durante todos os experimentos

realizados, respeitando as restrições do modelo físico adotado.

A descrição do *dataset* bem como a estrutura utilizada, as cenas fotografadas e a metodologia adotada serão descritos nas subseções a seguir.

4.1.1.1 Estrutura Desenvolvida

Para a aquisição deste *dataset* foi preparado um tanque de vidro, com dimensões de 1 x 0,80 x 1,30 metros (Largura x Altura x Comprimento). As laterais do tanque foram revestidas com um material translúcido e anti-reflexivo, a fim de permitir que a iluminação incidisse no tanque apenas pela parte superior. Isso permitiu que os raios de luz vindos de fora da estrutura não interferissem na cena.

A estrutura conta com uma iluminação composta por duas lâmpadas *Light Emitting Diode* - Diodo Emissor de Luz (**LED**) de 9,5 Watts de potência (equivalência de 75 Watts), com fluxo luminoso de 1.075 lúmens e cor de 6500 *kelvins*, equivalente a luz branca. As lâmpadas foram instaladas cada uma em uma luminária no interior de um *softbox* produzido com material refletor e difusor a fim de manter a luz contínua e uniforme durante todo experimento. Cada luminária foi disposta a 30 cm de distância das laterais e a uma altura de 1 metro da superfície do tanque.

No fundo do tanque foi posicionada uma bandeja, também de vidro, com dimensões de 70 x 2 x 100 centímetros (Largura x Altura x Comprimento) à uma altura de 5 centímetros (*cm*) da parte inferior. Essa bandeja é utilizada para a montagem do cenário a ser fotografado facilitando a composição e trocas das cenas.

A estrutura principal deste experimento descrita acima pode ser observada na Figura 4.

Para a captura das imagens foi utilizada uma câmera GoPro Hero3+ Black Edition ¹ presa a um suporte perpendicular a cena. A câmera foi mantida com a sua configuração padrão e setada para fotografar em uma resolução de 12 mega pixels (3000x4000). A câmera foi colocada perpendicular ao fundo do tanque a uma altura de 70 cm em um suporte projetado para se manter estático durante todo o experimento.

Nos experimentos onde foram capturadas as imagens dos conjuntos *GreenTea*, *Chlorophyll* e *Blue* a estrutura conta ainda com outra câmera GoPro Hero3+ Black com as mesmas configurações posicionada na parte interna presa a um suporte estático posicionado na lateral do tanque. A câmera encontra-se no suporte a uma altura de 35 cm do fundo do tanque e 40 cm das laterais posicionada para a frente.

A estrutura descrita acima foi colocada em uma sala totalmente escura onde a iluminação era obtida somente através das luminárias utilizadas.

¹ Para maiores informações sobre as especificações da câmera, o manual está disponível em: <http://goo.gl/O6htD3>



Figura 4 – Estrutura principal do experimento proposto composta por um tanque de vidro com dimensões de 1 x 0,80 x 1,30 metros (Largura x Altura x Comprimento), duas lâmpadas LED colocadas em um *softbox*, uma de cada lado a 30 cm do tanque e uma bandeja de vidro que pode ser utilizada para montar a cena a ser fotografada.

4.1.1.2 Cenas Fotografadas

Foram criadas três diferentes cenas a fim de diversificar as estruturas, texturas, cores objetos utilizados. As cenas foram criadas com o intuito de representar o máximo possível cenários encontrados em ambientes subaquáticos reais. Para isso, foram coletados diversos objetos em ambientes naturais tais como na Ilha da Pólvora e na praia do Cassino, ambos situados na cidade de Rio Grande no Rio Grande do Sul. Além disso, rochas marinhas mortas e decorações que imitam formas e características de corais foram utilizadas.

A seguir serão apresentados as três cenas criadas bem como os objetos utilizados em cada uma.

Cena 1

A cena 1 contém pedras brancas pequenas na base que caracterizam o fundo do oceano, pedras maiores de diversas cores, decorações que imitam as formas e características de corais, conchas, rochas marinhas mortas e alguns objetos feitos pelo homem como objetos de decoração nas cores verdes e vermelho e um cano de plástico na cor branca. As pedras e conchas contidas nessa cena assim como os objetos de decoração verdes e vermelhos foram retiradas de ambientes naturais na praia do Cassino localizada na cidade de Rio Grande no Rio Grande do Sul. A [Figura 5](#) exibe a organização e os objetos utilizados na cena 1.

Cena 2



Figura 5 – Cena 1 montada para a criação dos conjuntos de dados.

A cena 2 contém pedras brancas pequenas na base que caracterizam o fundo do oceano, pedras grandes de diferentes cores, uma boia utilizada em redes de pesca, uma vara de pescar, uma bota de couro, uma garrada de vidro azul, uma lata de alumínio na cor cinza, uma vara de madeira com corais mortos no seu entorno e objetos de decoração de alumínio nas cores azul e verde. Todos esses objetos contidos nessa cena foram retirados em ambientes naturais como na praia do Cassino e na Ilha da Pólvora ambos localizados na cidade de Rio Grande no Rio Grande do Sul. A [Figura 6](#) exibe a organização e os objetos utilizados na cena 2.

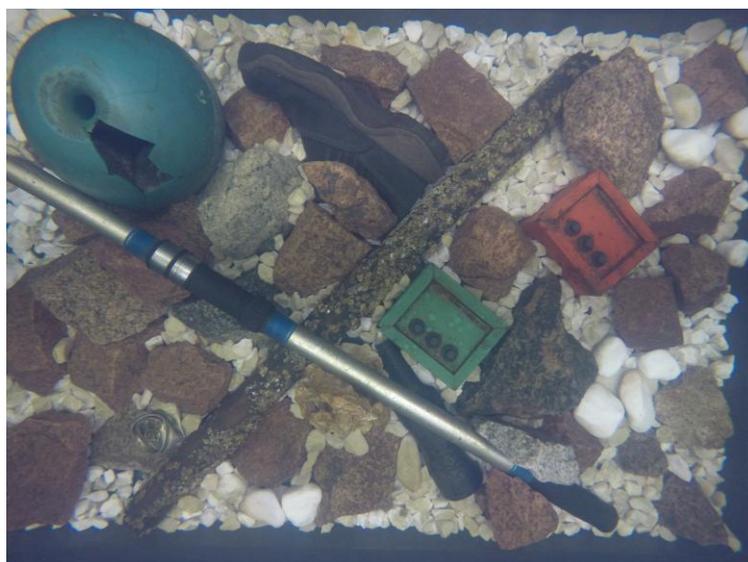


Figura 6 – Cena 2 montada para a criação dos conjuntos de dados.

Cena 3

A cena 3 é uma combinação das duas cenas anteriores com a adição de alguns objetos. A cena contém pedras brancas pequenas na base que caracterizam o fundo do

oceano, pedras grandes de diferentes cores, uma boia utilizada em redes de pesca, uma bota de couro, decorações que imitam as formas e características de corais, conchas, rochas marinhas mortas e alguns objetos feitos pelo homem como objetos de decoração nas cores verdes e vermelho e um cano de plástico na cor branca. A [Figura 7](#) exhibe a organização e os objetos utilizados na cena 3.



Figura 7 – Cena 3 montada para a criação dos datasets.

4.1.1.3 Metodologia adotada

A metodologia adotada durante todos os experimentos conta com cinco etapas como ilustrado na [Figura 8](#), sendo elas:

- 1) Organização e composição do cenário;
- 2) Organização da estrutura;
- 3) Adição consecutiva da substância escolhida para produção da turbidez;
- 4) Captura das imagens;
- 5) Organização das imagens finais no *dataset*;

1) Organização e composição do cenário:

A primeira etapa constitui na organização e composição da cena a ser fotografada. Os cenários montados bem como a descrição dos objetos utilizados em cada experimento pode ser visto na [subseção 4.1.1.2](#).

Os cenários utilizados foram montados em uma bandeja de vidro com dimensões de 70 x 2 x 100 *cm* (Largura x Altura x Comprimento) colocada posteriormente à uma

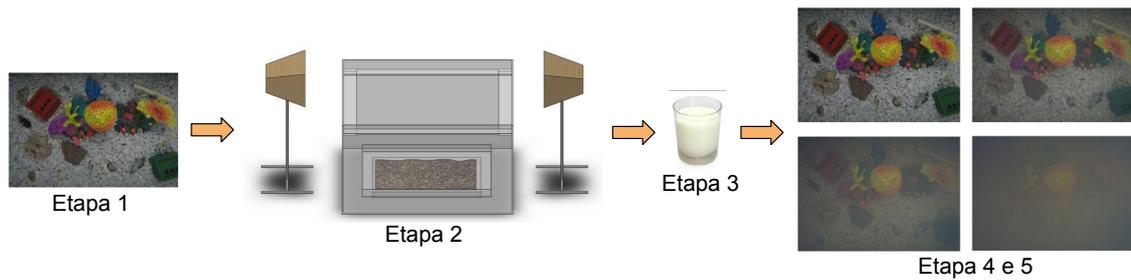


Figura 8 – Ilustração das etapas da metodologia adotada.

altura de 5 *cm* da parte inferior do tanque. Estas cenas foram organizadas e montadas para permanecerem estáticas durante todo o experimento.

2) Organização da estrutura:

Posteriormente, na segunda etapa o cenário criado é colocado dentro do tanque e a estrutura descrita na [subseção 4.1.1.1](#) é montada em uma sala totalmente escura. Para isso as luminárias foram posicionadas a 30 *cm* nas laterais do tanque e a uma altura de 1 metro da superfície.

Após a estrutura configurada, o tanque é preenchido com 910 litros de água pura, ficando o cenário e as câmeras utilizadas totalmente submersos como mostra na [Figura 9](#).

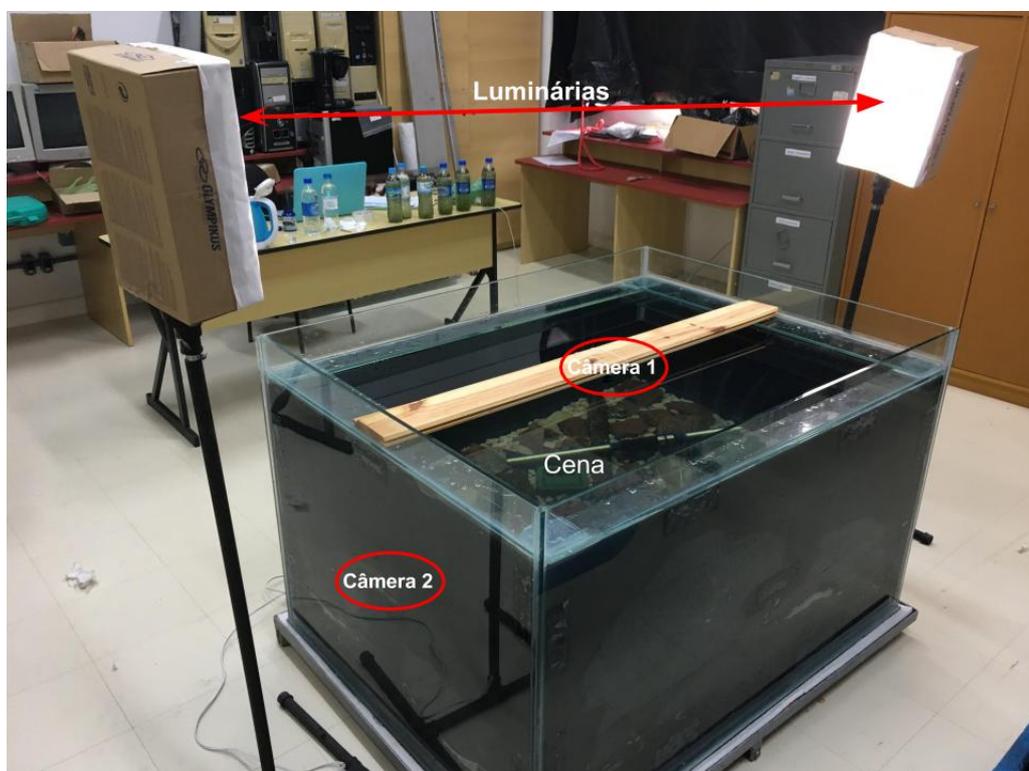


Figura 9 – Estrutura principal do experimento proposto com água e as câmeras posicionadas.

3) Adição consecutiva da substância escolhida para produção da turbi-

dez:

O objetivo principal deste experimento é criar um ambiente controlado onde o efeito de *backscattering* seja dominante. Segundo [Schechner e Karpel \(2005\)](#), em meios participativos o *backscattering* é o fator responsável pela maior parte da perda de contraste e visibilidade na formação da imagem. Sendo assim, a componente *forward scattering* tem uma pequena contribuição na degradação total da imagem.

Com isso, algumas substâncias que contenham partículas com dimensões suficientes para a produção desse fenômeno foram escolhidas e utilizadas durante os experimentos realizados, sendo elas:

Leite Integral:

Um estudo apresentado por [Narasimhan et al. \(2006\)](#) apontou que o leite integral possui uma grande quantidade de partículas que provocam o espalhamento grande angular da luz em ambientes subaquáticos, aumentando assim o efeito do *backscattering*. Esse procedimento foi utilizado e validado anteriormente nos trabalhos apresentados por [Garcia e Gracias \(2011\)](#) e [Codevilla et al. \(2015b\)](#). Para a indução deste efeito, foi utilizado um total de 190 mililitros (*ml*) de leite integral diluídos em diferentes quantidades no total 910 litros de água presente no tanque.

A quantidade de leite adicionada em cada etapa foi previamente avaliada em um experimento em pequena escala, onde foi possível definir a quantidade de leite necessária a ser diluída no montante de água para a produção da turbidez desejada.

A [Figura 4.1.1.3](#) apresenta a quantidade de leite utilizada para produzir os diferentes níveis de turbidez de cada uma das imagens capturadas que compõem o conjunto de imagens chamado de *Milk*. Essa mesma proporção foi utilizada nos conjuntos *Blue* e *DeepBlue*.

Chá Verde:

No segundo conjunto de imagens foi utilizado chá verde dissolvido em água para gerar os diferentes níveis de turbidez. A utilização de chá em geral para aumentar a turbidez da água é um procedimento comumente utilizado pela comunidade de pesquisas de recursos hidráulicos para calibração de instrumentos como turbidímetros. A utilização de chá verde em específico foi utilizada e validada no trabalho apresentado por [Nascimento, Campos e Barros \(2009\)](#). Segundo os autores, a utilização de chá verde é vantajosa pela homogeneidade e pela cor verde produzida muito comumente encontradas nos oceanos.

Durante o experimento para a produção dos 19 níveis de turbidez foi utilizado um total de 116 mg de chá verde diluídos em diferentes quantidades no total de água presente no tanque, produzindo assim o conjunto de imagens com diferentes níveis de degradação.

A quantidade de chá verde adicionada em cada etapa foi previamente avaliada em

Imagem (I_i)	Quantidade Adicionada de Leite	Quantidade Total de Leite
I_1	5 ml	5 ml
I_2	5 ml	10 ml
I_3	5 ml	15 ml
I_4	5 ml	20 ml
I_5	5 ml	25 ml
I_6	5 ml	30 ml
I_7	6 ml	36 ml
I_8	6 ml	42 ml
I_9	8 ml	50 ml
I_{10}	8 ml	58 ml
I_{11}	8 ml	66 ml
I_{12}	8 ml	74 ml
I_{13}	8 ml	82 ml
I_{14}	8 ml	90 ml
I_{15}	10 ml	100 ml
I_{16}	10 ml	110 ml
I_{17}	10 ml	120 ml
I_{18}	10 ml	130 ml
I_{19}	60 ml	190 ml

Tabela 1 – Quantidade de leite integral adicionado durante a captura de cada uma das imagens no experimento. Na primeira coluna é apresentado o nome da imagem produzida. Seguido pela quantidade correspondente de leite adicionado para a sua produção, apresentado na segunda coluna. Na terceira coluna é apresentado a quantidade total de leite diluído no montante de água. No total foram diluídos 190 ml de leite integral em diferentes porções em 910 litros de água para a produção de 19 diferentes níveis de turbidez da cena.

um experimento em pequena escala, onde foi possível definir a quantidade de chá verde necessária a ser diluída na água para a produção da turbidez desejada. As quantidades estabelecidas foram primeiramente colocadas em processo de infusão em 500 ml de água quente.

A [Tabela 4.1.1.3](#) apresenta a quantidade de chá diluído em água adicionado em cada etapa do experimento.

As imagens geradas pelo experimento onde essa substância foi utilizada compõem o conjunto de imagens *GreenTea*.

Clorofila:

No terceiro conjunto de imagens foi utilizado Clorofila em pó dissolvida em água para gerar os diferentes níveis de turbidez.

Clorofila é a denominação de um grupo de pigmentos verdes responsáveis pela a realização da fotossíntese presente em cianobactérias e os cloroplastos das algas e plantas. A intensa cor verde da clorofila se deve a suas fortes absorções das regiões azuis e vermelhas do espectro eletromagnético. Devido a essa absorção a luz refletida por esses pigmentos parece verde ([MACKINNEY, 1941](#)).

Imagem (I_i)	Quantidade de Chá Verde (mg)	Quantidade total de Chá Verde (mg)
I_1	3 mg	3 mg
I_2	3 mg	6 mg
I_3	3 mg	9 mg
I_4	3 mg	12 mg
I_5	3 mg	15 mg
I_6	3 mg	18 mg
I_7	4 mg	22 mg
I_8	4 mg	26 mg
I_9	6 mg	32 mg
I_{10}	6 mg	38 mg
I_{11}	6 mg	44 mg
I_{12}	6 mg	50 mg
I_{13}	6 mg	56 mg
I_{14}	6 mg	62 mg
I_{15}	9 mg	71 mg
I_{16}	9 mg	80 mg
I_{17}	9 mg	89 mg
I_{18}	9 mg	98 mg
I_{19}	18 mg	116 mg

Tabela 2 – Quantidade de chá verde adicionado durante a captura de cada uma das imagens no experimento. Na primeira coluna é apresentado o nome da imagem produzida. Seguido pela quantidade correspondente de chá verde diluído para a sua produção, apresentado na segunda coluna. Na terceira coluna é apresentado a quantidade total de chá verde adicionado no montante de água. No total foram diluídos 116mg de chá verde em diferentes porções em 910 litros de água para a produção de 19 diferentes níveis de turbidez da água.

Para a indução do efeito de *backscattering* e coloração da água a fim de representar os ambientes subaquáticos naturais que apresentam a cor verde na água verde, foi utilizado um total de 116 mg de clorofila na forma isolada.

Para a produção dos 19 níveis de turbidez gerados com essa substância, o total de 116 mg foi dissolvido em diferentes porções em 500 ml de água pura e adicionados separadamente no montante de 910 litros de água presentes no tanque.

A quantidade de clorofila adicionada em cada etapa foi previamente avaliada em um experimento em pequena escala, onde foi possível definir a quantidade necessária a ser diluída na água para a produção da turbidez desejada.

A [Tabela 4.1.1.3](#) apresenta a quantidade de clorofila em pó diluída em 500 ml de água e adicionado em cada etapa do experimento.

As imagens geradas pelo experimento onde essa substância foi utilizada compõem o conjunto de imagens *Chlorophyll*.

Azul de Metileno:

O cloreto de metiltionínio ou também conhecido popularmente como azul de meti-

Imagem (I_i)	Quantidade de clorofila (mg)	Quantidade total de clorofila (mg)
I_1	3 mg	3 mg
I_2	3 mg	6 mg
I_3	3 mg	9 mg
I_4	3 mg	12 mg
I_5	3 mg	15 mg
I_6	3 mg	18 mg
I_7	4 mg	22 mg
I_8	4 mg	26 mg
I_9	6 mg	32 mg
I_{10}	6 mg	38 mg
I_{11}	6 mg	44 mg
I_{12}	6 mg	50 mg
I_{13}	6 mg	56 mg
I_{14}	6 mg	62 mg
I_{15}	9 mg	71 mg
I_{16}	9 mg	80 mg
I_{17}	9 mg	89 mg
I_{18}	9 mg	98 mg
I_{19}	18 mg	116 mg

Tabela 3 – Quantidade de Clorofila adicionado durante a captura de cada uma das imagens no experimento. Na primeira coluna é apresentado o nome da imagem produzida. Seguido pela quantidade correspondente de Clorofila adicionado para a sua produção, apresentado na segunda coluna. Na terceira coluna é apresentado a quantidade total de Clorofila diluída no montante de água. No total foram diluídos 116 mg de Clorofila em pó em diferentes porções em 910 litros de água para a produção de 19 diferentes níveis de turbidez da cena.

leno é um composto aromático heterocíclico, sólido, de cor azul escura e solúvel em água. Este composto é comumente utilizado em muitas aplicações nas mais variadas áreas como na biologia e na química como um corante para diferentes procedimentos de coloração sendo ele de fácil solução de cor homogênea quando dissolvido em água ([DRUGS.COM, 2017](#)).

Apesar deste composto não conter partículas suficientes para a produção do efeito de *backscattering* desejado, a azul de metileno foi utilizado para a produção da cor azul na água, a fim de representar os ambientes subaquáticos naturais que apresentam essa característica. O efeito de *backscattering* foi produzido pela adição de leite integral ([NARASIMHAN et al., 2006](#)) em diferentes quantidades conforme apresentado pela [Figura 4.1.1.3](#).

O azul de metileno foi utilizado em diferentes porções para a produção de dois conjuntos de imagens presentes no *dataset* TURBID 3D. Foi utilizado uma quantidade de 15 ml de azul de metileno diluídos nos 910 litros de água para a produção do conjunto de imagens chamado de *Blue*. E posteriormente, foi utilizado o dobro desta quantidade, 30 ml também diluídos em 910 litros de água para a produção do conjunto de imagens chamado de *DeepBlue*.

4) Aquisição e Organização das imagens:

O procedimento para captura das imagens foi o mesmo para todos os experimentos realizados. Para a obtenção da imagem de referência foram fotografadas 30 imagens em água limpa, sem nenhuma adição de qualquer substância.

No caso dos conjuntos de imagens *Blue* e *DeepBlue* onde a substância chamada azul de metileno foi dissolvida no montante de água presente no tanque para a produção da cor azul um conjunto de 30 fotos foram capturadas também antes da adição de qualquer outra substância.

Logo após, a turbidez e conseqüentemente a quantidade de degradação foram elevados de uma maneira controlada por sucessivas adições da substância escolhida para a produção do efeito de *backscattering*. A quantidade necessária adicionada em cada experimento foi diferente conforme a necessidade.

Para cada concentração da substância adicionada, foram fotografadas 30 imagens com um intervalo de 10 segundos entre cada foto. Esse procedimento foi estabelecido para minimizar a possível variação de iluminação entre cada captura e estabelecer a completa diluição da substância na água.

Esse procedimento foi repetido 19 vezes com diferentes concentrações da substância utilizada. Com isso, foram produzidos 19 diferentes níveis de turbidez além da imagem de referência. Sendo assim, os conjuntos de imagens *Milk*, *GreenTea* e *Chlorophyll* contém 20 imagens, e os conjuntos *Blue* e *DeepBlue* onde foi obtido um estágio entre a imagem de referência e a adição da substância para a produção da turbidez, contém 21 imagens.

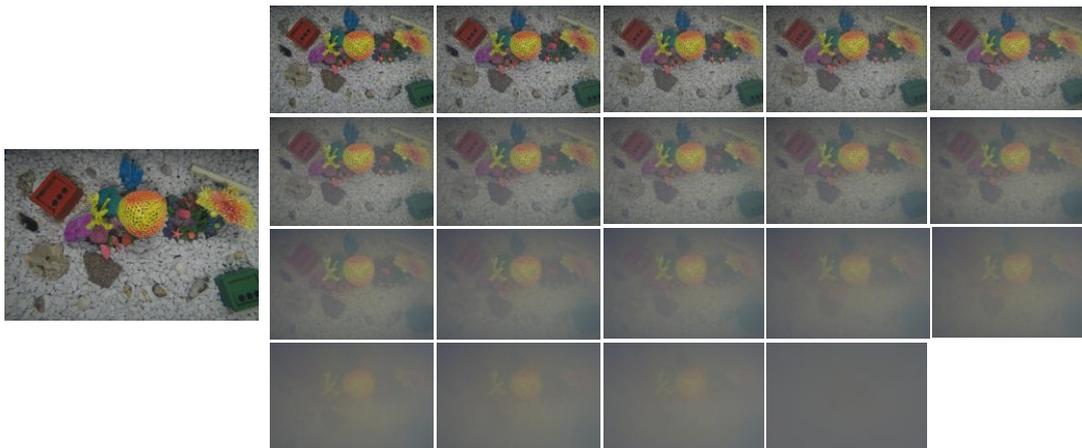
Para gerar o conjunto final de imagens com os diferentes níveis de turbidez, primeiramente foi calculado a média das 30 imagens intermediárias tiradas em água limpa. Essa imagem é considerada a imagem de referência do experimento, ou seja, a imagem sem degradação. O mesmo procedimento foi feito para os dois experimentos onde o azul de metileno foi adicionado em uma etapa intermediária. Depois disso, foi calculado também a média das 30 imagens tiradas em cada nível de turbidez produzido, criando assim as imagens $I_1 \dots I_{19}$.

Após esse procedimento, as imagens obtidas em cada experimento foram separadas em cinco diferentes grupos chamados de *Milk*, *GreenTea*, *Chlorophyll*, *Blue* e *DeepBlue*. Nos conjuntos de imagens *GreenTea*, *Chlorophyll* e *Blue* há ainda uma divisão interna entre a câmera com a visão da cena vista pela lateral e a câmera posicionada na parte superior perpendicular a cena. A [Tabela 4](#) apresenta como se deu a combinação entre cena e substância utilizada em cada um dos conjuntos de imagens.

As imagens presentes nos conjuntos *Milk*, *Chlorophyll* e *DeepBlue* podem ser vistas nas [Figuras 10](#), [11](#), [12](#) e [13](#) respectivamente.

	Substância Utilizada	Cena fotografada	Num. de Fotos
Milk	Leite Integral 3%	Cena 1	20
GreenTea	Chá Verde	Cena 3	40
Chlorophyll	Clorofila em pó	Cena 2	40
Blue	Azul de metileno + Leite integral	Cena 3	42
DeepBlue	Azul de metileno + Leite integral	Cena 2	21

Tabela 4 – Conjuntos de imagens disponíveis e suas respectivas informações

Figura 10 – Imagens contidas no conjunto *Milk* presente no *dataset* TURBID 3D.Figura 11 – Imagens contidas no conjunto *Chlorophyll* presente no *dataset* TURBID 3D.

Uma das contribuições deste trabalho se encontra na disponibilidade do *dataset* completo, que pode ser obtido pelo endereço eletrônico: www.amandaduarte.com.br/turbid. Além disso, o *dataset* e a metodologia apresentada neste trabalho podem ser utilizadas em diversas áreas da visão computacional que possuam a necessidade de dados com múltiplos níveis de turbidez.

A [Figura 14](#) apresenta algumas propriedades relevantes apresentadas em cada um dos *datasets* relacionados presentes no estado da arte comparados com o *dataset* TURBID 3D aqui apresentado.



Figura 12 – Imagens com vista lateral contidas no conjunto *Chlorophyll* presente no *dataset* TURBID 3D.



Figura 13 – Imagens contidas no conjunto *DeepBlue* presente no *dataset* TURBID 3D.

	TURBID	Nascimento et al.	Garcia et al.	TURBID 3D
Imagem de Referência	Sim	Sim	Sim	Sim
Níveis de Turbidez	19	2	12	19
Img Alta Resolução	Sim	Sim	Sim	Sim
Total de Imagens	60	9	36	163
Conjunto de imagens	3	3	1	5
Diferentes ângulos	Não	Não	Não	Sim
Disponível online	Sim	Não	Não	Sim

Tabela 5 – Propriedades dos *datasets* relacionados presentes no estado da arte em comparação com o TURBID 3D. Na primeira coluna a tabela apresenta as propriedades seguido dos *datasets* listados. As propriedades comparadas respectivamente são: A presença da imagem de referência entre as imagens do *dataset*, quantidade de níveis de turbidez presentes no *dataset*, a utilização de imagens de alta resolução, o total de imagens contidas no *dataset*, a quantidade de conjuntos de imagens diferentes é apresentado, a utilização de câmeras em diferentes ângulos e a disponibilidade para *download* das imagens. Nas colunas seguintes é apresentada as propriedades dos *datasets* proposto por [Codevilla et al. \(2015b\)](#), [Nascimento, Campos e Barros \(2009\)](#) e [Garcia e Gracias \(2011\)](#) respectivamente e na ultima coluna o *dataset* TURBID 3D apresentado no presente trabalho.

4.2 Simulação de Meios Participativos em Imagens Ópticas

A síntese de ambientes recriados computacionalmente com condições e características existentes no mundo real vem sendo amplamente utilizada em diversas áreas. Esta reprodução normalmente é feita quando o ambiente na qual pretende-se interagir é de difícil acesso ou possui particularidades que podem ser tratados e simulados por meio de algoritmos computacionais. Um exemplo desses ambientes são os meios participativos. Como visto na [seção 2.1](#), imagens capturadas em meios participativos são degradadas devido à interação dos raios de luz com as partículas suspensas no meio. Devido à essa interação, o sinal visual vindo do objeto pode ser corrompido ou sofrer uma perda de informação até chegar ao sensor, resultando em efeitos característicos como borramento, baixo contraste, ruído e distorção nas cores.

A aquisição de imagens nesses ambientes é dificultada devido às condições de acesso e recursos necessários para a captura. Além disso, o acesso ao *ground-truth* é praticamente impossível, ou seja, é inviável a captura de ambas imagens com e sem degradação da mesma cena sob as mesmas condições de luz. Em diversos casos, como por exemplo no treinamento de redes neurais ou na avaliação de métodos de restauração de imagens, *ground-truth* é extremamente necessário.

Com isso, foi desenvolvido um simulador capaz de sintetizar a degradação causada por ambientes participativos em imagens não degradadas levando em consideração o seu respectivo mapa de profundidade.

A seguir será descrito o desenvolvimento do simulador chamado aqui de TURBI, assim como as imagens geradas a partir dele.

4.2.1 Desenvolvimento do Simulador **TURBI**

O recente progresso em visão computacional tem sido direcionado pela alta capacidade de modelos de aprendizagem, tais como redes neurais treinadas em grandes conjuntos de dados ([KRIZHEVSKY; SUTSKEVER; HINTON, 2012](#); [SIMONYAN; ZISSERMAN, 2014](#); [SZEGEDY et al., 2015](#); [HE et al., 2016](#); [KARPATHY; FEI-FEI, 2015](#)). Entretanto, o treinamento de uma rede neural muitas vezes requer uma grande quantidade de dados rotulados, ou seja, entradas onde a *ground-truth* esteja disponível. No caso da restauração de imagens capturadas em meios participativos, este dado é composto por um par de imagens, ambas exatamente da mesma cena, capturadas sob as mesmas condições de luz, onde uma está corrompida pela degradação causada pelo ambiente e a outra é livre de degradação. Adquirir esses pares de imagens em uma quantidade grande o suficiente para treinar uma rede neural é uma tarefa extremamente difícil.

Para superar esse problema, durante este trabalho foi proposto simulador capaz de

gerar dados sintetizados aplicando degradação simulada sobre imagens sem degradação. Para o desenvolvimento do simulador, foi levado em consideração o modelo físico de formação de imagem, previamente elucidado na [seção 2.1](#), estendido de forma a incorporar diferentes comprimentos de onda, como mostra a [Equação 4.1](#). Para tal, adotou-se o padrão RGB de representação. Com isso é possível descrever o fenômeno através dos seus três canais de cores. O modelo físico de formação de imagem é dado por:

$$I^\lambda(x, y) = J^\lambda(x, y)e^{-c^\lambda d(x, y)} + A^\lambda(1 - e^{-c^\lambda d(x, y)}), \lambda \in \{R, G, B\}. \quad (4.1)$$

onde x, y são as coordenadas do pixel e $I(x, y)$ é a imagem degradada obtida no meio participativo. Na direita da igualdade, J é a irradiância da cena que é atenuada por e^{-cd} , onde d é a distância entre a cena e a câmera e c é o coeficiente de atenuação somado a $A(1 - e^{-cd})$ onde A é a luz global da cena. O termo e^{-cd} é conhecido também como a transmissão.

Essa equação representa uma versão simplificada ao modelo proposto por [Koschmieder \(1924\)](#). Neste estudo foi adotado este modelo que foi estabelecido como uma descrição dos efeitos atmosféricos apresentados sobre o observador ([SCHECHNER; NARASIMHAN; NAYAR, 2001; NARASIMHAN; NAYAR, 2002; HAUTIERE; TAREL; AUBERT, 2007; SCHECHNER; KARPEL, 2004](#)). Segundo esse modelo em condições de tempo limpo, ou seja, onde não há partículas suspensas no ambiente, ao ar livre, o brilho de um ponto de cena atingiria o observador quase inalterado. No entanto, quando a imagem é capturada em meios participativos, a irradiância observada em cada pixel é uma combinação linear da radiação do objeto da cena transmitida diretamente que será atenuada na linha de visão e dispersa em direção ao observador.

Com essa definição, métodos de restauração de imagens são desenvolvidos baseando-se no inverso desse modelo. Para gerar a degradação simulada, é necessário assim como na restauração a estimativa de alguns parâmetros, tais como: a cor da luz global da cena, o coeficiente de atenuação do ambiente e o mapa de profundidade da cena.

Por conseguinte, O modelo Koschmieder foi adaptado neste trabalho para condições de iluminação subaquática, como em ([DUNTLEY, 1963; SMITH; BAKER, 1981; ROSER; DUNBABIN; GEIGER, 2014](#)). Com isso, a luz global da cena à uma profundidade z está sujeita a uma atenuação da luz da seguinte forma:

$$A(z) = TI_0e^{-c^\lambda z}, \lambda \in \{R, G, B\}, \quad (4.2)$$

onde T coeficiente de transmissão na superfície da água, I_0 é a cor da luz na superfície e z é a profundidade até a superfície da água. Quando se lida com efeitos atmosféricos, como neblina ou névoa, c é normalmente assumido como sendo constante para todo o intervalo

visível. Contudo, em aplicações subaquáticas diferentes comprimentos de onda de luz λ , são absorvidos não linearmente e preferencialmente em relação a outros.

Além das constantes T e I_0 foi adicionado a equação um valor de l , onde l é uma constante parametrizada pela distância focal da lente da câmera. Assim, a equação utilizada para a estimativa dos parâmetros no simulador pode ser descrita por:

$$A(z) = lTI_0e^{-c^\lambda z}, \lambda \in \{R, G, B\}, \quad (4.3)$$

Na [Equação 4.3](#) assumiu-se que em uma lente com distância focal de $f = 20$ mm o l corresponde a $l = 1.06$ ([SCHECHNER; KARPEL, 2004](#)) e as constantes T tem um valor de 1.0 ([GOULD; ARNONE; SYDOR, 2001](#)) e I_0 é a luz branca.

Utilizando a [Equação 4.3](#) os parâmetros são estimados a partir de fragmentos extraídos de imagens degradadas capturadas em meios participativos com diferentes características. Esses fragmentos foram retirados de regiões da imagem onde o horizonte apresentava-se não ocluso. Esses fragmentos foram chamados no simulador de *turbid patches*.

A cada execução do simulador, um *turbid patch* é escolhido randomicamente dentre *patches* disponíveis. Cada imagem sem degradação recebe uma simulação diferente correspondente ao número de *turbid patches* disponíveis para uma melhor diversidade nas imagens geradas utilizando os parâmetros estimados.

Tendo em vista que o modelo físico de formação de imagem pode ser aplicado a qualquer meio participativo, o simulador TURBI pode ser utilizado para sintetizar diferentes tipos de degradação causados por diferentes meios participativos, somente modificando os *turbid patches* dados como entrada.

Para adquirir a profundidade da cena, foram utilizadas imagens limpas onde o mapa de profundidade estivesse disponível, como por exemplo os conjuntos de imagens disponibilizados pela Universidade de Nova Iorque (*New York University*) ([SILBERMAN; FERGUS, 2011](#)) ([SILBERMAN DEREK HOIEM; FERGUS, 2012](#)). Esses dois *datasets* chamados NYU-Depth Versão 1 e 2 são composto por uma sequência de vídeo de uma variedade de cenas internas gravadas por ambas as câmeras RGB e de profundidade do Microsoft Kinect V1. Outros conjuntos de imagens com o mapa de profundidade disponível podem ser encontrados na internet ([List of RGBD datasets, 2017](#)).

Com estes dados, é possível sintetizar a degradação desejada baseado na implementação do modelo de formação de imagem descrito na [Equação 4.1](#). O desenvolvimento foi realizado no *software* MATLAB ([MATLAB, 2010](#)). Suas entradas são uma imagem limpa, seu correspondente mapa de profundidade e o *turbid patch* para extração dos parâmetros. A saída do simulador é uma versão degradada da imagem de entrada em

função do seu mapa de profundidade, com a mesma resolução que a original. Com este simulador, o número de imagens simuladas é igual ao número de imagens de base com seus correspondentes mapas de profundidade multiplicadas pelo número de *turbid patches* utilizados .

4.2.2 Usando mapa de profundidade para sintetizar imagens turvas

Frente à relevância da criação de diferentes tipos de dados utilizando uma mesma plataforma somente com a modificação dos dados de entrada fornecidos, foram criados dois conjuntos de dados com diferentes tipos de degradação. Em um primeiro momento foram geradas imagens que apresentam degradação simulada com parâmetros estimados a partir de imagens turvas capturadas em ambientes subaquáticos.

Devido a falta de imagens de ambientes subaquáticos sem degradação e com seus respectivos mapas de profundidade, para a criação do *dataset* com imagens turvas foram utilizados como dados de entrada as imagens sem degradação com seus respectivos mapas de profundidades disponibilizadas pela segunda versão do *dataset* NYU-Depth (SILBERMAN DEREK HOIEM; FERGUS, 2012). Foram utilizadas 1449 imagens *indoor* sem degradação com seus respectivos mapas de profundidade. Como *turbid patches*, foram utilizados fragmentos retirados de 25 imagens diferentes capturadas em ambientes subaquáticos naturais com diferentes características, tais como: diferentes cor da água, diferentes condições de iluminação, diferentes intensidades de turbidez, etc. No total, esse *dataset* contém 36.225 imagens.

No segundo conjunto de dados foram geradas imagens que apresentam degradação simulada empregando os parâmetros estimados a partir de imagens capturadas em ambientes com névoa. Para a criação deste conjunto de dados, foi utilizado o mesmo *dataset*, NYU-Depth segunda versão, utilizado na geração do *dataset* de imagens turvas. Como *turbid patches*, foram utilizados fragmentos retirados de 5 imagens diferentes capturadas em ambientes *outdoor* com névoa com diferentes características em diferentes condições de iluminação.

No total, esse *dataset* contém 8.694 imagens.

Ambos *datasets* podem ser utilizados em diversas aplicações onde é necessário um grande número de imagens degradadas com seus respectivos *ground-truth*, como por exemplo no treinamento de redes neurais voltadas para a restauração de imagens.

Devido a grande quantidade de imagens contidas em ambos *datasets*, estes não estão disponíveis na internet. Porém o código do simulador assim com o conjunto de *Turbid Patches* utilizados em ambos *datasets* estão disponíveis em: https://github.com/mandacduarte/Underwater_Simulator.

5 Aplicações dos *Datasets* Propostos

Assim como a maioria dos *datasets* existentes, as aplicações na qual essas imagens podem ser utilizadas são diversas. Neste capítulo serão mostradas duas aplicações nas quais os *datasets* foram aplicados.

Na [seção 5.1](#) serão apresentadas as avaliações quantitativas de alguns dos métodos de restauração mais populares presentes no estado da arte quando aplicados a diferentes condições em um mesmo ambiente utilizando o conjunto de imagens *Milk* presente no *dataset* TURBID 3D.

Em seguida, na [seção 5.2](#) serão apresentado alguns exemplos de restauração de imagens capturadas em meios participativos utilizando uma rede neural convolucional treinada com um *dataset* gerado a partir do simulador TURBI, evidenciando a eficácia da abordagem utilizada para a sintetização de imagens turvas para o treinamento desse tipo de rede.

5.1 Avaliação de Métodos de Restauração de Imagens Aplicados à Ambientes Subaquáticos

Métodos de *enhancement* e restauração de imagens são propostos a fim de recuperar imagens que apresentam algum tipo de degradação. A maioria desses métodos são desenvolvidos para lidar com imagens que apresentam degradação causada pelas propriedades presentes em ambientes terrestres.

Entretanto, ambientes subaquáticos possuem propriedades particulares, como por exemplo diferentes partículas em suspensão, diferentes níveis de turbidez, diferentes cores do ambiente, diferentes condições de iluminação no mesmo ambiente, etc. Essas propriedades produzem cenas com uma degradação própria que normalmente imagens capturadas em ambientes terrestres não apresentam. Com essa grande diversidade de características sujeitas a variação, métodos de restauração de imagens quando aplicados em imagens capturadas nesse tipo de ambiente tem dificuldade em tratar todas essas variáveis.

Essas dificuldades refletem nas aplicações em tempo real, por exemplo, onde é necessário a utilização de métodos robustos capazes de lidar com diferentes condições em um mesmo ambiente, como é o caso da utilização em robôs em missões de navegação e exploração subaquática, dentre outros.

Porém, avaliar e comparar métodos de restauração quanto a sua robustez a diferentes condições em um mesmo ambiente quando aplicados em ambientes subaquáticos

não é uma tarefa fácil por diversas razões. Em primeiro lugar, ao contrário de outros ambientes, a imagem de referência, ou seja, a imagem sem degradação é de difícil acesso ou até mesmo impossível de se obter em ambientes naturais. Com isso, a avaliação objetiva da qualidade da restauração obtida pelos métodos é uma tarefa quase impossível. Isso se dá pelo fato de que a maioria das métricas de avaliação da qualidade de imagens leva em consideração a imagem de referência.

Como maneira alternativa, algumas métricas de avaliação subjetivas são comumente utilizadas. Porém, essas métricas possuem parâmetros não delimitados e definições ambíguas. Historicamente, esse tipo de avaliação pode levar a resultados quantitativos inconsistentes, prejudicando assim a avaliação de uma maneira geral. Por fim, a falta de um conjunto de avaliações pré-definidas torna difícil uma comparação justa entre os métodos.

Imagens com degradações simuladas são também utilizadas pelos desenvolvedores a fim de avaliar a restauração dos métodos desenvolvidos. Para isso, alguns tipos de degradação são simulados e aplicados em uma imagem não degradada utilizando técnicas computacionais de renderização. Entretanto, normalmente essas técnicas possuem um custo computacional muito alto e não são capazes de simular a complexidade do fenômeno de degradação causado em ambientes subaquáticos reais a fim de avaliar o quanto o método trata esse tipo de fenômeno.

Com a alternativa do acesso a imagem de referência e a imagens com diferentes condições em um mesmo ambiente subaquático, um conjunto de imagens presente no *dataset* TURBID 3D foi utilizado para a avaliação qualitativa de alguns dos métodos de restauração de imagens presentes no estado da arte.

Utilizando essa imagem de referência, é possível identificar o erro atual obtido pelos métodos quando aplicado a diferentes níveis de degradação, por exemplo, e com isso concluir sobre a eficiência deles, sabendo que o algoritmo está sendo avaliado em função da degradação da imagem.

5.1.1 Métodos Avaliados

Utilizando o conjunto de imagens *Milk* alguns dos métodos mais populares presentes no estado da arte além de novos métodos que apresentam diferentes paradigmas de restauração foram avaliados quanto a sua performance quando aplicados a imagens com diferentes níveis de turbidez.

Os métodos escolhidos foram:

- **métodos de *enhancement***: Equalização de histograma adaptativo - CLAHE (HUMMEL, 1977), Balanceamento de Branco - Shades of Gray (FINLAYSON; TREZZI, 2004), o método proposto por Ancuti et al. (2012);

- **métodos de restauração de imagens baseados no modelo físico de propagação da luz:** Dark Channel Prior - DCP (HE; SUN; TANG, 2011), Underwater Dark Channel Prior - UDCP Drews et al. (2016), Red Channel Prior - RCP (GALDRAN et al., 2015) e um método geral para restauração de imagens em meios participativos proposto por Gaya et al. (2016b) chamado aqui de Gaya.

Para restaurar as imagens degradadas utilizando os métodos *UDCP* e o método proposto por Codevilla et al, foram utilizados o código fonte providos pelos seus respectivos autores os métodos.

Os métodos **DCP**, **RCP** foram implementados em linguagem C++ em OpenCV (BRADSKI, 2000). Para obter os resultados dos métodos, Equalização de histograma adaptativo - **CLAHE**, o método proposto por Ancuti et al e o método de balanceamento de branco - **Shades of gray** foram utilizadas implementações feita em MATLAB.

Esses métodos foram implementados visto que não há disponibilidade dos códigos fontes originais providos pelos autores. Para a obtenção dos resultados mais semelhante possível aos originais, duas estratégias foram seguidas. Primeiro, alguns parâmetros foram modificados a fim dos resultados ficarem o mais próximo possível dos apresentados pelos autores em cada trabalho. Segundo, os métodos implementados foram aplicados nas imagens originais disponíveis por cada trabalho, afim de obter a restauração semelhante ao que foi apresentado por cada autor. Afim de promover uma comparação justa, foi evitado a estimativa da constante luz global do ambiente para a restauração, setando um valor fixo.

5.1.2 Avaliação Quantitativa

Para a avaliação quantitativa dos métodos de restauração escolhidos foram utilizadas sete métricas FR-IQ, dentre elas estão: **MSE** e **PSNR**, o **SSIM** (WANG et al., 2004), o **MSSSIM** (WANG; SIMONCELLI; BOVIK, 2003), o **FSIM** (ZHANG et al., 2011), e o **SRSIM** (ZHANG; LI, 2012).

O critério de escolha dessas métricas foi sua larga utilização na literatura e relatos de bom desempenho por pesquisadores. Além disso, as métricas utilizadas possuem seus respectivos códigos fonte disponibilizados.

5.1.3 Avaliação da Qualidade das Imagens Restauradas

Quando se trata de imagens capturadas em ambientes subaquáticos, a qualidade de uma imagem turva por ser entendida como a visibilidade da cena fotografada, também chamada de clareza visual da cena (GARCIA; GRACIAS, 2011).

A fim de avaliar qualitativamente a restauração obtida por cada um dos métodos, foram comparadas com a imagem de referência as imagens restauradas correspondentes aos 19 níveis de turbidez e a imagem sem turbidez utilizando as métricas de avaliação selecionadas. Essa avaliação foi realizada também no conjunto das imagens originais, ou seja, sem restauração, a fim de comparação.

Nos gráficos apresentados nas figuras a seguir cada índice apresenta no eixo y o valor correspondente a comparação de acordo com a métrica utilizada entre a imagem restaurada e a imagem de referência em função do aumento da degradação entre as 19 imagens, apresentado no eixo x .

Em verde, vermelho, preto, ciano, rosa e amarelo e preto são plotados os valores correspondentes apresentado entre as imagens restauradas com os métodos Ancuti (ANCUTI et al., 2012), Clahe (HUMMEL, 1977), Shades of gray (FINLAYSON; TREZZI, 2004), DCP (HE; SUN; TANG, 2011), RCP (GALDRAN et al., 2015), Gaya (GAYA et al., 2016b) e UDCP (DREWS et al., 2016) respectivamente e a imagem de referência utilizando a métrica em questão.

Para a obtenção desses resultados foram utilizados três conjuntos de imagens gerados durante o presente trabalho. Foram utilizados o conjunto de imagens *Milk* e *Chlorophyll* nas suas duas variações, com vista frontal e vista lateral da cena.

5.1.3.1 Resultados

O resultado visual da restauração das imagens do *dataset Milk* obtida pelos métodos Ancuti (ANCUTI et al., 2012), Clahe (HUMMEL, 1977), RCP (GALDRAN et al., 2015), Shades of gray (FINLAYSON; TREZZI, 2004), DCP (HE; SUN; TANG, 2011), RCP (GALDRAN et al., 2015), Gaya (GAYA et al., 2016b) e UDCP (DREWS et al., 2016) podem ser vistos na Figura 14. A imagem apresentada a restauração obtida na imagem de referência e nas imagens *I3*, *I7*, *I11*, *I15*, *I18* respectivamente.

A imagem de referência foi levada em consideração a fim de avaliar como o método se comportaria sem a presença de turbidez na imagem. O esperado nessas situações é que o método não adicione ruído nas imagens quando essa não apresentar turbidez. Entretanto, isso é o que acontece em alguns métodos, como pode ser observado na Figura 14.

A avaliação quantitativa da restauração de ambos *conjuntos de imagens* obtidas pelos métodos foi dada pelas métricas FR-IQ MSE e PSNR, o SSIM (WANG et al., 2004), o FSIM (ZHANG et al., 2011), o SRSIM (ZHANG; LI, 2012) e o MSSSIM (WANG; SIMONCELLI; BOVIK, 2003) respectivamente.

Em azul é apresentado o erro entre as imagens do conjunto original do *dataset*, ou seja, as imagens sem restauração a fim de comparação.

Nas Figuras 15, 16, 17, 18, 19, 20 estão os resultados obtidos conforme a métrica

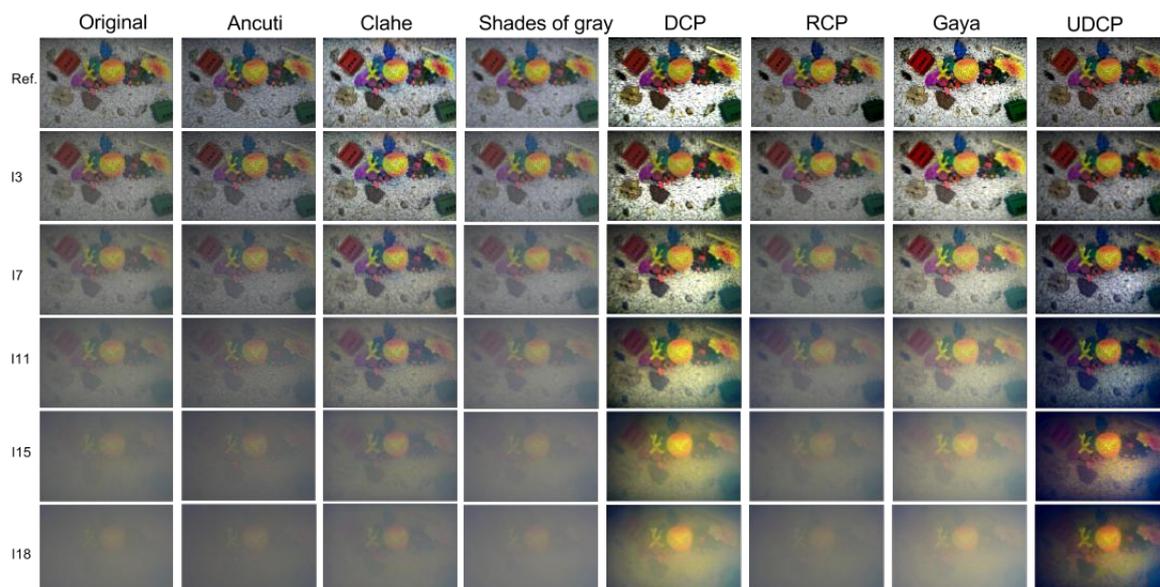


Figura 14 – Resultado visual da restauração da imagem de referência e das imagens *I3*, *I7*, *I11*, *I15*, *I18* do conjunto de imagens *Milk* utilizando os métodos - Ancuti (ANCUTI et al., 2012), Clahe (HUMMEL, 1977), Shades of gray (FINLAYSON; TREZZI, 2004), DCP (HE; SUN; TANG, 2011), RCP (GALDRAN et al., 2015), Gaya (GAYA et al., 2016b), UDCP (DREWS et al., 2016) e respectivamente.

utilizada para avaliação referentes a aplicação dos métodos no conjunto de imagens *Milk*.

A Figura 15 apresenta o gráfico obtido através da comparação dos resultados da restauração obtidos pelos métodos utilizando o erro quadrático médio - MSE entre as imagens restauradas pelos métodos e a imagem de referência. .

Em azul é apresentado o erro entre as imagens do conjunto original do *dataset*, ou seja, as imagens sem restauração a fim de comparação. É importante notar que quanto mais o erro estiver abaixo da linha das imagens não restauradas (representada pela linha azul) significa que o método realizou uma restauração mais eficaz, visto que a imagem em questão tendeu a ficar mais próxima da imagem de referência (ref ou I_0), ou seja, com menor degradação.

A Figura 16 apresenta os valores obtidos através da comparação dos resultados da restauração obtidos pelos métodos utilizando a relação sinal-ruído de pico - PSNR.

A Figura 17 apresenta os valores obtidos através da comparação dos resultados da restauração obtidos pelos métodos utilizando índice de similaridade estrutural - SSIM.

A Figura 18 apresenta os valores obtidos através da comparação dos resultados da restauração obtidos pelos métodos utilizando o índice de similaridade de características - FSIM.

A Figura 19 apresenta os valores obtidos através da comparação dos resultados da restauração obtidos pelos métodos utilizando o índice de similaridade SR-SIM - SRSIM.

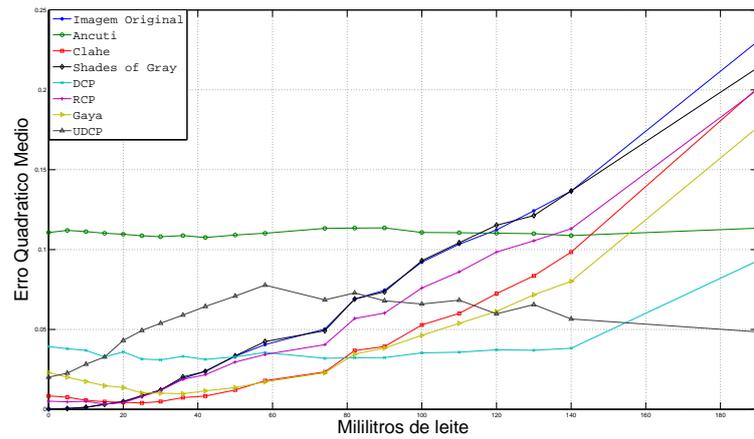


Figura 15 – O gráfico apresenta os valores correspondentes ao erro quadrático médio entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

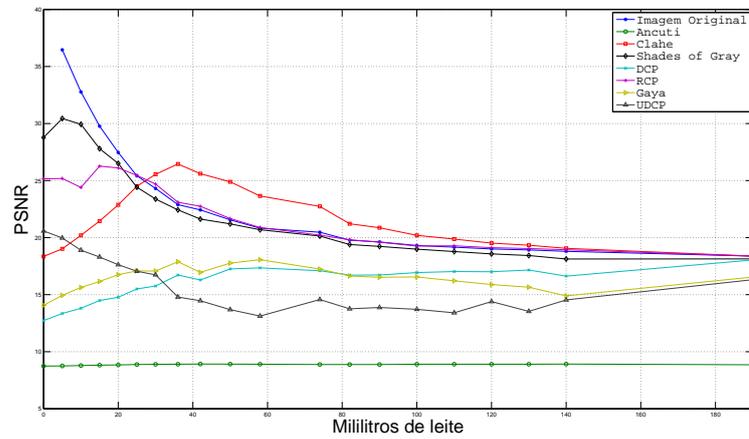


Figura 16 – O gráfico apresenta os valores correspondentes a relação sinal-ruído de pico entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

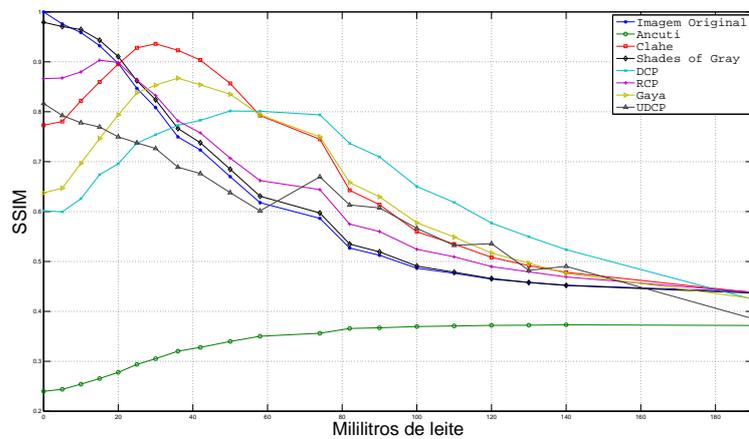


Figura 17 – O gráfico apresenta os valores correspondentes ao índice de similaridade estrutural entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

A Figura 20 apresenta os valores obtidos através da comparação dos resultados da restauração obtidos pelos métodos utilizando o índice de similaridade estrutural em

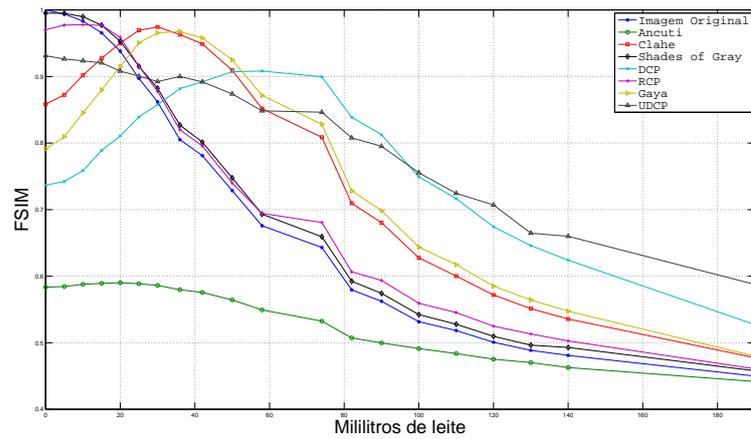


Figura 18 – O gráfico apresenta os valores correspondentes ao índice de similaridade de características entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência.

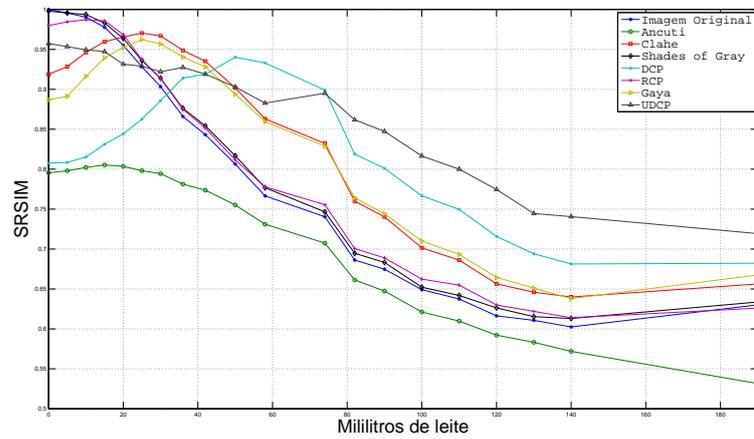


Figura 19 – O gráfico apresenta os valores correspondentes ao índice de similaridade SR-SIM entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

várias escalas - [MSSSIM](#).

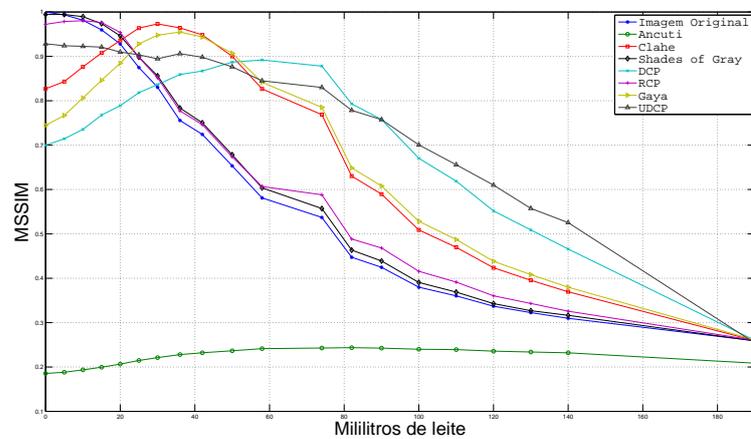


Figura 20 – O gráfico apresenta os valores correspondentes ao índice de similaridade estrutural em várias escalas entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

Nas 21, 22, 23, 24, 25, 26 serão apresentados os resultados obtidos conforme a métrica utilizada para avaliação referentes a aplicação dos métodos no conjunto de imagens *Chlorophyll* com vista frontal.

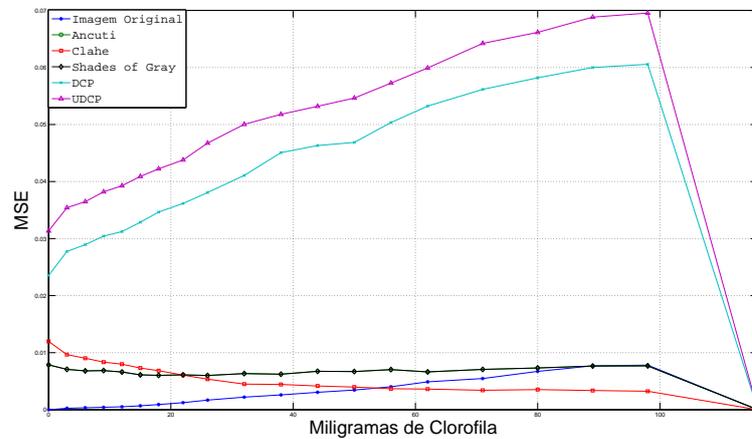


Figura 21 – O gráfico apresenta os valores correspondentes ao erro quadrático médio entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

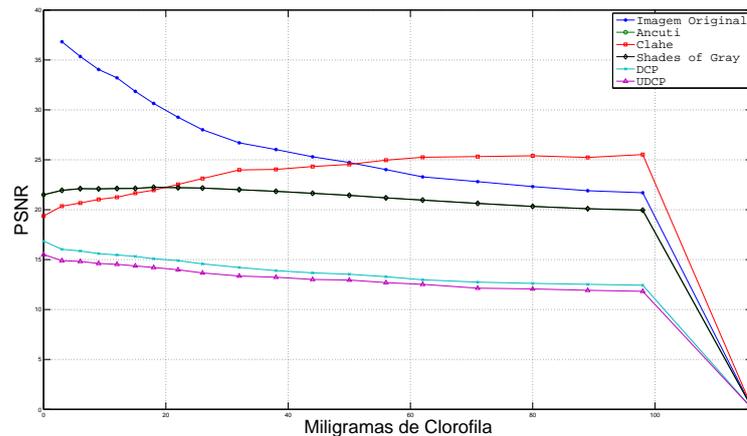


Figura 22 – O gráfico apresenta os valores correspondentes a relação sinal-ruído de pico entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

Nas 27, 28, 29, 30, 31, 32 serão apresentados os resultados obtidos conforme a métrica utilizada para avaliação referentes a aplicação dos métodos no conjunto de imagens *Chlorophyll* com vista lateral.

Com essas avaliações, pode ser visto que há uma diferença clara entre os métodos de restauração específicos e os métodos gerais de *enhancement*. Todos métodos de restauração DCP, RCP e [Gaya et al. \(2016b\)](#) são baseados em *priors*. Com esse comportamento é possível observar que a estimativa dessas *priors* necessitam de um certo nível de turbidez para serem estimadas corretamente.

A maioria dos métodos de restauração são associados com a estimativa de parâmetros, considerada uma tarefa complicada. Em níveis baixos de turbidez, quando esses parâmetros

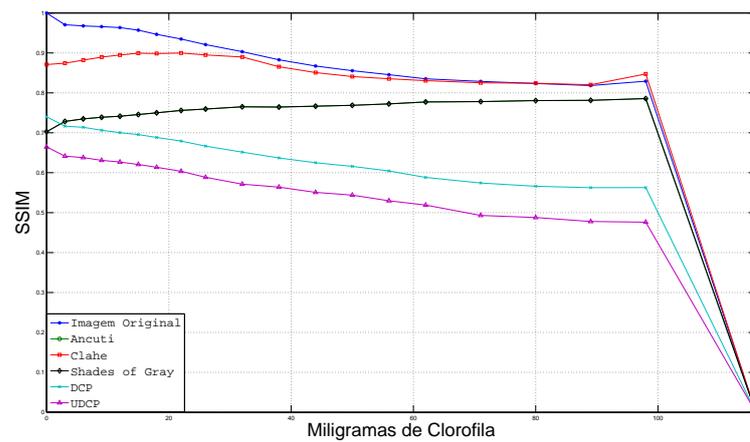


Figura 23 – O gráfico apresenta os valores correspondentes ao índice de similaridade estrutural entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

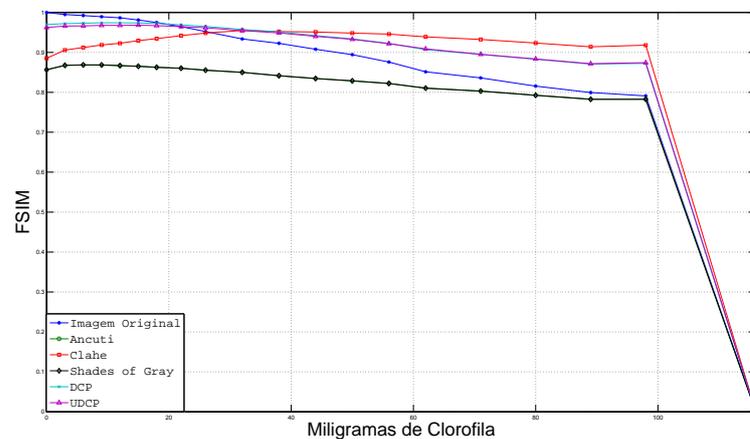


Figura 24 – O gráfico apresenta os valores correspondentes ao índice de similaridade de características entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência.

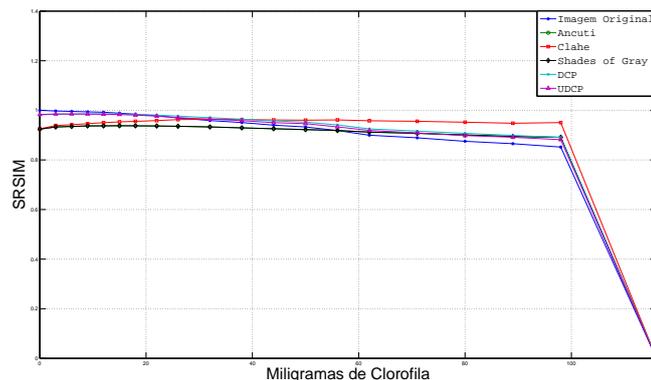


Figura 25 – O gráfico apresenta os valores correspondentes ao índice de similaridade SR-SIM entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

não são estimados corretamente, esses métodos tendem a incluir informações não pertencentes ao sinal original da imagem. Por outro lado, em níveis altos de turbidez, quando a visibilidade é baixa e as *priors* são estimadas corretamente, eles apresentam um compor-

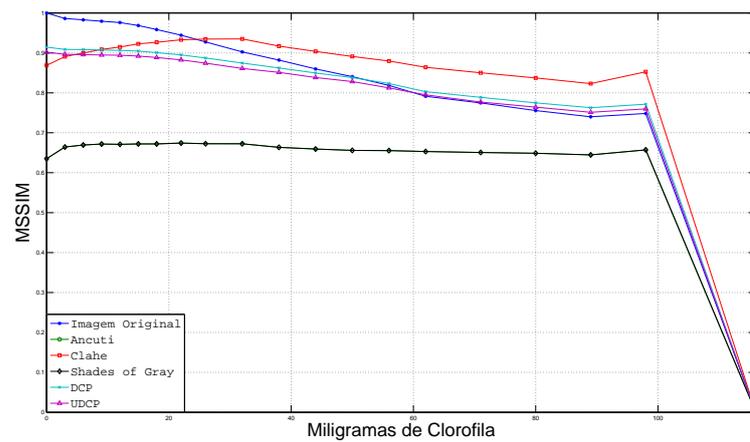


Figura 26 – O gráfico apresenta os valores correspondentes ao índice de similaridade estrutural em várias escalas entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

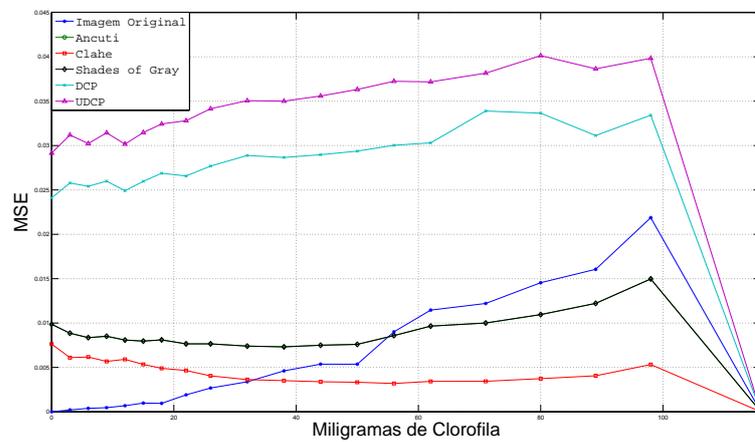


Figura 27 – O gráfico apresenta os valores correspondentes ao erro quadrático médio entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

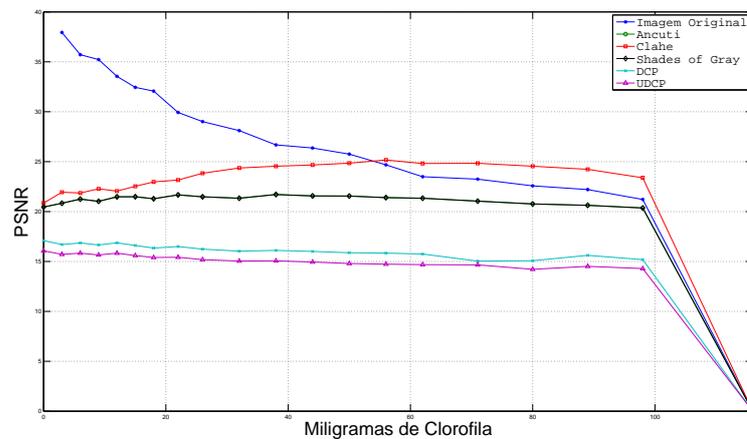


Figura 28 – O gráfico apresenta os valores correspondentes a relação sinal-ruído de pico entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

tamento mais eficaz. Entretanto, o [Ancuti et al. \(2012\)](#) não é baseado em *priors*, porém os resultados mostram que esse método também precisa de um certo nível de turbidez

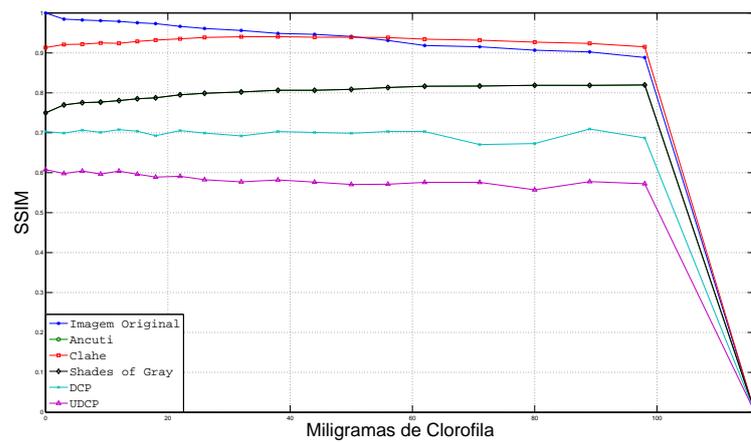


Figura 29 – O gráfico apresenta os valores correspondentes ao índice de similaridade estrutural entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

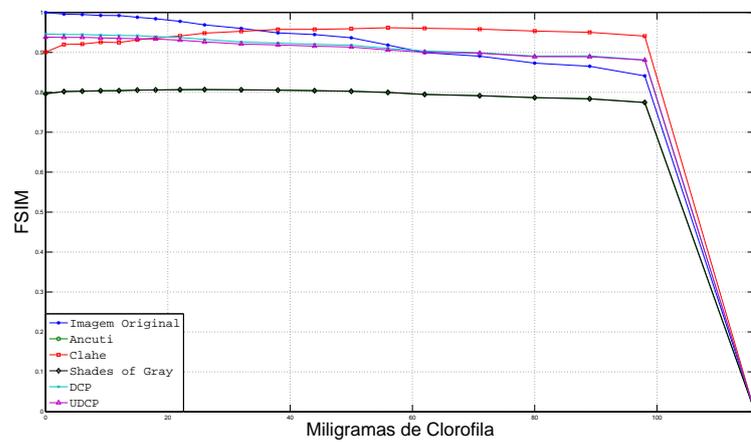


Figura 30 – O gráfico apresenta os valores correspondentes ao índice de similaridade de características entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência.

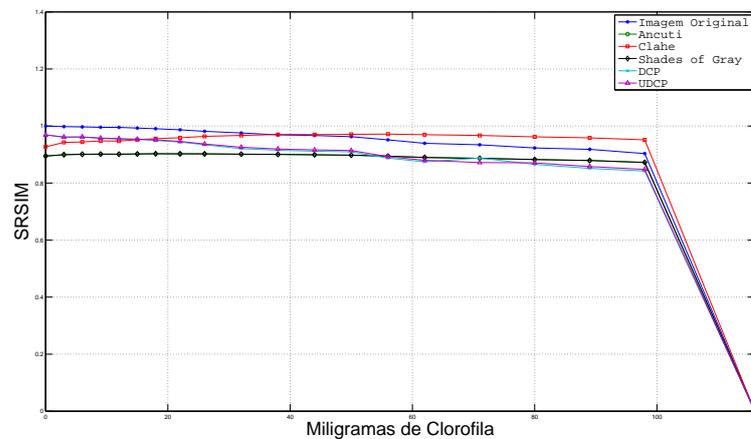


Figura 31 – O gráfico apresenta os valores correspondentes ao índice de similaridade SR-SIM entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

para mensurar os pesos dos presentes do método.

Os métodos de balanceamento de branco consideram na sua grande maioria as

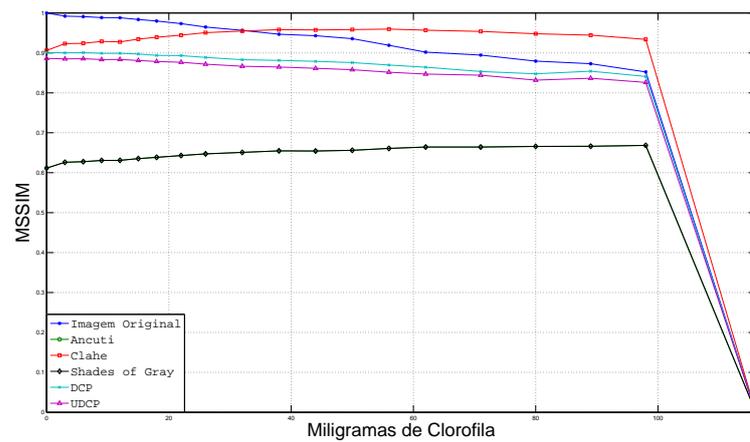


Figura 32 – O gráfico apresenta os valores correspondentes ao índice de similaridade estrutural em várias escalas entre as imagens restauradas utilizando os métodos de restauração e a imagem de referência

correções de luz na cena. Essa é uma boa solução para imagens capturadas em ambientes terrestres com estimativa de luz errada, porém em ambientes subaquáticos esse tipo de método não apresentou um desempenho suficiente na recuperação das propriedades originais da imagem sem degradação.

O método de equalização de histograma adaptativo CLAHE, apresentou-se como um método satisfatório para melhorar a visibilidade de imagens subaquáticas de uma maneira geral através do aumento da visibilidade das imagens. Um dos principais motivos para obtenção deste resultado se dá pelo fato de que esse método busca realçar os detalhes das imagens que tendem a serem perdidos com o aumento da turbidez. E diferentemente dos outros métodos utilizados, ele não adiciona informações não originais da cena em imagens pouco turvas.

5.2 Treinamento de Redes Neurais Convolucionais para Restauração de Imagens capturadas em Ambientes Participativos

Métodos de aprendizado de máquina, tais como redes neurais convolucionais são capazes, através de um processo chamado treinamento, de aprender automaticamente complexas relações de entrada-saída com base na observação de dados. Esses métodos podem ser capazes de aprender heurísticas mais complexas, que os humanos não conseguiam perceber. Essas heurísticas complexas podem levar a melhores resultados em uma gama mais ampla de situações.

Porém, um dos grandes problemas enfrentados por pesquisadores que desenvolvem métodos de aprendizado de máquina ou métodos de aprendizagem profunda, ou também chamados mais recentemente de métodos de *Deep Learning*, muitas vezes não são relacionados a redes neurais, um dos grandes problemas a ser resolvido é a obtenção dos dados certos no formato necessário (SKYMIND, 2016).

Aprendizagem profunda e aprendizado de máquina de forma mais geral, necessitam de um bom conjunto de treinamento para funcionar corretamente. Coletar construir um conjunto de treinamento - um corpo considerável de dados conhecidos - leva tempo e domínio de conhecimento específico de onde e como reunir informações relevantes. O conjunto de treinamento atua como a referência contra a qual as redes de aprendizagem profunda são treinadas. O conjunto de treinamento apresenta os dados que essas redes supostamente devem aprender a reconstruir antes de serem apresentadas aos dados que não viram antes. Para criar um conjunto de treinamento útil, é necessário entender o problema que está se tentando resolver, ou seja, o que as redes de aprendizagem devem prestar atenção.

Em muitos casos, esse treinamento requer uma grande quantidade de dados rotulados, ou seja, entradas onde a *ground-truth* esteja disponível. No caso da restauração de imagens capturadas em meios participativos, este dado é composto por um par de imagens, ambas exatamente da mesma cena, capturadas sob as mesmas condições de luz, onde uma está corrompida pela degradação causada pelo ambiente e a outra é livre de degradação.

No sentido de verificar a eficácia da abordagem utilizada para gerar dados para o treinamento de uma rede neural, foi utilizado um modelo de rede neural convolucional desenvolvido pelo grupo de pesquisa da Universidade Federal do Rio Grande, porém ainda não publicado, com o intuito de restaurar imagens capturadas em ambientes com névoa.

Um conjunto de dados ideal para o treinamento desse modelo seria composto por imagens ao ar livre com correspondentes mapas de profundidade de alta qualidade. No entanto, atualmente não encontra-se disponível um conjunto de dados com essas carac-

terísticas. Todos os conjuntos de dados com mapas de profundidades capturados ao ar livre atualmente disponíveis têm problemas que os tornam irrelevantes para a simulação de névoa utilizando o simulador TURBI, tais como mapas de profundidade de baixa resolução, alcance máximo limitado e bordas incompatíveis entre o mapa de profundidade e a imagem real. Com esses problemas, a simulação de névoa nessas imagens não seria de forma precisa, o que prejudicaria o treinamento da rede.

Dessa forma, foi utilizado o *dataset* com névoa simulada descrito na [subseção 4.2.2](#). Para a criação desse *dataset* foi utilizado o simulador TURBI com as imagens disponíveis pela versão 2 do *dataset* NYU-Depth (SILBERMAN DEREK HOIEM; FERGUS, 2012) e 5 *Turbid Patches* retirados de imagens com névoa capturadas em diferentes condições.

Cada imagem gerada foi segmentada em quatro partes iguais de 224×224 . Desse modo o conjunto de dados utilizados foi de aproximadamente 29.000 pares de imagens sendo 80 % delas utilizadas para o treinamento, 10% para validação e 10% para teste.

As figuras 33 e 34 apresentam os resultados da restauração obtida pela rede treinada com o conjunto de imagens gerado pelo simulador TURBI.



Figura 33 – Resultado da restauração feita por uma rede CNN de restauração de imagens com névoa treinada com dados simulados gerados pelo simulador TURBI.

A rede utilizada encontra-se em fase de testes e ajustes de parâmetros para ser finalizada. Porém, com os resultados apresentados pode-se concluir que a abordagem utilizada no desenvolvimento do simulador TURBI foi eficaz para a geração uma grande quantidade de dados simulados para o treinamento da rede.

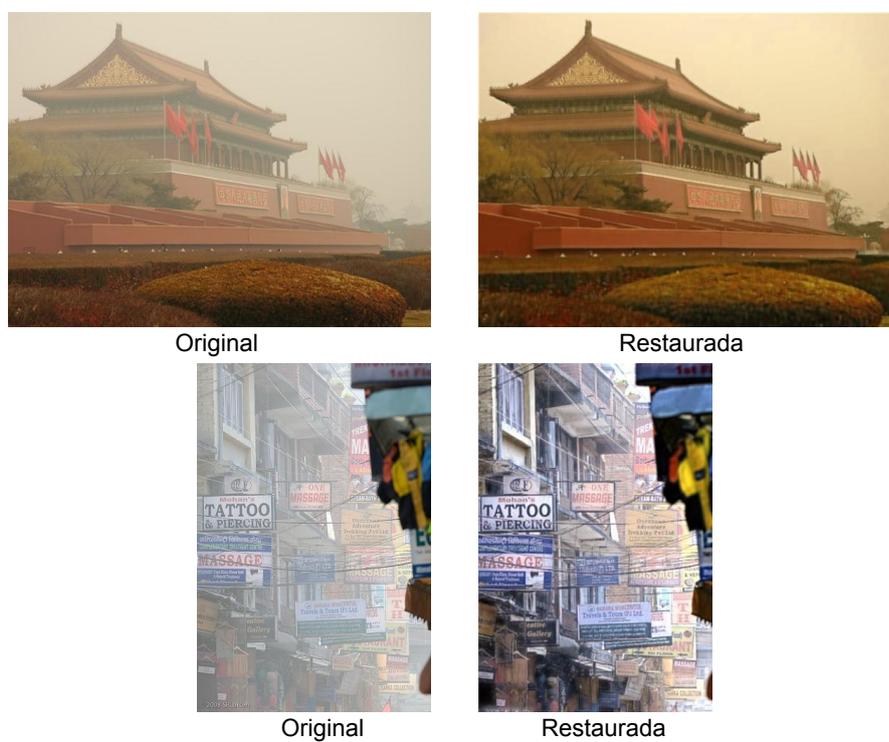


Figura 34 – Resultado da restauração feita por uma rede CNN de restauração de imagens com névoa treinada com dados simulados gerados pelo simulador TURBI.

6 Considerações Finais

A dissertação apresentada se desenvolveu em torno da criação de conjuntos de dados para a utilização em métodos de visão computacional e *machine learning* quando aplicados à imagens capturadas em ambientes subaquáticos. De maneira mais precisa o trabalho teve como objetivo a criação de dois tipos de conjuntos de dados: (i) imagens geradas a partir de um ambiente controlado onde as condições do meio e a degradação presente em cada imagem pudessem ser acessada e monitoradas, (ii) imagens geradas por meio da síntese da degradação causada por ambientes participativos a fim de criar conjuntos de dados simulados de larga escala.

Neste contexto foi proposta uma metodologia para a reprodução de ambientes subaquático. Estes ambientes contam com a presença de características encontradas em cenas naturais onde as condições do meio pudessem ser controladas. A partir dessa metodologia foi viável a criação de diferentes conjuntos de dados contendo imagens degradadas onde o acesso a imagem de referência, ou seja, a mesma imagem sem degradação, assim como a produção de diferentes tipos e intensidades de degradação foi possível.

Utilizando a estrutura proposta, foi gerado um *dataset* chamado de *TURBID 3D* o qual abrange cinco conjuntos de imagens com características distintas chamados neste trabalho de *Milk*, *GreenTea*, *Chlorophyll*, *DeepBlue*, *Blue*. Estes conjuntos de imagens formam gerados a partir da adição de diferentes substâncias na água para a produção do efeito *Backscattering*.

As imagens produzidas neste ambiente, podem ser utilizadas em diferentes aplicações dentro da visão computacional. Neste trabalho foi apresentada a avaliação de métodos de restauração de imagens presentes no estado da arte quando aplicados à imagens subaquáticas em função do aumento da degradação das imagens. Para este fim, foi empregado o uso de diferentes métricas de avaliação objetiva da qualidade das imagens restauradas. Essa avaliação foi possível pela disponibilidade da imagem de referência adquirida através da metodologia apresentada neste trabalho. Imagens capturadas em ambientes subaquáticos naturais, na maioria das vezes, não apresentam a sua respectiva imagem de referência.

Visto como uma das contribuições do trabalho apresentado, todas as imagens geradas neste ambiente, bem como as avaliações feitas nos métodos de restauração de imagens utilizando diferentes métricas podem ser acessadas a partir do endereço: www.amandaduarte.com.br/turbid.

Ao que se refere a segunda contribuição, foi criado um simulador, chamado de **TURBI**, capaz de simular em imagens não degradadas a degradação causada por meios

participativos a partir de fragmentos de uma imagem capturada em um ambiente natural.

A partir do desenvolvimento do simulador TURBI, foi possível a criação de *datasets* sintéticos de larga escala contendo pares de imagens (imagens degradadas e suas respectivas imagens de referencia). A criação de *datasets* de larga escala vem sendo hoje tratado com um dos grande problemas para o avanço no desenvolvimento de novas abordagens utilizando métodos de aprendizagem, como *Deep Learning*, que necessitam uma grande quantidade de imagens para serem treinados.

Um dos *datasets* proposto foi utilizado para treinar uma rede neural convolucional desenvolvida para restauração de imagens capturadas em ambientes com névoa. Para o treinamento desta rede, foram utilizado 28.000 pares de imagens com degradação sintetizada a partir do simulador TURBI.

Mesmo com a rede em fase ajustes e consolidação, o presente trabalho apresentou resultados satisfatórios quanto a restauração de cenas externas com presença de névoa utilizando uma rede neural treinada com um *dataset* gerado pelo simulador TURBI. Com isso, entende-se que a abordagem adotada para o desenvolvimento do simulador é válida e satisfatória.

Devido a grande quantidade de imagens presentes nos *datasets* gerados a partir do simulador TURBI, não foi possível a disponibilização do completo conjunto de imagens. Porém o código-fonte do simulador pode ser acessado a partir do endereço: https://github.com/mandacduarte/Underwater_Simulator.

Durante o desenvolvimento deste trabalho destaca-se a publicação de dois artigos e a apresentação em forma de *poster* em um *workshop* em conferência internacional:

- (DUARTE et al., 2016a) **A dataset to evaluate underwater image restoration methods**. OCEANS 2016-Shanghai, 2016.

- (GAYA et al., 2016a) **Vision-based Obstacle Avoidance Using DeepLearning**. 13th Latin-America Robotics Symposium, 2016.

- (DUARTE et al., 2016b) **TURBID: An Underwater Turbid Image Dataset**. Presented in European Conference on Computer Vision (ECCV) Workshop on Datasets and Analysis Performance in Early Vision, 2016.

Referências

- AHN, L. V.; DABBISH, L. Labeling images with a computer game. In: ACM. *Proceedings of the SIGCHI conference on Human factors in computing systems*. [S.l.], 2004. p. 319–326. Citado na página [32](#).
- ANCUTI, C. et al. Enhancing underwater images and videos by fusion. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012. Citado 5 vezes nas páginas [25](#), [56](#), [58](#), [59](#) e [64](#).
- AULINAS, J. et al. Vision-based underwater slam for the sparus auv. In: *Proceedings of the 10th International Conference on Computer and IT Applications in the Maritime Industries, Berlin, Germany*. [S.l.: s.n.], 2011. p. 2–4. Citado na página [14](#).
- BAKER, S. et al. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, Springer, v. 92, n. 1, p. 1–31, 2011. Citado na página [34](#).
- BARROS, W. Uma abordagem automática para restauração de imagens de cenas subaquáticas. Tese de Doutorado–UFMG. Belo Horizonte-MG, 2010. Citado na página [23](#).
- BAZEILLE S., Q. I. J. L.; MALKASSE, J. P. Automatic underwater image pre-processing. *Proceedings of the Characterisation du Milieu Marin*, 2006. Citado na página [25](#).
- BIANCO, N. C.; MOHAN, A.; EUSTICE, R. M. Initial results in underwater single image dehazing. In: IEEE. *OCEANS 2010*. [S.l.], 2010. p. 1–8. Citado na página [25](#).
- BOTELHO, S. S. d. C. et al. Appearance-based odometry and mapping with feature descriptors for underwater robots. *Journal of the Brazilian Computer Society*, SciELO Brasil, v. 15, n. 3, p. 47–54, 2009. Citado na página [14](#).
- BRADSKI, G. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000. Citado na página [57](#).
- BRANDOU, V. et al. 3d reconstruction of natural underwater scenes using the stereovision system iris. In: *OCEANS 2007 - Europe*. [S.l.: s.n.], 2007. p. 1–6. Citado na página [14](#).
- CAI, B. et al. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, v. 25, n. 11, p. 5187–5198, Nov 2016. ISSN 1057-7149. Citado na página [26](#).
- CARR, P.; HARTLEY, R. Improved single image dehazing using geometry. In: IEEE. *Digital Image Computing: Techniques and Applications, 2009. DICTA'09*. [S.l.], 2009. p. 103–110. Citado na página [23](#).
- CENTENO, M. *Rovfurg-ii: Projeto e construção de um veículo subaquático não tripulado de baixo custo*. Dissertação (Mestrado), 2007. Citado na página [13](#).

- CHAMBAH D. SEMANI, A. R. P. C. A. R. M. Underwater color constancy: Enhancement of automatic live fish recognition. *16th Annual symposium on electronic imaging*, IEEE, p. 157–168, 2004. Citado na página 24.
- CHIANG, J. Y.; CHEN, Y.-C. Underwater image enhancement by wavelength compensation and dehazing. *Image Processing, IEEE Transactions on*, IEEE, v. 21, n. 4, p. 1756–1769, 2012. Citado na página 25.
- CODEVILLA, F. et al. Geostatistics for context-aware image classification. In: SPRINGER. *International Conference on Computer Vision Systems*. [S.l.], 2015. p. 228–239. Citado na página 14.
- CODEVILLA, F. et al. Achieving turbidity robustness on underwater images local feature detection. *BMVC 2015*, v. 26th, p. 157, 2015. Citado 4 vezes nas páginas 37, 38, 44 e 50.
- Computer Vision Datasets. *Computer Vision Datasets*. 2017. Disponível em: <http://clickdamage.com/sourcecode/cv_datasets.php>. Citado na página 31.
- Computer Vision online. *Computer Vision online*. 2017. Disponível em: <<http://www.computervisiononline.com/datasets>>. Citado na página 31.
- CONCHA, A. et al. Real-time localization and dense mapping in underwater environments from a monocular sequence. In: IEEE. *OCEANS 2015-Genova*. [S.l.], 2015. p. 1–5. Citado na página 14.
- CV Datasets on the web. *CV Datasets on the web*. 2017. Disponível em: <<http://www.cvpapers.com/datasets.html>>. Citado na página 31.
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. 2009. Citado na página 33.
- DREWS-JR, P. L. J. *Contribuições à Restauração Automática de Imagens de Cenas em Meio Participativo*. Tese (Doutorado) — Universidade Federal de Minas Gerais — Departamento de Ciência da Computação, 2016. Citado na página 20.
- DREWS, P. L. et al. Underwater depth estimation and image restoration based on single images. *IEEE Computer Graphics and Applications*, IEEE, v. 36, n. 2, p. 24–35, 2016. Citado 4 vezes nas páginas 25, 57, 58 e 59.
- DRUGS.COM. *Bleu de méthylène*. 2017. Disponível em: <<https://www.drugs.com/international/bleu-de-méthylène.html>>. Citado na página 47.
- DUARTE, A. et al. A dataset to evaluate underwater image restoration methods. In: IEEE. *OCEANS 2016-Shanghai*. [S.l.], 2016. p. 1–6. Citado 2 vezes nas páginas 16 e 71.
- DUARTE, A. C. et al. Turbid: An underwater turbid image dataset. Presented in European Conference on Computer Vision (ECCV) Workshop on Datasets and Analysis Performance in Early Vision. 2016. Citado 2 vezes nas páginas 16 e 71.
- DUNTLEY, S. Q. Light in the sea*. *JOSA*, Optical Society of America, v. 53, n. 2, p. 214–233, 1963. Citado 2 vezes nas páginas 14 e 52.

- EUSTICE, R. et al. Uwit: Underwater image toolbox for optical image processing and mosaicking in matlab. In: IEEE. *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*. [S.l.], 2002. p. 141–145. Citado na página 24.
- EUSTICE, R. M.; PIZARRO, O.; SINGH, H. Visually augmented navigation for autonomous underwater vehicles. *Oceanic Engineering, IEEE Journal of*, IEEE, v. 33, n. 2, p. 103–122, 2008. Citado na página 14.
- EVERINGHAM, M. et al. The pascal visual object classes (voc) challenge. *International journal of computer vision*, Springer, v. 88, n. 2, p. 303–338, 2010. Citado na página 32.
- FATTAL, R. Single image dehazing. *ACM transactions on graphics (TOG)*, ACM, v. 27, n. 3, p. 72, 2008. Citado 2 vezes nas páginas 23 e 26.
- FATTAL, R. Dehazing using color-lines. *ACM Transactions on Graphics (TOG)*, ACM, v. 34, n. 1, p. 13, 2014. Citado na página 26.
- FEI-FEI, L.; FERGUS, R.; PERONA, P. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 28, n. 4, p. 594–611, 2006. Citado na página 31.
- FELLBAUM, C. *WordNet*. [S.l.]: Wiley Online Library, 1998. Citado 2 vezes nas páginas 31 e 33.
- FINK, M.; ULLMAN, S. From aardvark to zorro: A benchmark for mammal image classification. *International Journal of Computer Vision*, Springer, v. 77, n. 1-3, p. 143–156, 2008. Citado na página 31.
- FINLAYSON, G. D.; TREZZI, E. Shades of gray and colour constancy. In: SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY. *Color and Imaging Conference*. [S.l.], 2004. v. 2004, n. 1, p. 37–41. Citado 4 vezes nas páginas 24, 56, 58 e 59.
- FIRMAN, M. Rgb-d datasets: Past, present and future. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2016. p. 19–31. Citado na página 36.
- FRITSCH, J.; KUEHNL, T.; GEIGER, A. A new performance measure and evaluation benchmark for road detection algorithms. In: *International Conference on Intelligent Transportation Systems (ITSC)*. [S.l.: s.n.], 2013. Citado na página 35.
- GALDRAN, A. et al. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, Elsevier, v. 26, p. 132–145, 2015. Citado 4 vezes nas páginas 25, 57, 58 e 59.
- GARCIA, R.; GRACIAS, N. Detection of interest points in turbid underwater images. In: IEEE. *OCEANS, 2011 IEEE-Spain*. [S.l.], 2011. p. 1–9. Citado 4 vezes nas páginas 36, 44, 50 e 57.
- GARCIA, R.; NICOSEVICI, T.; CUFÍ, X. On the way to solve lighting problems in underwater imaging. In: IEEE. *OCEANS'02 MTS/IEEE*. [S.l.], 2002. v. 2, p. 1018–1024. Citado na página 24.

- GAYA, J. D. O. et al. Vision-based obstacle avoidance using deeplearning. In: IEEE. *13th Latin-America Robotics Symposium*. [S.l.], 2016. p. 1–6. Citado 3 vezes nas páginas 16, 26 e 71.
- GAYA, J. O. et al. Single image restoration for participating media based on prior fusion. *arXiv preprint arXiv:1603.01864*, 2016. Citado 5 vezes nas páginas 26, 57, 58, 59 e 62.
- GEIGER, A. et al. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013. Citado 2 vezes nas páginas 34 e 35.
- GIBSON, K. B.; VO, D. T.; NGUYEN, T. Q. An investigation of dehazing effects on image and video coding. *IEEE transactions on image processing*, IEEE, v. 21, n. 2, p. 662–673, 2012. Citado na página 23.
- GOULD, R. W.; ARNONE, R. A.; SYDOR, M. Absorption, scattering, and, remote-sensing reflectance relationships in coastal waters: Testing a new inversion algorithm. *Journal of Coastal Research*, Coastal Education Research Foundation, Inc., v. 17, n. 2, p. 328–341, 2001. ISSN 07490208, 15515036. Citado na página 53.
- GRIFFIN, G.; HOLUB, A.; PERONA, P. Caltech-256 object category dataset. California Institute of Technology, 2007. Citado na página 31.
- HAUTIERE, N.; TAREL, J.-P.; AUBERT, D. Towards fog-free in-vehicle vision systems through contrast restoration. In: IEEE. *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. [S.l.], 2007. p. 1–8. Citado na página 52.
- HE, K.; SUN, J.; TANG, X. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 33, n. 12, p. 2341–2353, 2011. Citado 5 vezes nas páginas 23, 25, 57, 58 e 59.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 51.
- HE, R. et al. Multiple scattering model based single image dehazing. In: IEEE. *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)*. [S.l.], 2013. p. 733–737. Citado na página 23.
- HENRIKSSON, L.; HYVÄRINEN, A.; VANNI, S. Representation of cross-frequency spatial phase relationships in human visual cortex. *Journal of Neuroscience*, Soc Neuroscience, v. 29, n. 45, p. 14342–14351, 2009. Citado na página 30.
- HUANG, G. B. et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*. [S.l.: s.n.], 2008. Citado na página 31.
- HULST, H. C.; HULST, H. C. van de. *Light scattering by small particles*. [S.l.]: Courier Corporation, 1957. Citado na página 36.
- HUMMEL, R. Image enhancement by histogram transformation. *Computer graphics and image processing*, Elsevier, v. 6, n. 2, p. 184–195, 1977. Citado 4 vezes nas páginas 24, 56, 58 e 59.

- JAFFE, J. S. Computer modeling and the design of optimal underwater imaging systems. *Oceanic Engineering, IEEE Journal of, IEEE*, v. 15, n. 2, p. 101–111, 1990. Citado 2 vezes nas páginas 20 e 22.
- KARPATHY, A.; FEI-FEI, L. Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2015. p. 3128–3137. Citado na página 51.
- KATSAGGELOS, A. K. *Digital image restoration*. [S.l.]: Springer Publishing Company, Incorporated, 2012. Citado na página 12.
- KIRK, J. T. *Light and photosynthesis in aquatic ecosystems, 3rd Edition*. [S.l.]: Cambridge University press, 1994. Citado na página 19.
- KOSCHMIEDER, H. Theorie der horizontalen sichtweite. *Beitr. zur Phys. d. freien Atm.*, p. 171–181, 1924. Citado 2 vezes nas páginas 23 e 52.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105. Citado na página 51.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado na página 13.
- List of RGBD datasets. *List of RGBD datasets*. 2017. Disponível em: <<http://www0.cs.ucl.ac.uk/staff/M.Firman/RGBDdatasets/>>. Citado 2 vezes nas páginas 31 e 53.
- LU, H. et al. Contrast enhancement for images in turbid water. *JOSA A*, Optical Society of America, v. 32, n. 5, p. 886–893, 2015. Citado na página 25.
- MACKINNEY, G. Absorption of light by chlorophyll solutions. *J. biol. Chem*, v. 140, n. 2, p. 315–322, 1941. Citado na página 45.
- MARR, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. [S.l.]: The MIT Press, 2010. Citado na página 30.
- MATLAB. *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010. Citado na página 53.
- MAVRIDAKI, E.; MEZARIS, V. No-reference blur assessment in natural images using fourier transform and spatial pyramids. In: *IEEE. Image Processing (ICIP), 2014 IEEE International Conference on*. [S.l.], 2014. p. 566–570. Citado na página 33.
- MCGLAMERY, B. A computer model for underwater camera systems. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Ocean Optics VI*. [S.l.], 1980. p. 221–231. Citado 2 vezes nas páginas 20 e 22.
- MENZE, M.; GEIGER, A. Object scene flow for autonomous vehicles. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. Citado 2 vezes nas páginas 34 e 35.
- MORRONE, M. C.; BURR, D. Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London B: Biological Sciences*, The Royal Society, v. 235, n. 1280, p. 221–245, 1988. Citado na página 30.

- NARASIMHAN, S. G. et al. Acquiring scattering properties of participating media by dilution. *ACM Transactions on Graphics (TOG)*, ACM, v. 25, n. 3, p. 1003–1012, 2006. Citado 2 vezes nas páginas 44 e 47.
- NARASIMHAN, S. G.; NAYAR, S. K. Vision and the atmosphere. *International Journal of Computer Vision*, Springer, v. 48, n. 3, p. 233–254, 2002. Citado 2 vezes nas páginas 23 e 52.
- NASCIMENTO, E.; CAMPOS, M.; BARROS, W. Stereo based structure recovery of underwater scenes from automatically restored images. In: IEEE. *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*. [S.l.], 2009. p. 330–337. Citado 4 vezes nas páginas 23, 37, 44 e 50.
- NEGAHDARIPOUR, S.; ZHANG, H.; HAN, X. Investigation of photometric stereo method for 3-d shape recovery from underwater imagery. In: IEEE. *OCEANS'02 MTS/IEEE*. [S.l.], 2002. v. 2, p. 1010–1017. Citado 2 vezes nas páginas 22 e 23.
- NOAA. *How much of the ocean have we explored?* 2014. Disponível em: <<http://http://oceanservice.noaa.gov/facts/exploration.html>>. Citado na página 13.
- OMAR, A. F. B.; MATJAFRI, M. Z. B. Turbidimeter design and analysis: a review on optical fiber sensors for the measurement of water turbidity. *Sensors*, Molecular Diversity Preservation International, v. 9, n. 10, p. 8311–8335, 2009. Citado na página 19.
- OMER, I.; WERMAN, M. Color lines: Image specific color representation. In: IEEE. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. [S.l.], 2004. v. 2, p. II–946. Citado na página 26.
- PHILLIPS, P. J. et al. The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, Elsevier, v. 16, n. 5, p. 295–306, 1998. Citado na página 31.
- QUEIROZ-NETO, J. P. et al. Underwater stereo. In: IEEE. *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*. [S.l.], 2004. p. 170–177. Citado na página 23.
- RASKAR, R.; AGRAWAL, A.; TUMBLIN, J. Coded exposure photography: motion deblurring using fluttered shutter. In: ACM. *ACM Transactions on Graphics (TOG)*. [S.l.], 2006. v. 25, n. 3, p. 795–804. Citado na página 33.
- REN, W. et al. Single image dehazing via multi-scale convolutional neural networks. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2016. p. 154–169. Citado na página 26.
- ROSER, M.; DUNBABIN, M.; GEIGER, A. Simultaneous underwater visibility assessment, enhancement and improved stereo. In: IEEE. *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. [S.l.], 2014. p. 3840–3847. Citado na página 52.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, v. 115, n. 3, p. 211–252, 2015. ISSN 1573-1405. Disponível em: <<http://dx.doi.org/10.1007/s11263-015-0816-y>>. Citado 2 vezes nas páginas 32 e 33.

- RUSSELL, B. C. et al. Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, Springer, v. 77, n. 1-3, p. 157–173, 2008. Citado na página 32.
- SAHU, P.; GUPTA, N.; SHARMA, N. A survey on underwater image enhancement techniques. *International Journal of Computer Applications*, Foundation of Computer Science, v. 87, n. 13, 2014. Citado na página 19.
- SCHECHNER, Y. Y.; KARPEL, N. Clear underwater vision. In: IEEE. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - CVPR 2004*. [S.l.], 2004. v. 1, p. 536–543. Citado 2 vezes nas páginas 52 e 53.
- SCHECHNER, Y. Y.; KARPEL, N. Recovery of underwater visibility and structure by polarization analysis. *Oceanic Engineering, IEEE Journal of*, IEEE, v. 30, n. 3, p. 570–587, 2005. Citado 2 vezes nas páginas 23 e 44.
- SCHECHNER, Y. Y.; NARASIMHAN, S. G.; NAYAR, S. K. Instant dehazing of images using polarization. In: IEEE. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. [S.l.], 2001. v. 1, p. I–I. Citado na página 52.
- SHI, J.; XU, L.; JIA, J. Discriminative blur detection features. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 2965–2972. Citado na página 33.
- SHIHAVUDDIN, A. et al. Automated classification and thematic mapping of bacterial mats in the north sea. In: IEEE. *OCEANS-Bergen, 2013 MTS/IEEE*. [S.l.], 2013. p. 1–8. Citado na página 14.
- SHOTTON, J. et al. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: SPRINGER. *European conference on computer vision*. [S.l.], 2006. p. 1–15. Citado na página 31.
- SILBERMAN DEREK HOIEM, P. K. N.; FERGUS, R. Indoor segmentation and support inference from rgb-d images. In: *ECCV*. [S.l.: s.n.], 2012. Citado 4 vezes nas páginas 34, 53, 54 e 68.
- SILBERMAN, N.; FERGUS, R. Indoor scene segmentation using a structured light sensor. In: *Proceedings of the International Conference on Computer Vision - Workshop on 3D Representation and Recognition*. [S.l.: s.n.], 2011. Citado 2 vezes nas páginas 34 e 53.
- SILVEIRA, L. et al. An open-source bio-inspired solution to underwater slam. *IFAC-PapersOnLine*, Elsevier, v. 48, n. 2, p. 212–217, 2015. Citado na página 14.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 51.
- SKYMIND. *DataSets and Machine Learning*. 2016. Disponível em: <<https://deeplearning4j.org/data-sets-ml>>. Citado na página 67.
- SMITH, R. C.; BAKER, K. S. Optical properties of the clearest natural waters (200–800 nm). *Applied optics*, Optical Society of America, v. 20, n. 2, p. 177–184, 1981. Citado na página 52.

- STOKES, M. D.; DEANE, G. B. Automated processing of coral reef benthic images. *Limnol. Oceanogr.: Methods*, v. 7, n. 157, p. 157–168, 2009. Citado na página 14.
- SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2015. p. 1–9. Citado na página 51.
- TAN, R. T. Visibility in bad weather from a single image. In: IEEE. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. [S.l.], 2008. p. 1–8. Citado na página 23.
- TAREL, J.-P.; HAUTIERE, N. Fast visibility restoration from a single color or gray level image. In: IEEE. *2009 IEEE 12th International Conference on Computer Vision*. [S.l.], 2009. p. 2201–2208. Citado na página 23.
- TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. *Sixth International Conference on Computer Vision, 1998.*, p. 839–846, 1998. Citado na página 24.
- TORRALBA, A.; FERGUS, R.; FREEMAN, W. T. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 30, n. 11, p. 1958–1970, 2008. Citado na página 31.
- TRUCCO, E.; OLMOS-ANTILLON, A. T. Self-tuning underwater image restoration. *IEEE Journal of Oceanic Engineering*, IEEE, v. 31, n. 2, p. 511–519, 2006. Citado 2 vezes nas páginas 22 e 23.
- WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, IEEE, v. 26, n. 1, p. 98–117, 2009. Citado na página 28.
- WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004. Citado 4 vezes nas páginas 28, 29, 57 e 58.
- WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multiscale structural similarity for image quality assessment. In: IEEE. *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*. [S.l.], 2003. v. 2, p. 1398–1402. Citado 4 vezes nas páginas 28, 29, 57 e 58.
- WEIJER, J. V. D.; GEVERS, T.; GIJSENIJ, A. Edge-based color constancy. *Image Processing, IEEE Transactions on*, IEEE, v. 16, n. 9, p. 2207–2214, 2007. Citado na página 24.
- WYNN, R. B. et al. Autonomous underwater vehicles (auvs): Their past, present and future contributions to the advancement of marine geoscience. *Marine Geology*, Elsevier, v. 352, p. 451–468, 2014. Citado na página 13.
- XIE, J.; XU, L.; CHEN, E. Image denoising and inpainting with deep neural networks. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2012. p. 341–349. Citado na página 13.

- XU, L. et al. Deep convolutional neural network for image deconvolution. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2014. p. 1790–1798. Citado na página 13.
- YAO, B.; YANG, X.; ZHU, S.-C. Introduction to a large-scale general purpose ground truth database: methodology, annotation tool and benchmarks. In: SPRINGER. *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. [S.l.], 2007. p. 169–183. Citado na página 32.
- ZHANG, J. et al. Local albedo-insensitive single image dehazing. *The Visual Computer*, Springer, v. 26, n. 6-8, p. 761–768, 2010. Citado na página 23.
- ZHANG, L.; LI, H. Sr-sim: A fast and high performance iqa index based on spectral residual. In: IEEE. *Image Processing (ICIP), 2012 19th IEEE International Conference on*. [S.l.], 2012. p. 1473–1476. Citado 4 vezes nas páginas 28, 30, 57 e 58.
- ZHANG, L. et al. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, IEEE, v. 20, n. 8, p. 2378–2386, 2011. Citado 4 vezes nas páginas 28, 30, 57 e 58.
- ZUIDERVELD, K. Graphics gems iv. In: HECKBERT, P. S. (Ed.). San Diego, CA, USA: Academic Press Professional, Inc., 1994. cap. Contrast Limited Adaptive Histogram Equalization, p. 474–485. ISBN 0-12-336155-9. Disponível em: <<http://dl.acm.org/citation.cfm?id=180895.180940>>. Citado na página 24.