



Felipe Codevilla Moraes

**Visão computacional em meio subaquático:
Um estudo sobre detecção de pontos de interesse
e classificação utilizando contexto.**

Brasil

5 de março, 2015

Felipe Codevilla Moraes

**Visão computacional em meio subaquático:
Um estudo sobre detecção de pontos de interesse
e classificação utilizando contexto.**

Universidade Federal de Rio Grande
Programa de Pós graduação em Computação
Mestrado em Engenharia de Computação

Orientador: Nelson Duarte Lopez Filho
Coorientador: Silvia Silva da Costa Botelho

Brasil
5 de março, 2015

Um estudo sobre detecção de pontos de interesse
e classificação utilizando contexto./ Felipe Codevilla Moraes. – Brasil, 5 de março,
2015-

Orientador: Nelson Duarte Lopez Filho
Felipe Codevilla Moraes Visão computacional em meio subaquático:
Dissertação de Mestrado – Universidade Federal de Rio Grande
Programa de Pós graduação em Computação
Mestrado em Engenharia de Computação, 5 de março, 2015.

1. Visão Computacional; 2. Visão Subaquática.

CDU 02:141:005.7

Felipe Codevilla Moraes

**Visão computacional em meio subaquático:
Um estudo sobre detecção de pontos de interesse
e classificação utilizando contexto.**

Nelson Duarte Lopez Filho
Orientador

Silvia Silva da Costa Botelho
Co-orientadora

Nuno Estrela Gracias
Convidado

Glauber Acunha Gonçalves
Convidado

Rafael Garcia
Convidado

Brasil
5 de março, 2015

Agradecimentos

Entre tantas dúvidas , resolvi apostar permanecer onde estava e, ao contrário do que se espera, não poderia ter acertado mais.

Primeiramente gostaria de agradecer a minha orientadora Silvia Botelho por todas as brilhantes ideias e oportunidades oferecidas, obviamente nada disso seria possível sem ela.

Agradeço a todos os colegas e amigos onde destaco Luan Silveira, Joel Gaya, Felipe Guth e, em especial, a ajuda fornecida pelo Pedro Ballester na última hora.

Agradeço a banca por comparecer e avaliar este trabalho.

I want to thank all the people from the ViCOROB Lab in Girona for receiving me so well during the 10 months I spent in Catalonia. Without this experience, the majority of this thesis would not be possible.

I also want to thank all my classmates and friends from Vibot (and Patryk) for all the fun and professional adventures we had together.

Agradeço o apoio financeiro da Agência Nacional do Petróleo, Gás Natural e Bio-combustíveis – ANP – , da Financiadora de Estudos e Projetos – FINEP – e do Ministério da Ciência e Tecnologia – MCT por meio do Programa de Recursos Humanos da ANP para o Setor Petróleo e Gás – PRH-ANP/MCT. Um agradecimento especial aos professores responsáveis pelo PRH - 27, Maria Isabel e Gilberto Griep.

Por fim, agradeço a minha família pelo apoio e carinho que sempre me deram, em especial ao meu pai, por sempre servir como um formidável exemplo ético em minha vida.

Resumo

A exploração e o monitoramento do bentos no ambiente marinho possuem importância econômica e ambiental crescente na sociedade atual. A qualidade da tecnologia de obtenção de imagens óticas subaquáticas tem melhorado consideravelmente devido ao advento dos *Remotely Operated Vehicles* (ROV) e dos *Autonomous Underwater Vehicles* (AUVs), o que tem possibilitado a coleta de milhares de dados visuais do fundo do oceano.

Técnicas de visão computacional, atualmente em franca utilização em ambientes terrestres, podem auxiliar a interpretação automática destas imagens, seja para minimizar o trabalho de identificação e monitoramento de feições e espécies, seja para fornecer subsídios a realização autônoma de missões.

Porém devido a presença do meio líquido, a propagação da luz no meio subaquático apresenta efeitos fotométricos que causam degradação na imagem, emergindo diversas questões a serem tratadas na classificação de imagens subaquáticas, as quais não estão presentes em outros ambientes.

Assim, o objetivo geral deste trabalho é estudar técnicas de visão computacional, e sua sensibilidade a presença do meio líquido. De forma mais precisa, duas técnicas de visão computacional são principalmente tratadas: a detecção de pontos de interesse e a adição das informações de contexto para classificação de objetos em ambientes *subsea*.

São aplicados e analisados diferentes algoritmos de detecção de pontos de interesse frente a imagens com diferentes níveis de turbidez. Um novo *dataset* foi proposto capaz de fornecer cenários com diferentes níveis de turbidez e objetos em cena, permitindo o testes múltiplos dos detectores mais usados na literatura e seu comportamento frente os fenômenos de degradação causados na imagem no meio subaquático. Foi encontrado que o algoritmo *DoG* se mostrou como uma melhor alternativa para resolver tal problema de forma invariante a escala.

Também foi estudada a questão da adição de contexto como forma de melhorar a taxa de acerto da classificação de imagens subaquáticas. Foi proposto um novo método para incluir contexto na classificação baseado em Geostatística e comparou-se com outras formas tradicionais de adição de contexto como os *Conditional Random Fields* (CRF).

Palavras-chaves: Visão Computacional, Geostatística, Visão Subaquática.

Abstract

The exploration and monitoring of the benthic sea zone has an important economic and environmental role in the nowadays society. The quality of the optical image acquiring technologies has become considerably better. This happened mainly due to the advent of the Remotely Operated Vehicles (ROV) and the Autonomous Underwater Vehicles (AUVs) and has opened the possibility to collect thousands of visual data from the seabed environment.

Computer vision techniques are today being largely used in over-land environments and can help the autonomous interpretation of images. These techniques can help to minimize the work of identifying and monitoring species and objects. Either having vision as a data acquiring source or to assist the automation of the operations.

However, due to the presence of the liquid media, the light propagation in underwater environments has photometric effects that cause degradation of the image. This degradation develops a lot of issues to be treated on underwater images that do not exist in other environments.

Thus, the objective of this work is to study computer vision techniques considering their sensibility to phenomenas of the underwater environments media. More precisely, mainly two computer vision techniques are considered: feature point detection and the adition context information for image classification, both on underwater images.

Different algorithms for feature point detection are applied for feature point detection under different turbidity levels. We provide a new dataset capable of providing different scenarios with different levels of turbidity. This dataset allowed the test of multiple feature detectors regarding their behavior with respect to the degradation effects of water turbidity. We found that, in this scenario, the DoG algorithm is the best alternative to solve scale invariant feature detection problems.

Finally, we studied the issue concerning the addition of context as a way to improve the accuracy of underwater image classification. We proposed a new method to include the context information on classification that is based on Geostatistics. This method was compared with an other traditional form of context addition that is the Conditional Random Fields (CRF).

Key-words: Computer Vision, Geostatistics, Underwater Vision.

Lista de ilustrações

Figura 1 – Comportamento da aplicação dos <i>kernels Hessian</i> e <i>Harris</i> para uma imagem teste (1a). (1b) mostra a saída da medida de <i>Harris</i> (Eq. 1.4). (1c) mostra a saída do determinante da matriz <i>Hessian</i> (Eq. 1.5) para a imagem teste. Tanto o <i>Hessian</i> como o <i>Harris</i> tem como saída as regiões de alta curvatura (Figura por Sojka (2003)).	31
Figura 2 – O processo para geração do espaço de escala pelo <i>DoG</i> . Ao invés de computar o <i>Laplacian</i> para cada escala, o mesmo é estimado pela diferença entre escalas consecutivas. Figura adaptada de (LOWE, 2004).	34
Figura 3 – Exemplo de um filtro caixa de tamanho 9x9 aplicado para geração de um espaço de escala equivalente a $\sigma = 1.2$. Outros espaços podem ser gerados usando caixas maiores.	34
Figura 4 – Alguns tipos de filtros utilizados para geração do espaço de escala pelo <i>CenSurE</i> . O filtro estrela, o filtro hexagonal e o filtro por diferença de caixas.	35
Figura 5 – Espaços de escala gerados. Primeira coluna mostra o espaço Gaussiano. Segunda coluna mostra o filtro média de caixas usado pelo <i>FastHessian</i> . Terceira coluna mostra um filtro poligonal estrelar de seis lados. A quarta Coluna mostra o espaço de escala anisotrópico usado pelo <i>KAZE</i>	37
Figura 6 – Janela considerada para a classificação usando contexto. No caso da integração de contexto diretamente nos classificadores (Fig. 6a), não são considerada as relações entre a vizinhança com si própria (Fig. 6b). Ou seja, se existem propriedades correlacionadas na vizinhança.	42
Figura 7 – A representação gráfica de um modelo CRF. Os quadrados em vermelho ($\varphi_i^u(x_i, \theta_u)$) são os fatores unitários calculados com o resultado dado pelo classificador. Os quadrados em azul são os fatores locais computados em cada aresta e utilizados para introduzir informação contextual. Os circulos verdes representam os superpixeis sendo classificados.	44
Figura 8 – Tendências que existem para as classes estarem próximas umas das outras, quanto mais claro, maior é a tendência existente. Por exemplo, é possível perceber que a classe <i>B</i> é provável de aparecer perto de uma classe <i>C</i> mas não próxima de uma classe <i>E</i>	46

Figura 9 – Divisão especificada em dois níveis de classificação. O nível unário $P_u(L \theta_u)$ onde somente a informação do superpixel segmentado é utilizada, apresentado em verde. E o nível local $P_l(L W)$, onde um determinado contexto local é incluído na classificação, representado pelo círculo azul.	50
Figura 10 – Figura do separador linear obtido pelo treinamento do SVM. Dado os conjuntos de dados já rotulados (Azuis e Vermelhos), o SVM determina o separador de máxima margem. A saída numérica do SVM já é própria para se ter um certo grau de confiança do classificador.	51
Figura 11 – Gráfico mostrando a probabilidade de acerto em função da máxima confiança retornada pelo classificador para um conjunto de dados. Em vermelho tem-se a função C_{l_i} treinada a partir do conjunto de dados em azul.	53
Figura 12 – Histograma mostrando a distribuição de probabilidades de saída de um classificador. Para o caso, a segunda classe, é a que obteve maior probabilidade, porém existe uma certa incerteza com relação a primeira classe.	53
Figura 13 – Diagrama geral da adição de contexto local utilizando Geoestatística. Primeiramente é medida a variabilidade entre as classes no contexto espacial. Tanto diretamente através das frequências de transição na imagem (taxa de transição medida), quanto através da inferência de propriedades estatísticas vindas da imagem (taxa de transição modelada). Em seguida são calculados os vetores de transição. Na segunda parte os vetores são utilizados para gerar pesos para imagem. Com isso, utilizando os pesos, o sistema SIS computa a adição de contexto local para cada superpixel.	54
Figura 14 – Medida feita do número de transições que uma classe faz para cada outra para múltiplas distâncias. Foi utilizada um kernel móvel e foram contadas as transições desde o centro (ponto vermelho) para todas as direções (representado pelos quadrados)	56
Figura 15 – A transição de probabilidade modelada para um determinado <i>dataset</i> . O eixo y apresenta a distância em pixels. As linhas verdes mostram as proporções para cada classe. Pode-se observar uma certa tendência na classe <i>Urchin</i> em transitar para categoria de <i>background</i> . Ainda, percebe-se que a classe de <i>background</i> tem um grande comprimento médio, dado que sua taxa de decaimento é bastante alta.	58

Figura 16	– Exemplo de uma vizinhança sendo considerada para um superpixel (apontado em vermelho). Um raio r é considerado e N pontos são amostrados nessa vizinhança (em azul). Cada um dos pontos amostrados irá influenciar no potencial do <i>superpixel</i> apontado em vermelho.	60
Figura 17	– Representação gráfica do modelo de Geoestatística (GS). Os fatores locais são representados em azul e usam a estatística de probabilidade de transição computada pela Eq. 3.3. Diferentemente do que no modelo da Fig. 7, vizinhos de diferentes distâncias também contribuem para calcular a distribuição de cada posição.	61
Figura 18	– Três trajetórias da luz até o plano da imagem. O componente direto, contendo a informação direta da cena. O <i>forward-scattering</i> , contendo informação da cena espalhada. Por fim, o <i>backscattering</i> contendo informações de fora da cena.	64
Figura 19	– Imagem de exemplo para as degradações do ambiente subaquático. É possível ver que existe uma variação conforme a distância e uma perda significativa da informação de cor.	65
Figura 20	– A sequência utilizada para classificação de imagens em meio subaquático.	66
Figura 21	– A cena criada para avaliar os algoritmos de avaliação de features. Ela é composta por lampadas fluorescentes e uma camera fotografando fotos impressas do assoalho do oceano.	72
Figura 22	– As imagens utilizadas no teste. As três imagens foram capturadas nas Bahamas em condições de turbidez próximas do ideal em uma resolução de 4928x3264 pixels	73
Figura 23	– As imagens capturadas sob diferentes níveis de degradação devido a turbidez, controlado pela adição de leite. Foram fotografadas três fotos impressas diferentes, P_1 (primeira coluna), P_2 (segunda coluna) e P_3 (terceira coluna). Na primeira linha foi mostrada a imagem limpa (sem leite) para cada foto capturada. A segunda linha apresenta o intervalo de <i>Baixa Turbidez</i> com por volta de 15ml de leite (T_4). O intervalo de <i>Média Turbidez</i> é mostrado na segunda linha e contém por volta de 50 ml de leite (T_{10}). Finalmente, na ultima (quarta) linha é mostrado o intervalo com <i>Alta turbidez</i> tendo por volta de 100 ml de leite (T_{16}). Quantidade de leite setada para uma caixa com 1000 litros de água. . .	76
Figura 24	– Repetibilidade (Taxa de Acerto) contra o índice de degradação estrutural normalizado (<i>NSDI</i>). As linhas em laranja indicam os intervalos de degradação. <i>Baixa Turbidez</i> 0 até 0.25; <i>Média Turbidez</i> , 0.25 até 0.75, e <i>Alta Turbidez</i> de 0.75 até 1.	79

Figura 25	– Comparação entre a geração de <i>um</i> nível do <i>kernel</i> do espaço de escala usado por quatro detectores diferentes. O kernel foi aplicado em níveis de turbidez diferentes para a imagem P_1 . Sendo que a primeira linha é a imagem limpa (T_0), a segunda linha é uma imagem com baixo nível de degradação (T_4), a terceira linha apresenta uma imagem com médio nível de degradação (T_{10}), a quarta linha apresenta imagens do nível de degradação alto (T_{16}). Para cada caso é mostrado o resultado de filtro equivalente a a aproximadamente um <i>kernel</i> gaussiano de $\sigma = 59.0$. Primeira Coluna: Gaussiano puro. Segunda Coluna: Borramento aproximado em caixas. Terceira Coluna: Difusão utilizando um polígono estrelar de seis pontas. Quarta Coluna: <i>kernel</i> anisotrópico g^2 do <i>KAZE</i> . É possível ver de certa forma estruturas mais definidas para o esquema de difusão usado pelo CenSurE (AGRAWAL; KONOLIGE; BLAS, 2008).	81
Figura 26	– Partes manualmente segmentadas utilizadas para treinamento do classificador. A esquerda são mostrados exemplos de nove amostras usadas para treinar o <i>dataset Redsea</i> . A direita são apresentadas nove amostras do <i>dataset Marker</i> .	85
Figura 27	– Curvas de confiança geradas no treinamento unário de cada classe para o <i>dataset Redsea</i> . A curva de confiança C_i treinada para cada uma das classes é mostrada, se bem como o grau de confiança obtido.	86
Figura 28	– Curvas de confiança geradas no treinamento unário de cada classe para o <i>dataset Marker</i> . A curva de confiança C_i treinada para cada uma das classes é apresentada, bem como o grau de confiança obtido.	87
Figura 29	– Vetores de transição obtidos na etapa de treinamento para o método de Geoestatística do Capítulo 3. Os vetores indicam a probabilidade de uma classe transitar para outra a uma determinada distância. O eixo x apresenta a distância em pixels. O eixo y dos gráficos apresenta as probabilidades de transição. Pode-se observar, por exemplo, uma certa tendência na classe <i>Urchin</i> em transitar para categoria de <i>background</i> .	88
Figura 30	– Mapa temático dos Mosaicos para o <i>dataset Redsea</i> . As figuras mostram a porcentagem de acerto relativa ao <i>GroundTruth</i> . As classes são representadas pelas seguintes cores: Verde <i>Brain Coral</i> ; Amarelo <i>Branchin Coral</i> ; Azul <i>Faviid Coral</i> ; Magenta <i>Urchin</i> e sem cor é o <i>background</i> . Os seguintes resultados são mostrados. (30a) classificação Unária. (30b) mostra a classificação Unária baseada nas curvas de confiança. (30c) classificação com adição de contexto baseada em Geoestatística. (30d) classificação com adição de contexto utilizando <i>CRF</i> .	90

Figura 31 – Mapa temático dos Mosaicos para o *dataset Marker*. As figuras mostram a porcentagem de acerto relativa ao *GroundTruth*. As classes são representadas pelas seguintes cores: Verde *Sand*; Amarelo *Sea Gorgon*; Azul *Corals* e sem cor é o background. Os seguintes resultados são mostrados. (31a) classificação Unária. (31b) mostra a classificação Unária baseada nas curvas de confiança. (31c) classificação com adição de contexto baseada em Geoestatística. (31d) classificação com adição de contexto utilizando *CRF*. 91

Figura 32 – Resultados de classificação para os datasets *Marker* e os datasets *Red-sea*. A primeira coluna apresenta a classificação unitária. A segunda coluna apresenta os resultados de Geoestatística. A terceira coluna apresenta os resultados para o CRF. Por fim, a última coluna apresenta o *GroundTruth*. Foi utilizada como peso para o potencial local w_i como sendo 0.4 para ambas as abordagens. Na primeira coluna foi possível perceber um resultado melhor para o CRF devido a uma maior suavização local. Na segunda linha, o método de Geoestatística obteve melhores resultados devido a suas medidas estatísticas de longa distância. Na última linha é mostrado os resultados para o *dataset Marker*, onde ambas as abordagens tiveram melhores resultados para esse caso. 92

Lista de tabelas

Tabela 1 – A quantidade de leite adicionada para cada nível de turbidez simulado.	74
Tabela 2 – Matriz de covariância que mostra as relações de proximidade entre as classes. Tais medidas são fatores que indicam correlação e não distribuições de probabilidade. Este resultado é normalizado ao final.	86
Tabela 3 – Matriz de covariância que mostra as relações de proximidade entre as classes. Tais medidas são fatores que indicam correlação e não distribuições de probabilidade. Este resultado é normalizado ao final.	87
Tabela 4 – Resultados para a taxa de acerto de diferentes segmentos para o <i>dataset Redsea</i> . Foram testados diversos segmentos quadrados amostrados aleatoriamente nos mosaicos. O tamanho do segmento é especificado pelo lado do quadrado	93

Sumário

Introdução	23
0.1 Motivação	23
0.2 Detectores de Pontos de Interesse	24
0.3 Classificação de Imagens do Assoalho Oceânico	24
0.4 Sumário desta Dissertação	25
1 Fundamentação Teórica 1: Detectores de Pontos de Interesse	27
1.1 Detectores de Única Escala	29
1.1.1 <i>Harris</i>	29
1.1.2 <i>Hessian e Laplacian</i>	30
1.1.3 Comparação	31
1.2 Detectores Invariantes a Escala	31
1.2.1 Hessian-Laplace e Hessian-Laplace	32
1.2.2 <i>Difference-of-Gaussians(DoG)</i>	33
1.2.3 <i>Fast Hessian</i>	33
1.2.4 <i>Center Surround Extrema Filters(CenSurE)</i>	35
1.2.5 KAZE	35
1.2.6 Comparação	36
2 Fundamentação Teórica 2: Classificação de Imagens Utilizando Contexto	39
2.1 Utilização do Contexto em Visão Computacional	39
2.1.1 Níveis de Contexto	40
2.1.2 Interações de Contexto	40
2.2 Integração de Contexto Na Classificação	41
2.2.1 Integrando contexto com base em Classificadores	42
2.2.2 Integrando contexto com base em Modelos Probabilísticos Gráficos	42
2.2.2.1 O Problema da Inferência Estatística	44
2.2.2.2 Aprendizado de parâmetros	45
2.3 Trabalhos utilizando Modelos Probabilísticos Gráficos	45
2.4 Sumário	47
3 Classificação Baseada em Contexto utilizando Geoestatística	49
3.1 Visão Geral da Proposta	49
3.2 Nível Unário $P_u(L \theta_u)$	50
3.2.1 Classificador	51
3.2.2 Treinando Curvas de Confiança	51

3.3	Distribuição de Probabilidades	52
3.4	Nível Local $P_l(L W)$	53
3.4.1	Medindo Transições de Probabilidades	54
3.4.1.1	Taxa de Transição Medida R_{mes}	55
3.4.1.2	Calculo da Matriz R_{mod}	57
3.4.2	<i>Sequential Indicator Simulation</i>	59
3.4.3	Computando o Potencial Final $P(L)$	59
3.5	Geoestatística e CRF	60
3.6	Sumário	61
4	Classificação de Imagens do Assoalho Oceânico	63
4.1	Propriedades de Imagens Subaquáticas	63
4.2	Classificação Autônoma de Imagens do fundo Oceânico	66
4.2.1	Pré-Processamento	67
4.2.1.1	Contraste	67
4.2.1.2	Correção de Cor	67
4.2.2	Segmentação	68
4.2.3	Descritores	68
4.2.4	Treinamento e Classificação	69
4.3	Conclusões	70
5	Testes e Resultados 1: Detecção de Pontos de Interesse em Ambiente Subaquático	71
5.1	Descrição do experimento	72
5.1.1	Cena Montada	72
5.1.2	Procedimento	74
5.2	Avaliando a degradação causada pela turbidez	75
5.3	Resultados	76
5.3.1	Procedimento de Avaliação	77
5.3.2	Comparação	78
5.4	Conclusões finais	80
6	Testes e Resultados 2: Contexto em Classificação Subaquática	83
6.1	<i>Datasets</i> Utilizados	83
6.2	Descrição do Geral do Sistema	83
6.2.1	Pré-Processamento	84
6.2.2	Segmentação e Descrição	84
6.2.3	Classificação	84
6.2.4	Adição de Contexto	84
6.3	Treinamento	84

6.3.1	Treinamento do Classificador	84
6.3.2	Treinamento Unário	85
6.3.3	Treinamento Potenciais Locais	86
6.4	Sistemas Testados	89
6.5	Computação do Mapa Temático	90
6.6	Conclusões	93
7	Conclusões Finais	95
7.1	Detectores de Pontos de Interesse em Imagens Subaquáticas Turvas	95
7.1.1	Contribuições Obtidas	95
7.1.2	Limitações e Trabalhos Futuros	95
7.2	Adição de Contexto Baseado em Geoestatística	96
7.2.1	Contribuições Obtidas	96
7.2.2	Limitações e Trabalhos Futuros	96
	Referências	97

Introdução

Este trabalho apresenta um estudo sobre técnicas de visão computacional considerando os aspectos fotométricos do meio subaquático. Dois pontos-chaves no processo são analisados: A extração autônoma de pontos de interesse; e a utilização do contexto espacial para a classificação.

0.1 Motivação

Visão computacional é a ciência que visa possibilitar às máquinas a capacidade de interpretação e representação de informações visuais.

Com tal capacidade, diversas aplicações podem ser então desenvolvidas, como: a inspeção industrial autônoma, a reconstrução de cenas em três dimensões, a localização de robôs, a rotulação ou classificação de objetos em imagens, entre outras. Tais aplicações podem ser implementadas nos mais diversos ambientes, desde o domínio industrial ou doméstico, em ambientes fechados ou abertos, em localidades sobre a terra ou no fundo do mar, etc.

Neste trabalho se dá atenção especial ao ambiente marinho, o qual, cobrindo em torno de 70% da terra, e contendo cerca de 90% de sua biodiversidade, é de evidente importância. O advento dos *Remotely Operated Vehicles (ROV)* e dos *Autonomous Underwater Vehicles (AUVs)*, tem possibilitado a coleta de milhares de imagens para monitoramento do oceano, ampliando as possibilidades de aplicações em visão computacional em ambientes subsea.

Diversas das aplicações para visão computacional em terra podem ser facilmente extrapoladas para utilização no meio subaquático. Como exemplo, tem-se o caso da classificação autônoma de imagens do assoalho oceânico. Para tal aplicação, tem-se o caso dos recifes de corais, os quais desde 1980 sofrem de massivas perdas devido a poluição, pesca excessiva e espécies invasivas (NEMETH et al., 2008). Uma classificação autônoma é fundamental, dada a grande área monitorada e a necessidade de reduzir o tempo necessário de especialistas para se classificar as espécies.

O monitoramento é também uma realidade que gera demanda para sistemas robóticos autônomos. Isso gera margem para utilização de sistemas visuais nas mais diversas aplicações como: a localização de robôs subaquáticos, a inspeção e rastreamento de *risers* e *flows* na indústria de óleo e gás, entre outras.

Devido a dificuldade de desenvolvimento e instrumentação *subsea*, o uso de *ROVs* e *AUVs* é recente, implicando em uma limitada quantidade de estudos relacionados ao

domínio da visão computacional para o ambiente subaquático. Nesse ambiente existem desafios específicos que não necessitam ser tratados em outros ambientes. A propagação da luz em meio subaquático apresenta efeitos fotométricos associados o que causa degradação na formação da imagem. Efeitos como borramento, espalhamento da informação luminosa e atenuação de cor na imagem, são alguns exemplos que precisam ser considerados em aplicações subaquáticas

Além disso, quando se trata das cenas capturadas em tais ambientes, a monotonicidade do ambiente, dada pela falta de diversificação dos objetos e a falta de estruturas geométricas bem definidas, muito causada pela erosão, dificulta a interpretação visual, o que por sua vez, acarreta no aumento da complexidade das aplicações em visão computacional.

Neste contexto, o objetivo deste trabalho é analisar duas aplicações fundamentais para visão computacional em meio subaquático: a detecção de pontos de interesse e a consideração de contexto para classificação de grande extensões de mosaicos de imagens do assoalho oceânico.

0.2 Detectores de Pontos de Interesse

A detecção de pontos de interesse é de fundamental importância para diversas áreas fundamentais de uso no meio subaquático, como a classificação de imagens (PADMAVATHI; MUTHUKUMAR; THAKUR, 2010), reconstrução 3D (BEALL et al., 2010) (NICOSEVICI et al., 2009), localização de robôs (AULINAS et al., 2011), etc.

Este trabalho propõe um novo *dataset* de imagens subaquáticas, o qual é usado para apontar os detectores de pontos de interesse mais adaptados ao meio subaquático. Tal *dataset* deve ser capaz de isolar a degradação causada pelo comportamento da propagação da luz em meio subaquáticos como a principal fonte de degradação.

Serão testados detectores, considerando diversos paradigmas para detecção, com respeito a sua robustez a degradação das imagens subaquáticas. Um especial tratamento será dado aos detectores invariantes a escala por sua comprovada baixa performance neste meio (GARCIA; GRACIAS, 2011).

0.3 Classificação de Imagens do Assoalho Oceânico

A outra aplicação analisada diz respeito ao uso do contexto espacial, muito pouco utilizado para classificação de imagens subaquáticas e, de fundamental importância, a medida que grandes extensões passam a ser monitoradas.

Na classificação de grandes extensões de recifes de corais, por exemplo, é natural

que as diferentes espécies possam estar inseridas dentro de um contexto. A utilização da informação de contexto pode auxiliar a interpretação da cena (BAR, 2004). Assim, analisam-se diversos algoritmos para a utilização de contexto em situações e cenários genéricos.

Com isso, este trabalho tem como objetivo também propor um novo algoritmo para adição de contexto inspirado em Geoestatística, área que modela a variabilidade de grandezas no espaço geométrico, de tal forma o método proposto seja capaz de mitigar a falta de informação anotada no meio subaquático.

0.4 Sumário desta Dissertação

O Capítulo 1 apresenta a fundamentação teórica sobre a detecção de pontos de interesse. Primeiramente, uma definição formal de pontos de interesse é realizada e suas principais características desejadas são apontadas. São apresentadas as definições dos detectores *Harris* e *Hessian*, *Harris-Laplace*, *Hessian-Laplace*, *Difference of Gaussians (DoG)*, *FastHessian*, *CenSurE* e *KAZE*.

O Capítulo 2 apresenta o problema de classificação de imagens usando o contexto. Uma definição de como representar as relações de contexto em uma imagem é apresentado. Em seguida é feita uma revisão dos métodos de visão computacional os quais incorporam esses conceitos. Um destaque é dado aos métodos que utilizam os *Conditional Random Field (CRF)* para incorporar o contexto.

No Capítulo 3 é apresentado um sistema de classificação de imagens baseado em Geoestatística, a qual é uma área da estatística que busca modelar a variabilidade espacial de uma determinada grandeza. Nesse capítulo é proposta uma extensão deste conceito para adição de informação de contexto na classificação de imagens.

O Capítulo 4 apresenta os problemas existentes na classificação de imagens subaquáticas e também, os aspectos especiais que existem para aplicações em visão subaquática. Também se apresenta uma breve revisão dos resultados já obtidos na classificação de mosaicos do bentos sem utilização do contexto e, também, uma visão geral do sistema proposto por Shihavuddin et al. (2013) o qual foi usado como base para os testes e resultados.

O Capítulo 5 apresenta um estudo sobre o comportamento dos detectores apresentados no Capítulo 1 quanto a sua robustez à degradação causada em imagens subaquáticas. É especificado um experimento realizado para geração de diferentes níveis de degradação nas imagens. Por fim, são apresentados os detectores mais adaptados ao meio.

O Capítulo 6 apresenta um estudo de caso da aplicação de contexto para classificação no meio subaquático. Os resultados gerados para o novo método proposto baseado em Geoestatística são mostrados, e também a sua comparação com os demais métodos

do estado da arte em adição de contexto.

Por fim, no Capítulo 7 as conclusões deste trabalho são apresentadas.

1 Fundamentação Teórica 1: Detectores de Pontos de Interesse

Este capítulo apresenta a fundamentação teórica sobre detectores de pontos de interesse. Formalmente pontos de interesse são definidos como um padrão de uma imagem que difere de sua vizinhança imediata (TUYTELAARS; MIKOLAJCZYK, 2008). Normalmente são pontos com particularidades de uma imagem as quais possuem alguma característica visual relevante. Vale notar que, apesar do termo utilizado ser "pontos de interesse", não é utilizada a definição matemática de um ponto infinitesimal sendo definidos como pequenas regiões. Pontos de interesse servem como âncoras de regiões da imagem, determinando quais posições podem ser descritas para se ter uma representação confiável da mesma. Distintas aplicações fazem uso dos pontos de interesse como: a classificação de imagens (PADMAVATHI; MUTHUKUMAR; THAKUR, 2010), reconstrução 3D (BEALL et al., 2010) (NICOSEVICI et al., 2009), mapeamento e localização (GIL et al., 2010), rastreamento (CORKE et al., 2007), etc.

Um exemplo de pontos de interesse seriam as quinas, as quais são responsáveis por boa parte do processo de interpretação visual de um objeto (TUYTELAARS; MIKOLAJCZYK, 2008). Muitas vezes, os pontos de interesse apresentam uma relação semântica mais estreita com a aplicação. Por exemplo, ao classificar faces, as regiões do olho ou da boca são de grande interesse para a classificação.

A utilização de pontos de interesse locais traz as seguintes vantagens, em contraste com o uso do contexto geral da imagem:

- Redução significativa do custo computacional;
- Descarte de parte do ruído presente na imagem pois somente os pontos relevantes são utilizados;
- Obtenção e uso de apenas características mais distintas da imagem;
- Possibilidade de reconhecimento de cenas sem a necessidade de segmentação.

Porém, para um ponto de interesse ser eficaz, a presença de algumas propriedades são de fundamental importância (TUYTELAARS; MIKOLAJCZYK, 2008). Entre as mais importantes tem-se:

- Repetibilidade: Um ponto de interesse deve representar características que possam ser encontradas em determinados objetos, independente da configuração em que tal

objecto foi fotografado. Dado duas imagens de um mesmo objeto ou cena, o que foi visto em ambas as cenas deve ser detectado como ponto de interesse em ambas as cenas.

- Distintividade: Um ponto de interesse deve representar características que sejam distintas, com destaque sobre as demais características e que sejam específicas de um determinado objeto. Só assim este objeto pode ser discriminado com relação aos demais.

A repetibilidade, de fato a propriedade mais importante ([TUYTELAARS; MIKOLAJCZYK, 2008](#)), pode ser atingida tendo os pontos de interesse sendo invariantes a determinadas transformações que uma imagem pode sofrer, tais como:

- Rotação: Um ponto que pertence a uma cena, deve ser encontrado independente da orientação que a cena foi capturada.
- Translação: Se o ponto representa o mesmo objeto, o mesmo deve ser encontrado independente da posição na imagem onde ele foi capturado.
- Escala: Independente da distância em que a cena foi capturada, o mesmo ponto deverá ser encontrado.

Para outras transformações que a imagem possa sofrer, muitas vezes é interessante que um ponto seja somente robusto. Ou seja, capaz de ser encontrado somente até um determinado nível da transformação. Alguns efeitos, ou transformações, a se ter robustez são: efeitos de discretização, artefatos causados por compressão, borramento devido a movimento, ruído branco, distorção de perspectiva, etc.

Diversos algoritmos são desenvolvidos para encontrar pontos os quais apresentam as propriedades acima descritas. São eles chamados os *Detectores de Pontos de Interesse*. Os Detectores são desenvolvidos de forma a terem um valor de retorno alto em relação a certas *estruturas* presentes na imagens. Define-se estrutura como um determinado padrão com respeito a variação de intensidade luminosa em uma região da imagem.

Divide-se os *Detectores de Pontos de Interesse* com respeito as determinadas propriedades as quais os mesmos possuem invariância.

Neste trabalho, selecionou-se nove detectores principais encontrados na bibliografia a serem fundamentados. Primeiramente, na seção 1.1, são expostos os detectores capazes de responder a estruturas possuindo invariância a rotação e translação. Tais detectores são chamados também de detectores de única escala pois, não possuindo invariância a escala, somente analisam a imagem em uma única escala.

Por fim, na seção 1.2, são apresentados detectores que convivem também com a invariância a escala. Estes simulam múltiplas escalas de forma a encontrar pontos invariantes a escala. Tais detectores são chamados de *detectores de múltipla escala*.

1.1 Detectores de Única Escala

Os detectores apresentados nessa seção possuem invariância a translação ou rotação, podendo possuir em algum nível, também invariância a escala. Os detectores apresentados podem também ter robustez a diversos tipos de ruído.

Normalmente um detector é implementado como uma função, ou *kernel*, o qual é *convoluída* com a imagem e produz uma imagem de saída a qual apresenta o resultado da aplicação deste *kernel*.

Como já explicado, existem diversas características em uma imagem a serem buscadas como pontos de interesse. Neste trabalho, tanto para o caso de única, como de múltipla escala, seleciona-se características baseadas na alta curvatura de uma região, calculada através do gradiente da imagem. Duas estruturas são escolhidas, quinas e *blobs*. Ambas são bastante utilizadas como pontos de interesse pelos detectores mais populares.

As quinas, são estruturas que não necessariamente representam uma quina de fato. São estruturas as quais possuem gradientes de alta intensidade em pelo menos duas direções distintas.

Blobs são definidos como regiões que são diferentes em intensidade da região ao redor. Normalmente são associados com algum ponto de extremo na intensidade da imagem (LINDBERG; EKLUNDH, 1991).

1.1.1 Harris

O detector *Harris* (HARRIS; STEPHENS, 1988) é um dos mais populares detectores de quinas encontrados na literatura. Uma quina é detectada quando existir variação em duas direções principais de uma função analítica de auto-correlação na imagem. Tal função indica a variação de intensidade em todas as direções para uma imagem $I(x, y)$ e pode ser definida pela equação 1.1.

$$M = \sigma_D^2 g(\sigma_I) * \begin{bmatrix} I_x^2(x, y, \sigma_D) & I_x(x, \sigma_D)I_y(x, y, \sigma_D) \\ I_x(x, y, \sigma_D)I_y(x, y, \sigma_D) & I_y^2(x, y, \sigma_D) \end{bmatrix} \quad (1.1)$$

onde:

$$I_x(x, y, \sigma_D) = \frac{\partial}{\partial x} g(\sigma_D) * I(x, y) \quad (1.2)$$

e g é uma função gaussiana definida por:

$$g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1.3)$$

A quina pode ser computada por uma análise dos autovalores da matriz M . Quando os dois autovalores tiverem valores altos, isso indica a existência de uma quina. Uma maneira de medir a intensidade dos autovalores sem a necessidade de computar os autovalores de M diretamente, é através da medida de *Harris* dada por:

$$cin = det(M) - k \text{traco}(M), \quad (1.4)$$

onde k é uma constante normalmente setada entre 0.04 e 0.06. Quando a medida cin da Eq. 1.4 for alta, a presença de quinas também o será (HARRIS; STEPHENS, 1988). Um ponto da imagem é considerado uma quina, se a saída da aplicação da Eq. 1.4 for maior que um limiar t .

Harris já foi avaliado como sendo o detector com melhor repetibilidade quando comparado com outros detectores de única escala (SCHMID; MOHR; BAUCKHAGE, 2000).

1.1.2 Hessian e Laplacian

O detector *Hessian*, proposto inicialmente por (BEAUDET, 1978), é um método bastante usado para detecção de *blobs* em imagens. Para uma imagem $I(x, y)$, os *blobs* podem ser calculado pelo determinante da matriz Hessiana:

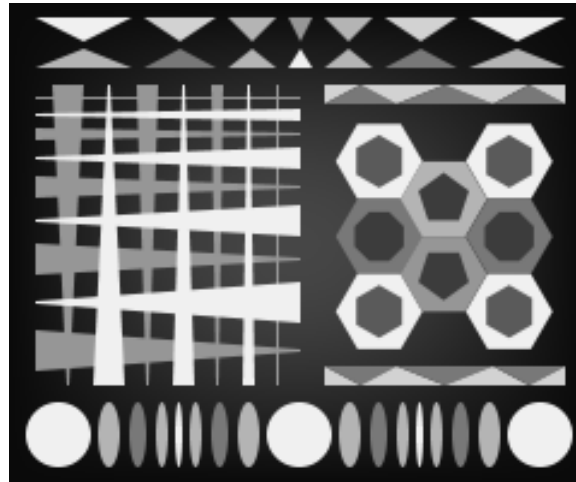
$$H = \begin{bmatrix} I_{xx}(x, y, \sigma_D) & I_{xy}(x, y, \sigma_D) \\ I_{xy}(x, y, \sigma_D) & I_{yy}(x, y, \sigma_D) \end{bmatrix} \quad (1.5)$$

O determinante responde aos gradientes em múltiplas direções da imagem e tende a revelar *blobs* de alta curvatura, o que representa uma região distinta. O detector, determinante de *Hessian*, ou simplesmente detector *Hessian* é dado selecionando os pontos os quais tem uma saída com respeito a matriz H maior que um valor de limiar t .

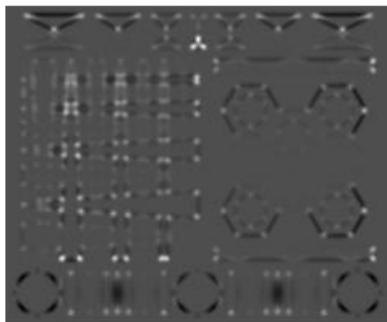
Uma variação do *Hessian* é a aplicação de um *kernel Laplacian* o qual é computado pelo traço da matriz H ($I_{xx} + I_{yy}$). Porém o *Laplacian* tende também a responder a bordas (TUYTELAARS; MIKOLAJCZYK, 2008). Bordas não são bons pontos de interesse pois, não possuem uma aceitável invariância a rotação (TUYTELAARS; MIKOLAJCZYK, 2008).

1.1.3 Comparação

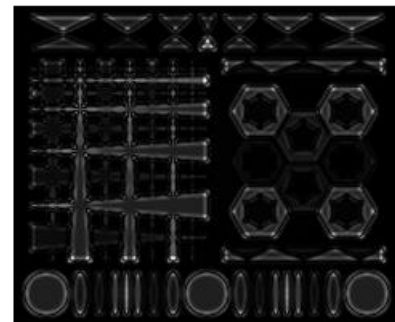
A Figura 1 mostra um exemplo de aplicação do *Hessian* e *Harris* em uma imagem. As *blobs* podem ser vistos como a área mais elevada em *morros* de intensidade. As quinas podem ser juntas "T" ou "L", também podendo ter formato mais arredondado.



(a) Original



(b) Harris



(c) Hessian

Figura 1 – Comportamento da aplicação dos *kernels Hessian* e *Harris* para uma imagem teste (1a). (1b) mostra a saída da medida de *Harris* (Eq. 1.4). (1c) mostra a saída do determinante da matriz *Hessian* (Eq. 1.5) para a imagem teste. Tanto o *Hessian* como o *Harris* tem como saída as regiões de alta curvatura (Figura por Sojka (2003)).

Percebe-se que há semelhança entre ambos, dado que ambos são associados a regiões de alto gradiente.

1.2 Detectores Invariantes a Escala

A noção de escala é crucial na interpretação de uma imagem (LINDBERG, 1994). Alguns objetos só são entidades visuais significativas em uma determinada escala. Sendo assim, uma modelagem explícita de cada nível de escala se torna necessário para o pro-

cessamento (LINDBERG, 1998). Ou seja, uma imagem não mais pode ser representada como uma matriz $I(x, y)$ e passa a ter um terceiro componente de escala s , sendo assim determinada como a função $L(x, y, s)$.

Para gerar o conjunto espaços de escala $L(x, y, s)$, pode-se utilizar o princípio da difusão (LINDBERG, 1994). O qual determina que uma família de escalas L pode ser determinada através da equação da difusão:

$$\partial_\sigma L = \frac{1}{2} \nabla^2 L = \frac{1}{2} (\partial_{xx} + \partial_{yy}) L \quad (1.6)$$

O que representa o fato de que, à medida que a escala se torna menos detalhada, a informação visual tende a se dispersar.

Portanto, para a geração do espaço de escala de uma imagem $L(x, y, s)$ deve ser proposta uma equação que atenda a Equação 1.6. Inicialmente, foi adotado que a função gaussiana seria a única a ser uma solução da equação 1.6. Posteriormente, outras funções foram colocadas como possíveis para geração do espaço de escala (LINDBERG, 1997). Considerando determinada uma escala s , definida igual a um parâmetro de difusão σ , a geração de um espaço de escala σ é dada por:

$$L(x, y, \sigma) = g(x, y, \sigma) * I(x, y) \quad (1.7)$$

sendo a função gaussiana $g(x, y, \sigma)$ calculada como na Eq.1.3.

De forma a atingir a invariância a escala, os detectores passam a considerar essa função $L(x, y, \sigma)$ para se detectar os pontos de interesse. Porém, (LINDBERG, 1994) determinou que é possível realizar a escolha de uma escala, e tal escala será sempre escolhida independente do ambiente e sem a necessidade de escolha de parâmetros. Caracterizando uma escala onde existe invariância.

Foi sugerido que os pontos de extremo de funções gradientes das estruturas entre as escalas tem propriedades invariantes. Isso representa a escala com o máximo de sensibilidade a função. Tal escala é chamada de escala característica.

Nesta seção apresentam-se alguns detectores invariantes a escala. A ideia de máximo de uma determinada função gradiente entre escalas é usada por todos os métodos apresentados.

1.2.1 Hessian-Laplace e Hessian-Laplace

Um das primeiras extensões para detectores de múltipla escala foram feitas para as funções *Harris* e *Hessian* (MIKOLAJCZYK; SCHMID, 2004) Nestes métodos, o espaço

de escala é gerado por uma equação gaussiana tal como na Eq. 1.7. Os pontos de extremo entre um conjunto de escalas σ_n são computados conforme a Eq. 1.8.

$$|LoG(x, y, \sigma_n)| = \sigma^2 |L_{xx}(x, y, \sigma_n) + L_{yy}(x, y, \sigma_n)| \quad (1.8)$$

sendo a Eq. 1.8 uma representação da função *Laplacian* em múltiplas escalas. Desta forma, são selecionados os pontos extremos que tem alta resposta a função *Harris*, para o caso do Harris-Laplace ou da função *Hessian* para o caso do *Hessian-Laplace*.

1.2.2 *Difference-of-Gaussians(DoG)*

O detector *DoG* é uma otimização a aplicação do *Hessian-Laplace*. É o detector proposto pelo método *SIFT* (LOWE, 2004).

Ao invés de computar o *Laplacian* para cada escala, neste aplica-se o *Laplacian* pela diferença, $D(x, y, \sigma)$, entre múltiplos níveis do espaço gaussiano $L(x, y, \sigma)$. Sendo assim:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1.9)$$

Diversos níveis de escala são gerados. A cada determinado número de imagens, chamado oitava, é feito um redimensionamento na imagem. Dentro de uma oitava, as imagens diferentes são criadas pela aplicação do filtro gaussiano. A função D é gerada a partir da diferença entre níveis vizinhos. O processo utilizado pelo algoritmo é mostrado na Figura 2.

Para se encontrar a escala característica, basta encontrar o máximo na função $D(x, y, \sigma)$ variando σ . Ao final, os extremos do espaço, os quais tem baixa resposta à função *Hessian* são eliminados.

1.2.3 *Fast Hessian*

Trata-se de um método que busca fazer uma otimização ainda maior em relação ao *DoG* para geração do espaço de escala (BAY et al., 2008). Trata-se de um filtro que não usa o filtro gaussiano para geração do espaço de escala. Os filtros gaussianos são aproximados por filtros caixas. Um filtro caixa basicamente computa a média de uma imagem dado uma janela de convolução, podendo ser computado rapidamente pela utilização de Imagens Integrais (DERPANIS; LEUNG; SIZINTSEV, 2007).

É possível neste caso fazer a abordagem de diferença de caixas, o que permite juntar a aplicação do filtro *Hessian* com a geração do espaço de escala. Uma aproximação do *Hessian* já é computada diretamente ao se aplicar as diferenças de caixas.

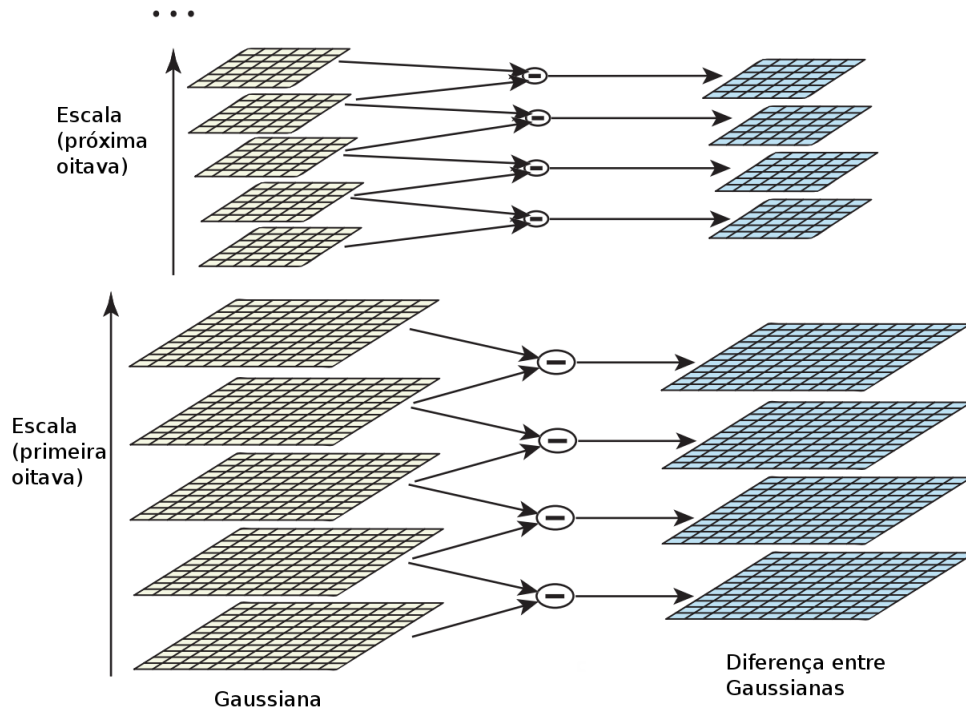


Figura 2 – O processo para geração do espaço de escala pelo *DoG*. Ao invés de computar o *Laplacian* para cada escala, o mesmo é estimado pela diferença entre escalas consecutivas. Figura adaptada de (LOWE, 2004).

Para um determinado tamanho de caixa de aresta N , a resposta do *Hessian* é dada por:

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (1.10)$$

A Figura 22 mostra um exemplo de filtros D_{xx} e D_{xy} que são aplicados. Enquanto aplicar a diferença entre caixas, produz o *Hessian*, a computação em blocos aplica a difusão na imagem.

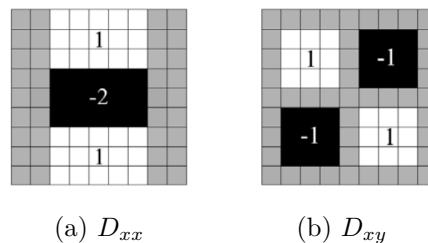


Figura 3 – Exemplo de um filtro caixa de tamanho 9x9 aplicado para geração de um espaço de escala equivalente a $\sigma = 1.2$. Outros espaços podem ser gerados usando caixas maiores.

Para relacionar com o espaço gaussiano, basta saber que uma imagem de filtro gaussiano $\sigma = 1.2$ é equivalente a utilização de um filtro caixa 9x9. Então, para geração

do espaço de escala basta gerar a resposta de vários tamanhos de caixa $N = 9, 11, 13..$ etc.

Para encontrar os pontos característicos basta encontrar o máximo para todos os níveis de escala.

1.2.4 Center Surround Extrema Filters(*CenSurE*)

O Filtro de centro e arredores (AGRAWAL; KONOLIGE; BLAS, 2008) tem uma abordagem similar a utilizada pelo detector *Fast Hessian*, porém realizando a diferença entre múltiplos níveis como no caso do *DoG*. Este processo visa também uma aproximação do *Laplacian*. Os espaços de escala são criados pela geração de polígonos de múltiplos tamanhos. Tal como o filtro caixa, um filtro poligonal representa a media através de uma janela de convolução.

De maneira similar ao *FastHessian*, um filtro de polígono lado $N = 2$ é equivalente a um espaço gaussiano de $\sigma = 1.88$ A Figura 4 mostra alguns tipos de polígonos usados para gerar o espaço de escala. Os polígonos podem ter estruturas estreladas, poligonais, entre outras.

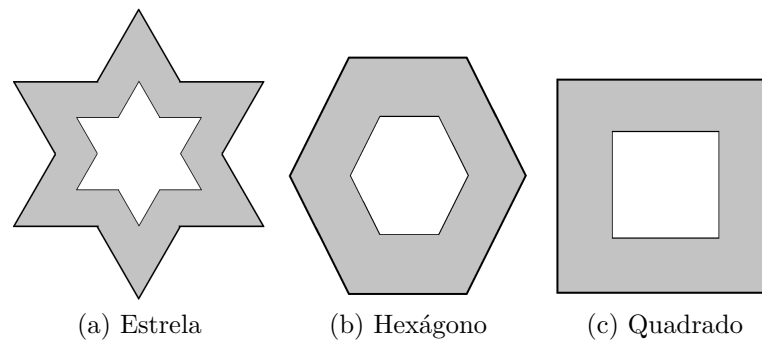


Figura 4 – Alguns tipos de filtros utilizados para geração do espaço de escala pelo *CenSurE*. O filtro estrela, o filtro hexagonal e o filtro por diferença de caixas.

Para gerar o *Laplacian* uma imagem a qual teve aplicada um filtro maior é subtraída de uma imagem com um polígono menor aplicado.

O máximo deste espaço gerado é encontrado como pontos de interesse.

Por fim uma função *Harris* é aplicada, eliminando os pontos que obtiveram baixa resposta. Isso segue, pelo fato do *Harris* ter sido determinado como uma função com melhor repetibilidade.

1.2.5 KAZE

A ideia do *KAZE* (ALCANTARILLA; BARTOLI; DAVISON, 2012) é gerar um espaço de escala suavizando de maneira diferente em locais com alta intensidade de bordas. Tal abordagem trata-se de uma difusão não linear.

A geração do espaço de escala é dada pela solução para equação da difusão não linear:

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \nabla L), \quad (1.11)$$

Abordagens que aplicam uma difusão não linear podem obter resultados melhores para o caso da segmentação de imagens e remoção de ruído (WEICKERT; ROMENY; VIERGEVER, 1998).

A ideia principal é que, durante a formação da escala, as bordas das estruturas devem se manter mais do que de fato acontece com o filtro gaussiano. Desta forma, primeiramente uma função ∇I que responde as bordas é aplicada. Sendo ∇I basicamente um gradiente da imagem. Com base nessa saída, uma definição alternativa da função gaussiana é aplicada para geração do espaço de escala. Perona e Malik (1990), descreveram algumas possíveis formulações de funções:

$$c1 = \exp\left(\frac{|\nabla I_\sigma|^2}{k^2}\right), \quad c2 = \exp\left(\frac{1}{1 + \frac{|\nabla I_\sigma|^2}{k^2}}\right) \quad (1.12)$$

sendo k um parâmetro que controla o nível de difusão. Alcantarilla, Bartoli e Davison (2012) propôs uma terceira formulação de *kernel*:

$$c3 = \begin{cases} 1 & , |\nabla I_\sigma|^2 = 0 \\ 1 - \exp\left(\frac{3.315}{(|\nabla I_\sigma|/k)^8}\right) & , |\nabla I_\sigma|^2 > 0 \end{cases} \quad (1.13)$$

Levando em conta os kernels definidos, cada nível do espaço de escala $Lk(x, y, t)$ é gerado pela aplicação da seguinte função recursiva:

$$Lk(x, y, t + 1) = (I - (t_i + 1 - t_i).c(x, y, t) * Lk(x, y, t))^{-1} Lk(x, y, t) \quad (1.14)$$

onde t é um parâmetro de escala temporal facilmente relacionado a σ . Ao final, também são desconsideradas as regiões que tem baixa resposta a aplicação de uma matriz *Hessian*.

1.2.6 Comparação

A Figura 5 apresenta exemplos de geração do espaço de escala baseado em 4 funções diferentes. O gaussiano, utilizado pelo *DoG*, *Harris-Laplace* e *Hessian-Laplace*, a função de caixas utilizada pelo *FastHessian*, uma função poligonal estrelar de seis lados, utilizadas pelo *CenSurE* e a função de difusão anisotrópica utilizado pelo *KAZE*.

Pode-se perceber que determinadas estruturas se mantem mais que outras para espaços diferentes. Claramente algumas aplicações se beneficiariam do uso de um espaço de escala diferente. No Capítulo 5 se estudam os melhores detectores para o campo de estudo de imagens subaquáticas com presença de turbidez.

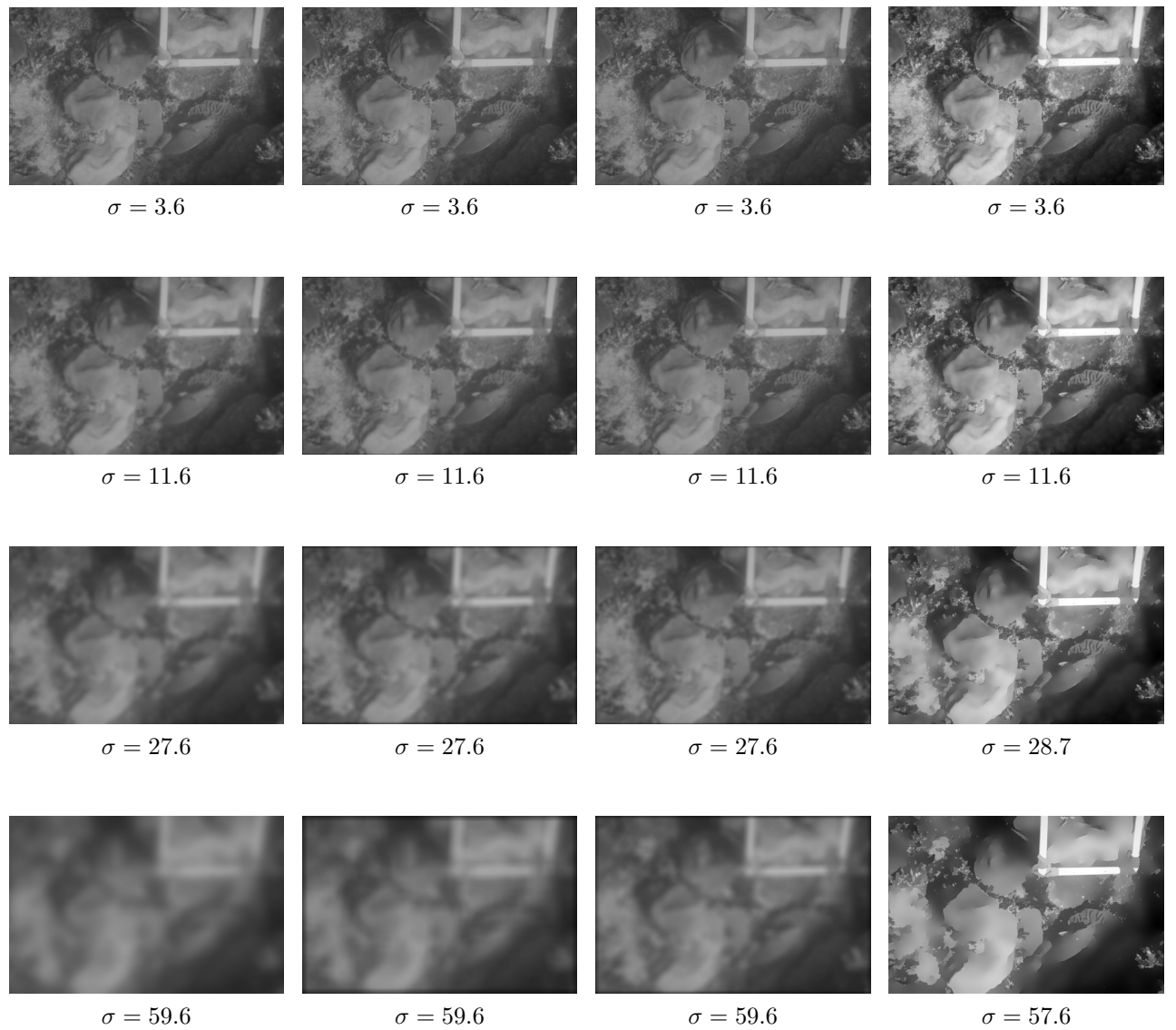


Figura 5 – Espaços de escala gerados. Primeira coluna mostra o espaço Gaussiano. Segunda coluna mostra o filtro média de caixas usado pelo *FastHessian*. Terceira coluna mostra um filtro poligonal estrelar de seis lados. A quarta Coluna mostra o espaço de escala anisotrópico usado pelo *KAZE*.

2 Fundamentação Teórica 2: Classificação de Imagens Utilizando Contexto

Este capítulo apresenta a fundamentação teórica utilizada neste trabalho associada a utilização de contexto para a classificação de imagens. Inicialmente são postuladas as definições de como representar as relações de contexto em uma imagem. Também é feita a definição de classificação de imagens incorporando o conceito de contexto, bem como uma revisão dos métodos de visão computacional que os tratam são apresentados. Um destaque é dado aos métodos que utilizam os *Conditional Random Field (CRF)*.

2.1 Utilização do Contexto em Visão Computacional

Existem diversos descritores capazes de discriminar os objetos com base em suas características visuais, como textura, cor e forma. Tais características buscam capturar a variabilidade dos objetos para sua classificação ([GALLEGUILLOS; BELONGIE, 2010](#)). Porém, estudos envolvendo o sistema perceptivo visual humano trazem novas perspectivas no que tange a como as típicas configurações dos objetos em uma cena podem contribuir para a percepção, de tal forma que o reconhecimento de objetos no sistema visual humano considera não somente os aspectos locais referentes a interpretação da cena, mas também a situação geral onde um objeto foi encontrado.

Biederman ([BIEDERMAN; MEZZANOTTE; RABINOWITZ, 1982](#)) estabelece alguns tipos de relações contextuais importantes que são fundamentais para o reconhecimento de objetos no sistema visual humano. Estas relações estabelecem níveis semânticos tais como: *i* suporte (onde os objetos tendem a sustentar ou ser sustentados por outros), interposição (relativo a relações de oclusão), *ii* probabilidade (objetos tendem a aparecer na mesma situação), *iii* posição (objetos tendem a ficar em determinada posição relativa com outros) e *iv* tamanho (objetos tendem a ter certo tamanho se comparado com outros em uma dada escala).

Vários modelos computacionais já fizeram uso destas relações semânticas as quais podem ser usadas para classificar objetos. Essas relações normalmente são resumidas em três tipos de contexto principais: semântico, posição e escala.

O contexto semântico, tende a incluir as relações de ocorrência entre objetos. Ao encontrar um determinado objeto em uma cena, o qual se possui certeza de sua presença, considera-se uma maior probabilidade de presença de outros objetos. Por exemplo, a existência de um bule de chá implica em uma maior probabilidade de existência de

outros utensílios de cozinha como talheres ou um fogão (FISCHLER; ELSCHLAGER, 1973) (HANSON; RISEMAN, 1978). Como exemplo, Rabinovich et al. (2007), incorporou a informação anotada pelos *Google Sets* indicando objetos que tendem a aparecer em situações semelhantes de forma a melhorar a classificação.

O contexto de posição indica que os objetos tendem a ter uma relação espacial na imagem. Como por exemplo, o céu em uma imagem tende a estar acima do chão. Já o contexto de escala está associado às relações de tamanho entre objetos na cena (TORRALBA, 2003) (KUMAR; HEBERT, 2005) (TORRALBA; MURPHY; FREEMAN, 2004). Pois, existe já, uma relação de tamanho típica que os objetos possuem entre eles (GALLEGUILLOS; BELONGIE, 2010).

Para incluir tais tipos de contexto na classificação de imagens, alguns aspectos fundamentais devem ser considerados. Primeiramente, qual nível de contexto será classificado. Se as relações a serem consideradas serão apenas entre objetos próximos, ou no domínio geral de uma imagem. Ou ainda, por meio de que tipo de estrutura visual encontrada na imagem, a interação de contexto ocorre.

2.1.1 Níveis de Contexto

Os sistemas que adicionam contexto na classificação normalmente dividem o contexto em dois níveis (GALLEGUILLOS; BELONGIE, 2010): local e global.

O contexto local é onde somente as interações de vizinhança são utilizadas para adicionar o contexto a um determinado objeto. O contexto local está relacionado aos objetos que cercam outros objetos. Vale notar que a aplicação de contexto é recursiva, ou seja, a própria vizinhança possui também suas próprias relações de contexto. Isso faz que não somente as relações estritamente próximas façam parte do contexto local.

O contexto global está relacionado às interações de contexto presentes ao longo de toda a imagem utilizada. O contexto global normalmente está associado ao ambiente onde os objetos estão posicionados. Por exemplo, se as relações contextuais indicam que os objetos estão em uma cozinha, isso implica em uma alta probabilidade de um dos objetos ser uma panela.

2.1.2 Interações de Contexto

Não necessariamente cada componente da imagem deve ser um objeto com um conceito semântico relacionado. Na literatura se estabelecem três níveis básicos de interação nos quais o contexto pode ser integrado (GALLEGUILLOS; BELONGIE, 2010). Além de objetos, as interações também se dão entre pixels ou regiões.

A interação em nível de pixel estabelece que pixels vizinhos tendem a ter a mesma classe. Tais interações ajudam a inferir as bordas existentes na imagem. Vale notar a

utilização de tais interações são mais computacionalmente intensas, dado que existem diversas combinações entre pixels da imagem. Ressalta-se que o uso de níveis mais baixos de contexto não necessariamente implica na perda da informação semântica. Ou seja, encontrar que pixels de determinado objeto são próximos, é também identificar a alta probabilidade de proximidade de tais objetos.

O conceito de pixel pode ser estendido para o nível de representação de regiões. Ao utilizar regiões, tende-se a reduzir a complexidade levantada pelo grande número de pixels existentes na imagem. Uma estrutura de região bastante utilizada é a consideração de pequenas regiões adaptadas a estrutura local da imagem. Tais regiões chamadas, *superpixels*, capturam a redundância dos dados, facilitando a utilização do contexto (FULKERSON; VEDALDI; SOATTO, 2009).

Já a interação em nível de objetos é a representação mais natural para reconhecimento de contexto humano (BAR, 2004). Sabendo-se já a classe do objeto é possível treinar as relações de contexto. No trabalho de (TORRALBA; MURPHY; FREEMAN, 2004), os objetos mais fáceis de classificar ajudam, através do contexto, a obter a classe de objetos mais difíceis. Se por um lado usar objetos tende a capturar melhor as interações existentes na cena, o uso do contexto em nível de objetos implica já o conhecimento prévio (classificação) dos objetos existentes na imagem. A interação entre regiões, por outro lado, ajuda a reduzir a quantidade de combinações existentes na interação de pixels, sem a necessidade de um conhecimento maior sobre a imagem (GALLEGUILLLOS; BELONGIE, 2010).

2.2 Integração de Contexto Na Classificação

Nesta seção são apresentadas algumas abordagens para integração do contexto na classificação de imagens. São escolhidos métodos com integração baseada em *superpixels*, com foco para integração local de contexto. Quanto aos tipos de contexto, por considerar o nível de interação como *superpixels*, os principais tipos integrados são os de posição e escala.

Dado a representação da imagem como uma matriz $SP(x, y)$ onde cada elemento sp é um *superpixel*, a definição de classificação é dada pela determinação de um rótulo l_i dentre um conjunto possível de rótulos $L = l_1, l_2, \dots, l_n$ para cada sp . Para classificações sem contexto, apenas o *superpixel* sp é considerado, já para o caso apresentado nesta seção, a vizinhança de sp é também importante para determinar um rótulo l_i .

Nesta seção são especificadas duas formas de incorporar o contexto. Utilizando a vizinhança de um *superpixel* sp diretamente no classificador ou através de modelos probabilísticos gráficos (MPGs).

2.2.1 Integrando contexto com base em Classificadores

As informações locais advindas de uma análise de contexto podem ser incorporadas diretamente aos sistemas de classificação, considerando uma janela de contexto em torno da região a ser classificada (Figura 6)

Fink e Perona (2003) incorporou o contexto local usando a janela da região para treinamento de classificador fracos em um esquema de *boosting*.

Kruppa e Schiele (2003), visando melhorar a classificação de rostos, incorporou a descrição dos descritores da vizinhança local da face em um sistema *Naive Bayes*

O principal problema é que tais aplicações não levam em conta as possíveis correlações entre os vizinhos. Este problema é demonstrado na Figura 6, a vizinhança só afeta o que foi considerado no centro da imagem, sem afetar a si própria. Tais problemas são parcialmente resolvidos criando-se interações mais conectadas, como no caso dos modelos probabilísticos gráficos a serem explicados na próxima seção.

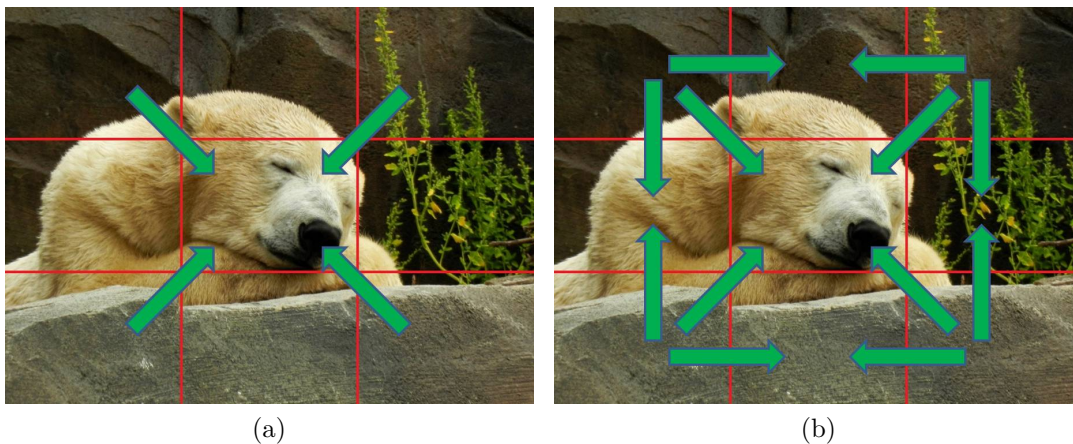


Figura 6 – Janela considerada para a classificação usando contexto. No caso da integração de contexto diretamente nos classificadores (Fig. 6a), não são consideradas as relações entre a vizinhança com si própria (Fig. 6b). Ou seja, se existem propriedades correlacionadas na vizinhança.

2.2.2 Integrando contexto com base em Modelos Probabilísticos Gráficos

Nesta seção, serão apresentados os principais conceitos associados aos Modelos Probabilísticos Gráficos (MPGs) e seu uso na integração contextual em classificação de imagens.

Uma forma natural de representar a dependência entre variáveis é utilizando os Modelos Probabilísticos Gráficos (MPGs) (SUTTON; MCCALLUM, 2006). Estes modelos representam algumas fatorizações de uma função de probabilidades como o *Markov Random Fields* MRF ou *Conditional Random Fields* CRF.

Um MPG é usado para capturar as correlações existentes dentro de um conjunto de dados. Baseado neste modelo, é possível calcular uma função potencial. Esta abordagem, quando baseada em modelos probabilísticos não direcionadas, é usada no MRF e no CRF (SUTTON; MCCALLUM, 2006).

O MRF modela a função de probabilidade $p(y, x)$ de um dado conjunto de rótulos y e os conjunto de descritores de entrada x . Esse modelo necessita um alto custo computacional para classificação de imagens (CARBONETTO; FREITAS; BARNARD, 2004). Ainda, dado que deve seguir a premissa de Markov, nenhuma característica global deve ser adicionada. Para o caso de Markov, a computação de $p(y, x)$ necessita a computação de $p(y)$ e também $p(x)$, o qual não se tem conhecimento sobre, pois está relacionado a probabilidade das descrições de entrada aparecerem.

Uma abordagem mais comumente utilizada para classificação de imagens é o modelo CRF. Neste modelo, somente a distribuição condicional, $p(y|x)$, é computada. Normalmente o CRF tem uma melhor associação aos dados, dado que não é necessário computar a probabilidade a priori para os dados de entrada ($p(x)$) (SUTTON; MCCALLUM, 2006).

Considerando um dado modelo, é definida a probabilidade para um conjunto de rótulos serem atribuídos. Esta probabilidade é estabelecida como uma função de um fator unitário e um fator local. Deste modo, define-se a probabilidade de uma imagem possuir um certo conjunto de rótulos $L = l_1, l_2, \dots, l_n$ dado um grafo de um modelo G e um conjunto de parâmetros θ_u e θ_l como a Eq. 2.1.

$$\log P(L|G, \theta) = w_u \sum_{x_i \in X} \varphi_i^u(x_i, \theta_u) + w_l \sum_{(x_i, x_j) \in \varepsilon} \varphi_{ij}^L(x_i, x_j, \theta_l) \quad (2.1)$$

onde X é um conjunto de vértices no modelo probabilístico gráfico, cada um relacionado a um superpixel da imagem e ε é o conjunto de arestas no grafo de adjacência $G(X, \varepsilon)$. $\varphi_i^u(x_i, \theta_u)$ é a distribuição de probabilidades a priori de um rótulo, para este caso, o conjunto de parâmetros θ_u está associado com o treinamento da geração da distribuição a priori. $\varphi_{ij}^L(x_i, \theta_l)$ é o fator local associado com a probabilidade de duas classes serem vizinhas uma da outra. Neste caso, o parâmetro θ_l está associado as matrizes de covariância, treinadas para indicar as probabilidades proximidade entre os rótulos do conjunto $L = l_1, l_2, \dots, l_n$ considerando sua conexão G .

Os pesos w_u e w_l facilitam a calibração empírica do modelo, determinando a importância de cada termo na Eq. 2.1. A Figura 7, mostra uma representação visual de parte da modelagem usando CRF para a aplicação de interesse que é a classificação e segmentação de imagens.

No modelo CRF também é possível incluir diferentes aspectos baseado em proprie-

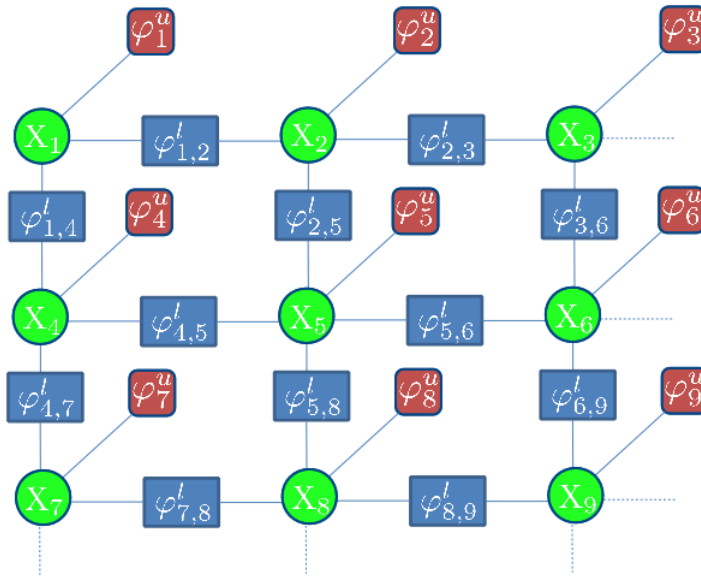


Figura 7 – A representação gráfica de um modelo CRF. Os quadrados em vermelho ($\varphi_i^u(x_i, \theta_u)$) são os fatores unitários calculados com o resultado dado pelo classificador. Os quadrados em azul são os fatores locais computados em cada aresta e utilizados para introduzir informação contextual. Os círculos verdes representam os superpixels sendo classificados.

dades da imagem. A função de bordas de Potts (SHOTTON et al., 2009) (FULKERSON; VEDALDI; SOATTO, 2009) reforça nodos que não são separados por bordas a pertencerem a mesma classe. Isto é implementado incluindo o atributo g_{ij} em φ_i^l . Onde g_{ij} é definido pela Eq. 2.2.

$$g_{ij} = \left[\begin{array}{c} \exp(-\beta \|x_i - x_j\|^2) \\ 1 \end{array} \right] \quad (2.2)$$

Onde β é uma função de contraste dependente da imagem que pode ser facilmente estimada como explicado em (SHOTTON et al., 2009).

2.2.2.1 O Problema da Inferência Estatística

Dado um modelo probabilístico gráfico e uma função de probabilidades, uma das principais dificuldades é encontrar um conjunto de rótulos L' que maximize uma função de probabilidades como a função da Eq. 2.1. Em outras palavras, seria encontrar a configuração de classificação na imagem mais provável, dado um modelo probabilístico. Este problema é considerado *NP-Hard*, dado que existe uma combinação de rótulos exponencialmente grande. O problema de inferência é especialmente difícil quando se utiliza a abordagem MRF, dado que existem muito mais casos para computar a distribuição de probabilidades conjunta.

Algumas aproximações são introduzidas de forma reduzir o custo computacional.

Por exemplo, a abordagem *loopy belief propagation (LBP)* propaga as informações de distribuição de probabilidades dos vértices ao longo do grafo através de mensagens e obtém boa performance (WEISS, 2000), entretanto, a convergência não pode ser garantida. Outra estratégia é o corte de grafos baseado no *alpha-cut* (BOYKOV; JOLLY, 2001). Este algoritmo produz melhores resultados apesar de possuir maior complexidade.

2.2.2.2 Aprendizado de parâmetros

É necessário estimar os parâmetros θ_u and θ_l . Estes parâmetros são a matriz de covariância que representa as tendências das classes serem vizinhas (θ_l), a matriz θ_l esta associada às relações espaciais entre as classes.

Os parâmetros podem ser estimados utilizando a técnica de máximo *a-posteriori* (MAP). Esta técnica seleciona os parâmetros que maximizam os resultados para a Eq. 2.1. Isto é custoso, dado que existe a necessidade de computar a inferência diversas vezes. Porém, é possível realizar a estimativa, parte a parte, dividindo os parâmetros os quais maximizar (SHOTTON et al., 2009), então reduzindo o custo computacional.

A Figura 10 mostra um exemplo de uma matriz de covariância estimada, sendo quanto mais claro for o quadrado mais relacionadas espacialmente as classes estão. É possível perceber que a classe *B* tem uma probabilidade muito maior de ficar próxima ao *C* mas não necessariamente a classe *E*.

2.3 Trabalhos utilizando Modelos Probabilísticos Gráficos

Diversos trabalhos já utilizaram os modelos probabilístico gráficos (MPGs) para adição de contexto. Apresenta-se aqui alguns relevantes para elaboração deste trabalho.

Carbonetto, Freitas e Barnard (2004) foi um dos primeiros trabalhos a usar MPGs para classificação de imagens . O autor utilizou uma versão usando um modelo *MRF* com o contexto local e propôs uma forma de reduzir o tempo de inferência usando uma técnica de *expectation maximization* (EM) (DEMPSTER; LAIRD; RUBIN, 1977).

Shotton et al. (2009) utilizou uma combinação de descritores diretamente dos descritores de textura, cor e localização como fatores unários e adicionou a informação de contexto local usando a medida de potts. O nível de interação foi baseado em regiões utilizando um novo esquema de representação de imagens através "canais de textura". A inferência foi feita utilizando o algoritmo de *alpha-cut* (BOYKOV; JOLLY, 2001).

Fulkerson, Vedaldi e Soatto (2009), utiliza uma representação usando *SIFT* (LOWE, 2004) para *bag-of-words* (SIVIC; ZISSERMAN, 2006) para o fator unário. Como nível de interação, foram utilizadas regiões baseadas em *superpixels*. Em seguida, aplica-se um sistema CRF similar ao proposto por (SHOTTON et al., 2009).

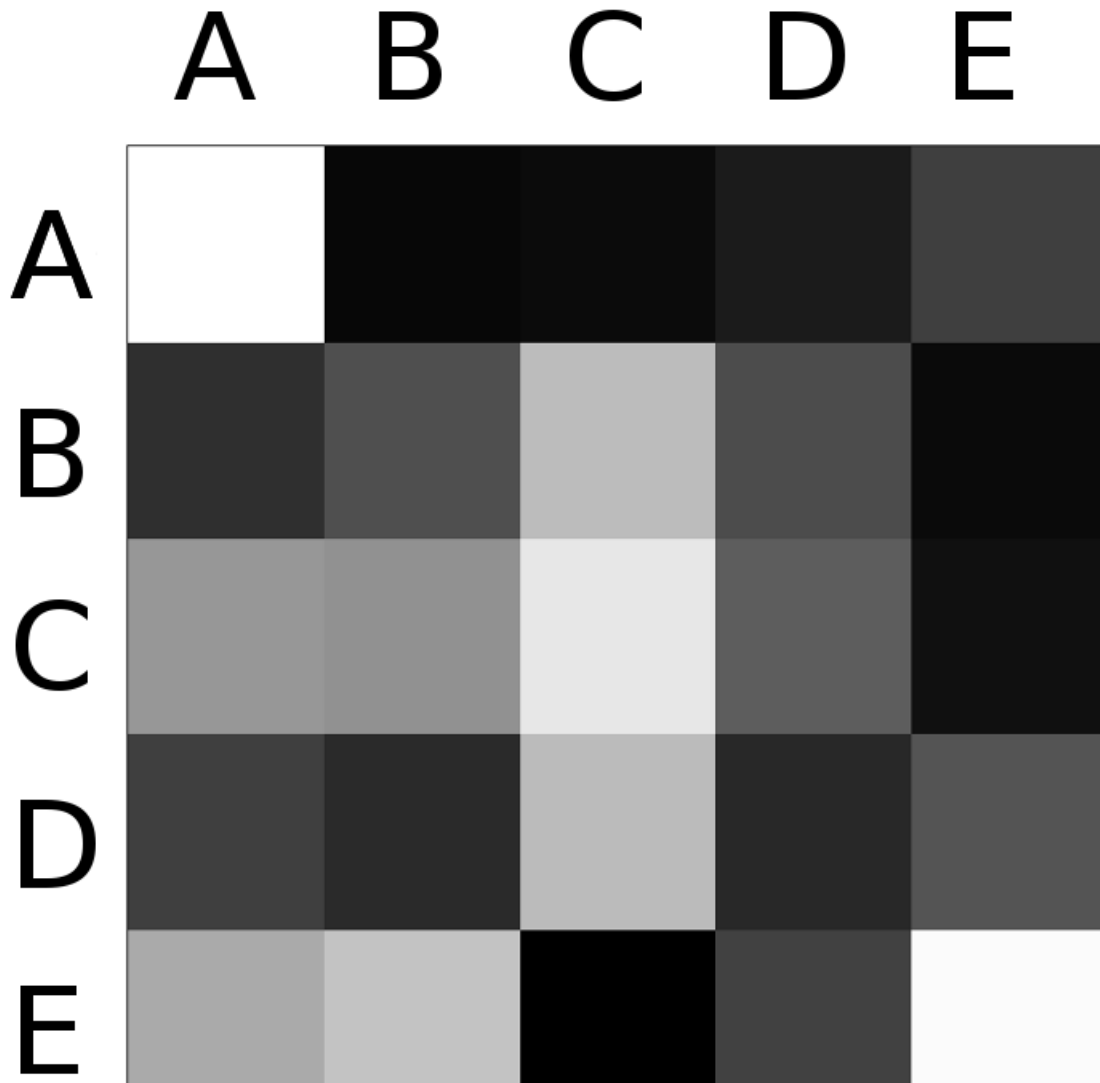


Figura 8 – Tendências que existem para as classes estarem próximas umas das outras, quanto mais claro, maior é a tendência existente. Por exemplo, é possível perceber que a classe *B* é provável de aparecer perto de uma classe *C* mas não próxima de uma classe *E*.

Koltun e Vladlen (2011), utiliza um *CRF* com uma maior conectividade entre os vértices e utiliza um nível de interação por pixel. Neste caso, dado o conjunto de pixels, cada par possível de pixels é conectado. O aumento da complexidade de inferência é resolvido com um sistema aproximado baseado em médias.

Boix et al. (2012) propõem adicionar o contexto global ao *CRF*. Para isso, a Equação 2.1 pode ser estendida para a Eq. 2.3.

$$\log P(L|G, \theta) = w_u \sum_{x_i \in X} \varphi_i^u(x_i, \theta_u) + w_l \sum_{(x_i, x_j) \in \epsilon} \varphi_{ij}^L(x_i, x_j, \theta_l) + w_g \sum_{(x_i, x_g) \in \epsilon_g} \varphi_{ig}^G(x_i, x_g, \theta_g) \quad (2.3)$$

onde a porção $\varphi_{ig}^G(x_i, x_g, \theta_g)$ representa as conexões com um nodo global que, tendo esti-

mado seu conjunto de parâmetros θ_g , indica as configurações mais prováveis entre todas as porções da imagem. Com tal modelo, as relações de contexto global, como o conjunto típico de objetos possíveis em cena, podem ser incorporadas.

Por fim, [Lucchi et al. \(2011\)](#) critica o funcionamento do CRF, comparando a utilização do CRF com pontos de interesse globais ([BOIX et al., 2012](#)), concluindo que não há ganho significativo. Além disso, o contexto local adicionado pelo CRF a tende apenas melhorar a suavização da classificação local. Ou seja, dado uma pequena região da imagem, a variabilidade de classes é reduzida.

2.4 Sumário

Este capítulo apresentou os conceitos de utilização de contexto para visão computacional. Apresentou-se em quais níveis o contexto pode ser utilizado, sendo eles globais ou locais. Também foi apresentado quais níveis de interação o contexto podem se dar, sendo eles no nível de pixels, regiões ou objetos.

Nesse âmbito, formalizou-se a classificação utilizando contexto, considerando o nível de interação baseado em regiões, no caso, *superpixels*. Em seguida, foram apresentadas formas de utilizar o contexto. Primeiramente foram apresentadas formas de aplicar o contexto diretamente no classificador. Em seguida, foi apresentado o uso de modelos probabilísticos gráficos para a aplicação de contexto.

Por fim, alguns dos principais trabalhos em modelos probabilísticos gráficos foram apresentados.

3 Classificação Baseada em Contexto utilizando Geoestatística

Tendo em vista as limitações existentes no CRF (LUCCHI et al., 2011) e o conhecimento obtido através de estudos em Geoestatística, neste capítulo, busca-se propor um novo método para adição de contexto baseado em Geoestatística.

Tal abordagem agrega duas áreas com aplicações distintas mas conceitos semelhantes. No campo da modelagem geológica, uma abordagem baseada em geostatística primeiramente busca modelar a variabilidade espacial de uma determinada medida com o objetivo de interpolar este comportamento para áreas desconhecidas (CARLE; FOGG, 1996). Esta estratégia é bastante utilizada para aplicações como modelagem de reservatórios em campos de extração óleo (BEATTIE; MILLS; MAYO, 1998) ou mapeamento geológico (PURKIS; VLASWINKEL; GRACIAS, 2012).

Porém, neste trabalho busca-se também mostrar que este conceito se aplica para o caso de adição de contexto classificação de imagens. A abordagem apresentada é capaz de assegurar a suavização de estruturas espaciais de maneira similar que o CRF, porém, o método proposto, também considera correlações em longa distância entre rótulos.

3.1 Visão Geral da Proposta

A Equação 3.1 apresenta a adição do contexto espacial utilizando Geoestatística:

$$P(L) = w_u P_u(L|\theta_u) + w_l P_l(L|W), \quad (3.1)$$

sabe-se que a probabilidade, P de um dado conjunto de rótulos L , é dada pela soma ponderada, pelos pesos w_u e w_l , das probabilidades unária $P_u(L|\theta_u)$, da parte segmentada, e a probabilidade do contexto local $P_l(L|W)$. A matriz W está associada ao peso atribuído aos *superpixels* da vizinhança. θ_u está associado aos parâmetros de usados para obtenção da distribuição unária.

Como nível de interação, parte-se de uma segmentação baseada em *superpixels* em um grid retangular e uniforme onde cada superpixel tem aproximadamente o mesmo tamanho. O sistema de classificação proposto neste capítulo, é mostrado na Figura 10. Nesta proposta, o nível de interação, em uma imagem segmentada ocorre através de *Turbopixels* (LEVINSHTEIN et al., 2009).

No nível unário $P_u(L|\theta_u)$, somente as informações visuais descritas de um único

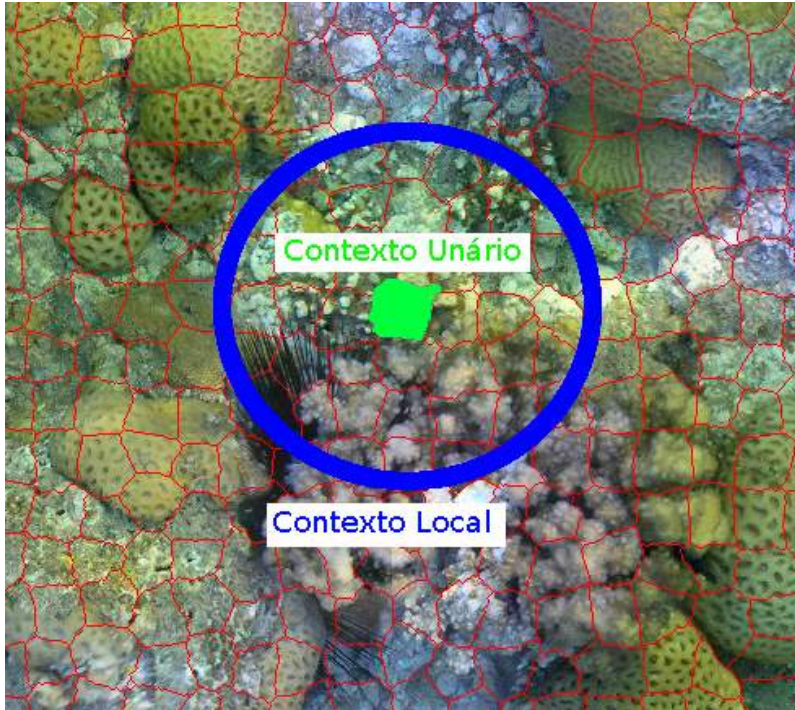


Figura 9 – Divisão especificada em dois níveis de classificação. O nível unário $P_u(L|\theta_u)$ onde somente a informação do superpixel segmentado é utilizada, apresentado em verde. E o nível local $P_l(L|W)$, onde um determinado contexto local é incluído na classificação, representado pelo círculo azul.

superpixel são relevantes para a classificação do mesmo. Na seção 3.2, mostra-se a computação do nível unário e a necessidade do mesmo de produzir uma distribuição confiável.

No nível local $P_l(L|W)$, se considera as conexões de uma determinada área onde medidas estatísticas são utilizadas (Círculo Azul Fig. 10). O nível $P_l(L|W)$ é apresentado na seção 3.4.

3.2 Nível Unário $P_u(L|\theta_u)$

Para o sistema proposto, o foco da classificação unária é obter uma distribuição de probabilidades previa para cada superpixel. Essa probabilidade a priori vai ser usada para inferir a vizinhança.

Em um dado *superpixel* o qual pode ser classificado como *um* dentre um conjunto de rótulos $L = l1, l2..ln$ busca-se obter uma saída $P_u(L|\theta_u) = P_u(l1|\theta_u), P_u(l2|\theta_u)...P_u(ln|\theta_u)$ onde $\sum P_u(L|\theta_u) = 1$. Onde θ_u é conjunto de parâmetros usados para se ter essa saída.

Para se chegar em tal resultado, a geração do nível unário é dividida em duas etapas. A primeira corresponde ao treinamento da função de discriminação $f(x)$, do classificador. No caso, é feito o treinamento de um *kernel* linear para uma *Support Vector Machine (SVM)*. A segunda etapa é a determinação das curvas de confiança, que corres-

ponde ao grau de certeza da classificação. Ou seja, o grau de certeza C_{l_i} é dado como uma função treinada, e é usado para gerar $P_u(L|\theta_u)$ (ABFALG et al., 2007).

3.2.1 Classificador

Como classificador, foi utilizado uma *Support Vector Machine* (*SVM*) com um *kernel* linear. A ideia do algoritmo é encontrar uma função de hiperplano $f_{l_i}(x)$, para cada classe l_i que separa linearmente um conjunto de dados previamente rotulados, porém maximizando uma determinada margem. Trata-se de uma abordagem de classificação supervisionada.

Para dada aplicação é necessário que o classificador produza uma distância de um objeto à borda de classe mais próxima a borda entre classes (ABFALG et al., 2007). Tal resultado é obtido diretamente pelo *SVM* dado que sua função de hiperplano já maximiza a margem entre classes. A saída numérica do *SVM* já é própria para se ter um certo grau de confiança do classificador.

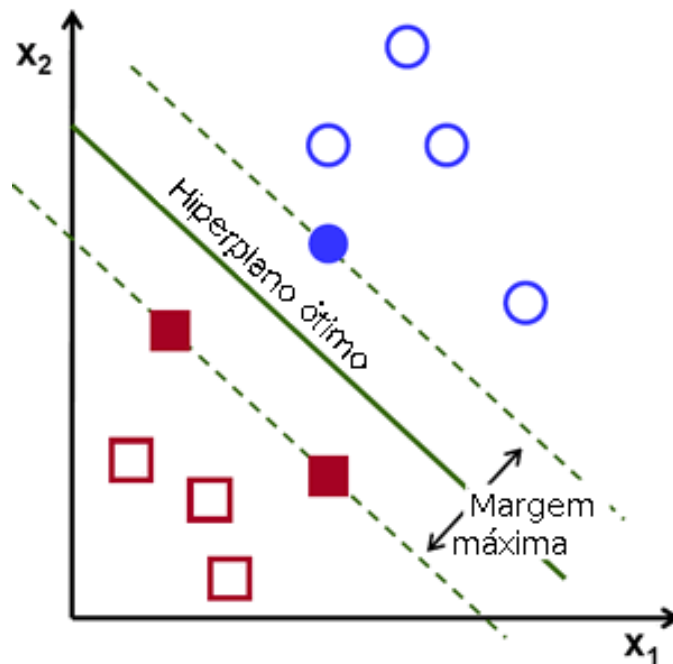


Figura 10 – Figura do separador linear obtido pelo treinamento do SVM. Dado os conjuntos de dados já rotulados (Azuis e Vermelhos), o SVM determina o separador de máxima margem. A saída numérica do SVM já é própria para se ter um certo grau de confiança do classificador.

3.2.2 Treinando Curvas de Confiança

Somente as distâncias de saída do SVM (CRISTIANINI; SHAW-TAYLOR, 2000), não estabelecem diretamente o grau de confiança de uma classificação (ABFALG et al., 2007). Ou seja, uma distância de valor número 5 para o SVM pode ser para alguns casos

uma saída confiável, para outros não. É necessário treinar para quais distâncias existe uma grande probabilidade da predição ser correta. Isso depende do *dataset* que foi utilizado, da classe (rótulo) e também do classificador.

O estudo apresentado por Platt (PLATT, 1999) demonstra que comportamento sigmoidal pode modelar a distribuição de probabilidade do SVM.

Sendo assim, para cada classe l_i é feito um ajuste de uma função sigmoidal C_{l_i} . Esta função C_{l_i} não mais retorna uma distância e sim uma *probabilidade* de uma determinada entrada na função $f_{l_i}(x)$ ser correta.

A Figura 11 apresenta a saída esperada para a curva de confiança treinada C_{l_i} , para um dado rótulo l_i . Dado um conjunto de validação em que já se possui os retornos $f_{l_i}(x)$, o treinamento é feito da seguinte forma: Obtém-se os pontos em azul os quais indicam a porcentagem de acertos que se tem utilizando apenas *as parcelas* de dados as quais seu retorno vindo do classificador ($f_{l_i}(x)$) é de *no máximo* o que é mostrado no eixo x . Por exemplo, na Figura 11, para um conjunto de dados com uma distância do classificador ($f_{l_i}(x)$) de até dois, tem-se uma porcentagem de acerto de 70%. Dado esse conjunto de pontos, o algoritmo de *Levenberg-Marquardt* (MARQUARDT, 1963) de otimização é utilizado para encontrar os coeficientes, α_k e β_k da função sigmoidal:

$$C_{l_i} = N_{l_i} \frac{0.5}{1 + \exp(\alpha_{l_i} * f_{l_i}(x) + \beta_{l_i})} + 0.5 \quad (3.2)$$

onde N_{l_i} é uma constante de normalização para a classe l_i . A função inicia de 0.5 pois é a probabilidade inicial de uma classe ser acertada aleatoriamente. A abordagem proposta contrasta com a proposta apresentada por ABfalg et al. (2007) por propor uma curva de confiança para cada classe. Isso também leva em conta as diferenças existentes em cada classe. θ_u está relacionado aos parâmetros α_k e β_k de $P_u(L|\theta_u)$.

3.3 Distribuição de Probabilidades

Normalizando o grau de confiança para a distância com relação a todas as classes, se obtém a distribuição de probabilidades para um determinado objeto. Sendo assim, um classificador não mais produz somente uma saída, mas também uma chance de cada item de um conjunto de dados a pertencer a cada uma das classes. A Figura 12 mostra a saída da classificação de um exemplo calculado. Sendo que cada barra representa a chance do objeto pertencer a tal classe.

Observa-se na Figura 12 que a saída do classificador mostra a classe mais provável mas existe uma incerteza significativa para uma segunda classe ser a correta.

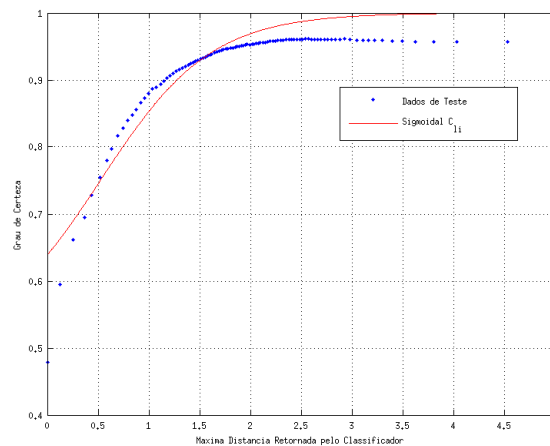


Figura 11 – Gráfico mostrando a probabilidade de acerto em função da máxima confiança retornada pelo classificador para um conjunto de dados. Em vermelho tem-se a função C_{1i} treinada a partir do conjunto de dados em azul.

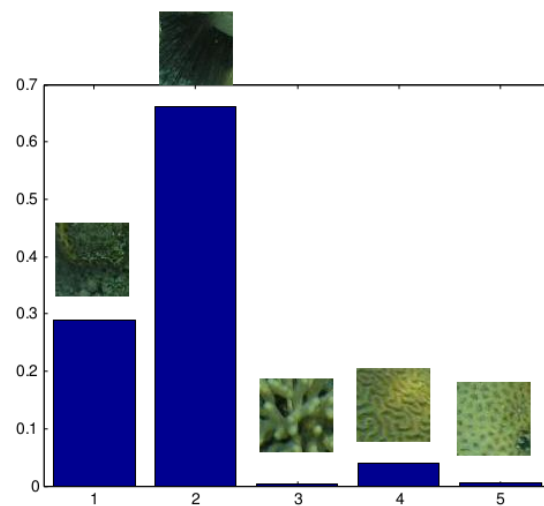


Figura 12 – Histograma mostrando a distribuição de probabilidades de saída de um classificador. Para o caso, a segunda classe, é a que obteve maior probabilidade, porém existe uma certa incerteza com relação a primeira classe.

3.4 Nível Local $P_l(L|W)$

A Figura 13 mostra o conjunto de passos para calcular os parcelas locais utilizando o contexto baseado em Geoestatística.

A primeira parte do método (lado direito da Figura 13) é estimar e modelar a incerteza existente nos conjunto de dados os quais se quer adicionar o contexto. Essa modelagem se da através da estimativa da matriz de probabilidade de transição (T) entre os possíveis rótulos presentes no conjunto de dados. Isso é feito em uma etapa de treinamento *offline* do método A estimativa é feita em duas partes que são combinadas: analisando

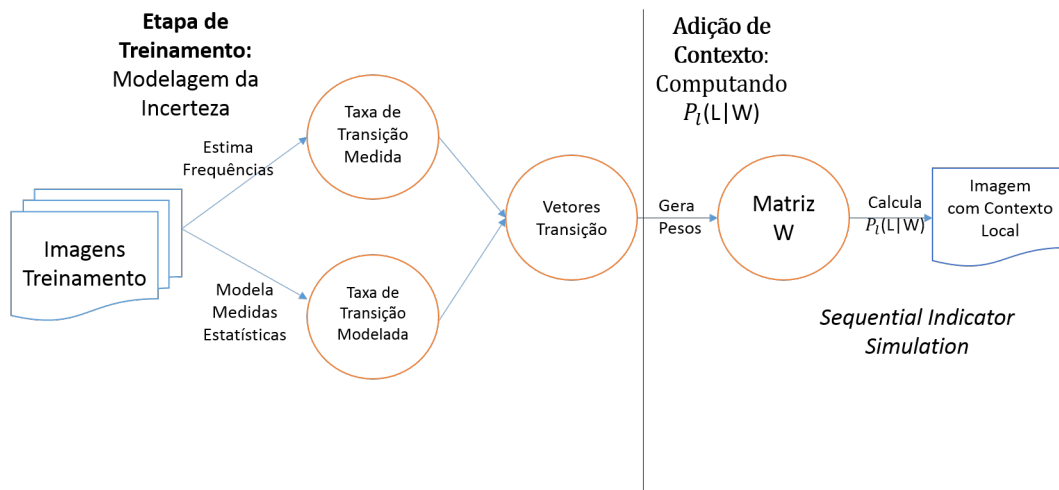


Figura 13 – Diagrama geral da adição de contexto local utilizando Geoestatística. Primeiramente é medida a variabilidade entre as classes no contexto espacial. Tanto diretamente através das frequências de transição na imagem (taxa de transição medida), quanto através da inferência de propriedades estatísticas vindas da imagem (taxa de transição modelada). Em seguida são calculados os vetores de transição. Na segunda parte os vetores são utilizados para gerar pesos para imagem. Com isso, utilizando os pesos, o sistema *SIS* computa a adição de contexto local para cada superpixel.

as frequências de transições entre as classes de um conjunto de imagens (CARLE et al., 1998) e medindo propriedades estatísticas nos dados como proporções e espessuras.

A segunda parte do método é a geração do contexto local $P_l(L|W)$. Deste modo, para cada *superpixel* que se deseja computar, a matriz de probabilidades de transição é usada para computar os pesos W , para servir como entrada em um sistema de SIS, *Sequential Indicator Simulation* (Indicador de simulação sequencial) (EMERY, 2004) o qual computa $P_l(L|W)$. Cada um dos processos apontados na Figura 13 são detalhados no restante desta seção.

3.4.1 Medindo Transições de Probabilidades

Primeiramente, um sistema de transição de probabilidades baseado em cadeias de *Markov* é medido. Este modelo representa a variabilidade espacial existente juntamente com os dados da imagem.

A estratégia proposta é calcular uma matriz T , onde cada componente é a função $t_{ij}(h\phi)$ a qual modela a probabilidade de uma classe i de transitar para a classe j em uma distância h considerando a direção ϕ . É necessário obter tal medida para cada par de classes i e j presente no conjunto de dados.

Neste método, assume-se que os dados são isomórficos. Portanto, para uma dada transição, todas as direções são consideradas como idênticas. Não obstante, a abordagem pode ser utilizada em casos não isomórficos, considerando duas ou mais direções, cada

uma com sua própria matriz de probabilidade de transições.

Foi assumido que a transição de probabilidades tem um comportamento exponencial, como proposto por (CARLE et al., 1998). As equação 3.3 mostra como calcular a transição de probabilidades entre classes diferentes e também para a mesma classe (auto transição).

$$t_{ij} = \begin{cases} t_{ij} = e^{r_{ij}h\phi} + p_i & \text{se } [i = j] \\ t_{ij} = p_j - e^{-r_{ij}h\phi} & \text{se } [i \neq j] \end{cases} \quad (3.3)$$

A função de transição na Equação 3.3 também depende da distância h e da probabilidade a priori da classe p_j . Cada fator r_{ij} é um componente da matriz R . Essa matriz é a taxa de transição entre as classes, para k classes, R pode ser calculada como:

$$R = \begin{bmatrix} r_{11} & \dots & r_{1i} & \dots & r_{1k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{j1} & \cdot & r_{ji} & \cdot & r_{jk} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{k1} & \dots & r_{ki} & \dots & r_{kk} \end{bmatrix} \quad (3.4)$$

Cada elemento da matriz representa a taxa na qual ocorre a transição, sendo assim:

$$r_{ij} = \frac{\partial t_{ij}(0)}{\partial h} \quad (3.5)$$

A matriz T não pode ser diretamente calculada a partir dos dados (AGTERBERG, 1988). Para tal, primeiramente, é necessario estimar a matriz R . Carle et al. (1998), propõem obter o calculo de R através da multiplicação elemento a elemento das medidas da correlação entre as configurações espaciais diretas (R_{mes}) e da medida de conceitos estatísticos (R_{mod}) extraídos das imagens:

$$R = R_{mes} * R_{mod}. \quad (3.6)$$

Esta foi a técnica adotada neste trabalho para calcular R , porém com algumas modificações. Na seção a seguir, mostra-se o processo para calcular as matrizes R_{mes} e R_{mod} .

3.4.1.1 Taxa de Transição Medida R_{mes}

R_{mes} é calculado medindo a frequência de transição cumulativa da matriz F . Para computar F , foi somado o número de vezes que cada classe transita para cada outra classe. Foi considerado um conjunto de direções e um conjunto de distâncias h . Tal treinamento é

feito em um conjunto de imagens já previamente classificadas. Este processo é apresentado na Figura 14. Foi utilizado um *kernel* de janela deslizante que iterativamente se desloca ao longo de toda a imagem.

Para cada posição do *kernel*, foram contadas todas as transições que o rótulo do ponto central do *kernel* faz. Isto é feito para diversas distâncias, o qual é representado pelos quadrados coloridos da Fig. 14. Ao final, cada linha de F é normalizada.

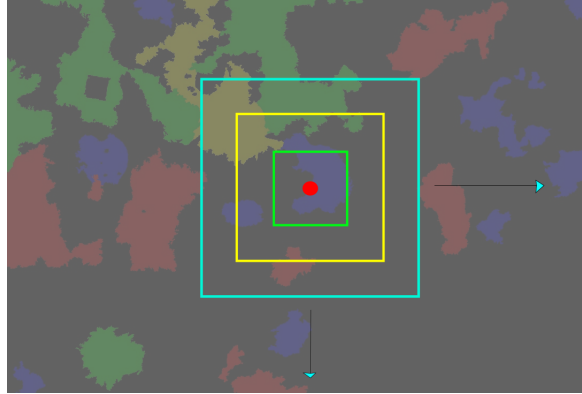


Figura 14 – Medida feita do número de transições que uma classe faz para cada outra para múltiplas distâncias. Foi utilizada um kernel móvel e foram contadas as transições desde o centro (ponto vermelho) para todas as direções (representado pelos quadrados)

A equação 3.7 mostra um exemplo da matriz F feitas para um "dataset" exemplo.

$$F = \begin{bmatrix} 0.762 & 0.007 & 0.037 & 0.060 & 0.132 \\ 0.639 & 0.066 & 0.048 & 0.061 & 0.184 \\ 0.607 & 0.010 & 0.214 & 0.052 & 0.115 \\ 0.594 & 0.006 & 0.032 & 0.261 & 0.104 \\ 0.642 & 0.010 & 0.036 & 0.054 & 0.255 \end{bmatrix} \quad (3.7)$$

A matriz estimada F é afetada pelas incertezas nas premissas de probabilidade assumidas. Por exemplo, a premissa do isomorfismo assumida não é perfeitamente verdadeira. Para reduzir o efeito das incertezas e encontrar um padrão na representação, uma análise de autovetores e autovalores é aplicada na Eq. 3.7 (CARLE; FOGG, 1996).

A partir disso, é possível computar R_{mes} aplicando a equação 3.8.

$$R_{mes} = \sum_{k=1}^L \theta_k Z_k \quad (3.8)$$

onde o termo θ_k de $k = 1, \dots, L$ denota os autovalores de F e Z_k denota os componentes espectrais das matrizes desde a análise de auto-vetores. Foi calculado Z_k como mostrado

na Eq. 3.9.

$$Z_k = \frac{\prod_{m \neq k} (\theta_k I - F)}{\prod_{m \neq k} (\theta_m - \theta_k)} \quad k = 1, \dots, L \quad (3.9)$$

Esta computação consiste em uma medida inicial que congrega as tendências de verossimilhança espacial entre as classes. Contudo, esta medida ainda contém muita imprecisão para ser usada como entrada para a simulação. A medida pode ser ainda mais estabilizada adicionando a computação de R_{mod} assim como mostrado na Eq. 3.6.

3.4.1.2 Cálculo da Matriz R_{mod}

Computa-se R_{mod} utilizando estatísticas extraídas dos dados, como: proporções das classes, comprimentos médios das classes e as tendências de justaposição.

A proporção de uma classe l_i é a probabilidade a priori desta classe aparecer. Em outras palavras, a proporção é a chance de selecionar uma parcela da classe l_i aleatoriamente da imagem classificada (CARLE; FOGG, 1996).

O comprimento médio é calculado pela quantidade média de pixels contínuos de uma certa classe ao longo de uma determinada direção. Como assume-se isomorfismo nos dados, esta direção é arbitrária. Considerando em termos de transição de probabilidades, o comprimento médio $L_{h\phi}$ é a taxa de decaimento da curva de transição da função $t_{ii}(h\phi)$ na direção ϕ . O comprimento médio é mostrado na equação 3.10.

$$-\frac{\partial t_{ii}(0)}{\partial h} = \frac{1}{L_{h,\phi}} \quad (3.10)$$

Isso é análogo a taxa de uma classe transitar para si mesma, como mostrado na Eq. 3.11 (CARLE; FOGG, 1996).

$$\tilde{r}_{ii} = -\frac{1}{L_{h,\phi}} \quad (3.11)$$

O conceito de tendência de justaposição modela as probabilidades de uma classe transitar fora de si mesmo e depois em outra dado uma distância. Considerando r_{ii} como a taxa que a uma certa classe transita para si mesma, r_{ij} depende das proporções de j como mostrado na Eq. 3.12.

$$\tilde{r}_{j,k}(h\phi) = \frac{p_k}{L_j\phi(1-p_j)} \quad (3.12)$$

Para o caso de um *dataset* de cinco classes, a matrix R_{mod} tem a seguinte estrutura:

$$R_{mod} = \begin{bmatrix} \frac{1}{L_{11}} & \tilde{r}_{11} & \tilde{r}_{13} & \tilde{r}_{14} & \tilde{r}_{15} \\ \tilde{r}_{21} & \frac{1}{L_{22}} & \tilde{r}_{23} & \tilde{r}_{24} & \tilde{r}_{25} \\ \tilde{r}_{31} & \tilde{r}_{32} & \frac{1}{L_{33}} & \tilde{r}_{34} & \tilde{r}_{35} \\ \tilde{r}_{41} & \tilde{r}_{42} & \tilde{r}_{43} & \frac{1}{L_{44}} & \tilde{r}_{45} \\ \tilde{r}_{51} & \tilde{r}_{52} & \tilde{r}_{53} & \tilde{r}_{54} & \frac{1}{L_{55}} \end{bmatrix} \quad (3.13)$$

Finalmente, usando R_{mes} e R_{mod} é possível computar a Eq 3.6. Na Figura 15 são mostradas as transições de probabilidades calculadas para um *dataset*. Os gráficos da Figura 15 foram computados aplicando a Eq. 3.3 variando a distância h . Como um exemplo, pode-se perceber que o comprimento médio da classe *background* é bem alto. Isso acontece por que o seu decaimento exponencial é muito baixo.

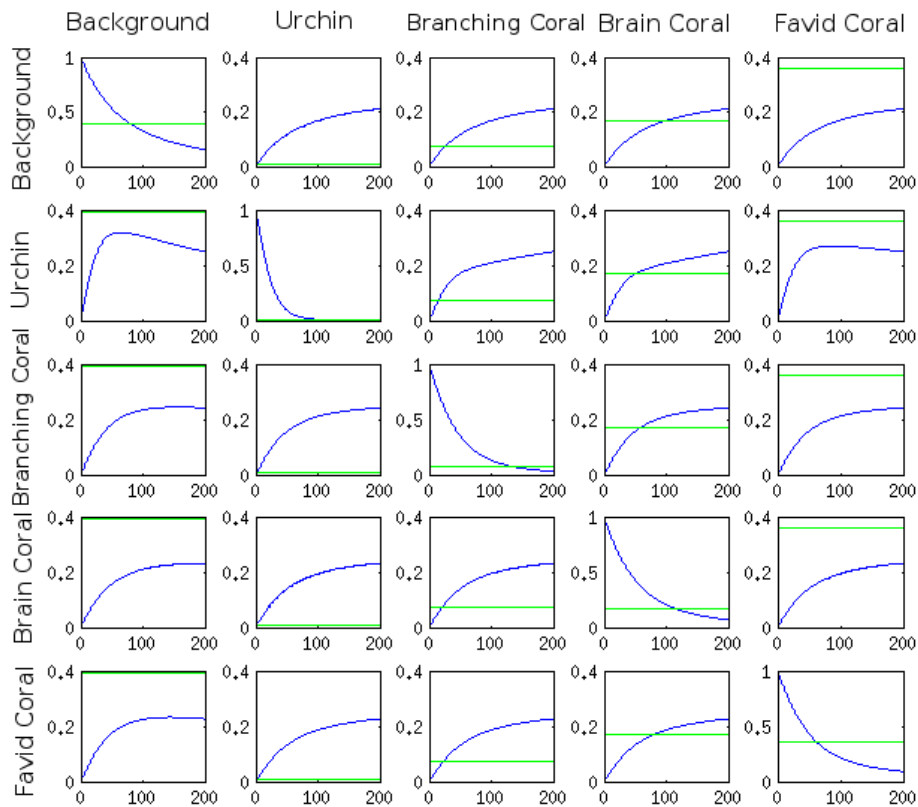


Figura 15 – A transição de probabilidade modelada para um determinado *dataset*. O eixo y apresenta a distância em pixels. As linhas verdes mostram as proporções para cada classe. Pode-se observar uma certa tendência na classe *Urchin* em transitar para categoria de *background*. Ainda, percebe-se que a classe de *background* tem um grande comprimento médio, dado que sua taxa de decaimento é bastante alta.

3.4.2 Sequential Indicator Simulation

Dado que a matriz de transição T já foi calculada para um *dataset*, o algoritmo *Sequential Indicator Simulation* (SIS) tenta simular o $P_l(L|W)$ de um *superpixel* com base em sua vizinhança espacial. Para simular os fatores locais de um certo *superpixel* x_0 , um certo número N de posições aleatórias amostradas x_α são computados em torno da região em um raio r . Cada uma das posições amostradas vai contribuir para o computar o fator local, sendo que a contribuição é feita de forma a minimizar a variância desta vizinhança com respeito ao modelo.

Com isso, a probabilidade relacionada com o contexto espacial para cada classe k em uma certa parcela x_0 é computada como:

$$P_l(x_0 = k | P_u(X_\alpha)) = P_u(x_0 = k) \sum_{\alpha=1}^N \sum_{j=1}^K P_u(X_\alpha = j) w_{jk,\alpha} \quad (3.14)$$

onde P_u é a probabilidade a priori (unário) de uma região, sendo $P_u(X_0 = k)$ o *superpixel* em questão e $P_u(X_\alpha = j)$, os amostrados. $w_{jk,\alpha}$ é o peso da posição α para a classe j transitar para a classe k . Ou seja a probabilidade local de um *superpixel* é função da distribuição do mesmo ($P_u(X_0 = k)$) e o quanto cada posição amostrada contribui para este *superpixel* ($\sum_{\alpha=1}^N \sum_{j=1}^L P_u(X_\alpha = j)$). O fator é controlado pelo peso $w_{jk,\alpha}$.

Os pesos para cada posição amostrada formam o conjunto de matrizes W_N e são calculados resolvendo o sistema linear da Eq. 3.15 :

$$\begin{bmatrix} T(x_1 - x_1) & \dots & T(x_N - x_1) \\ \cdot & \cdot & \cdot \\ T(x_1 - x_N) & \dots & T(x_N - x_N) \end{bmatrix} \begin{bmatrix} W_1 \\ \cdot \\ W_N \end{bmatrix} = \begin{bmatrix} T(x_0 - X_1) \\ \cdot \\ T(x_0 - X_N) \end{bmatrix} \quad (3.15)$$

onde:

$$W_i = \begin{bmatrix} w_{11,\alpha} & \dots & w_{1L,\alpha} \\ \cdot & \cdot & \cdot \\ w_{L1,\alpha} & \dots & w_{LL,\alpha} \end{bmatrix} \quad (3.16)$$

A Figura 16 mostra o exemplo de um *superpixel* arbitrário e sua respectiva região amostrada, para a computação do potencial local. Para tal região a Eq. 3.15 será aplicada de forma a encontrar o peso para cada uma das posições. O peso encontrado é o que tornaria a região o mais homogênea possível.

3.4.3 Computando o Potencial Final $P(L)$

Depois de obter uma saída da curva de confiança para cada *superpixel*, primeiramente se busca os *superpixeis* com uma saída bem alta de confiança. Foi decido computar

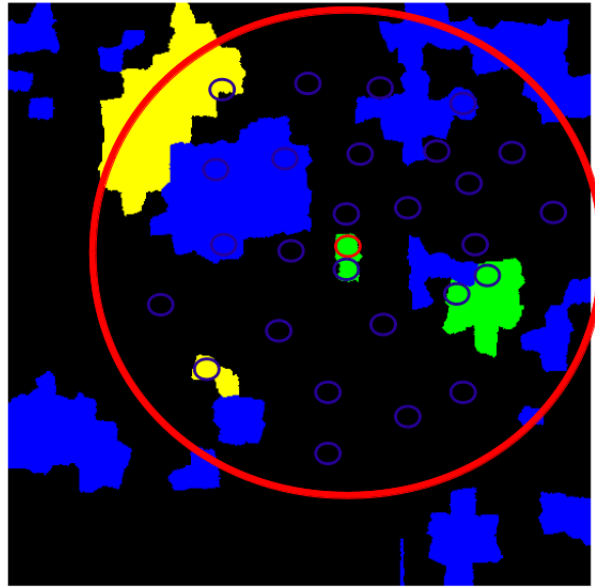


Figura 16 – Exemplo de uma vizinhança sendo considerada para um superpixel (apontado em vermelho). Um raio r é considerado e N pontos são amostrados nessa vizinhança (em azul). Cada um dos pontos amostrados irá influenciar no potencial do *superpixel* apontado em vermelho.

o potencial local apenas para superpixeis onde a confiança está abaixo de um limiar t . O limiar é selecionado como a confiança máxima, dado pelo conjunto de validação.

O processo do *SIS* é repetido para cada *superpixel* presente na imagem em ordem aleatória e os pesos já são atualizados. Isso garante que a correlação entre a própria vizinhança seja considerada.

Os pesos são obtidos diretamente pela Eq. 3.15. Dois parâmetros devem ser escolhidos para este método, o número de amostras N e o raio r onde vai ser feita a amostragem. Experimentos preliminares mostraram que não existe vantagem pratica em usar mais de 25 amostras. Também, o raio r passa e se tornar irrelevante a partir de uma certa distância, dado que às transições de probabilidade tendem a ser iguais as proporções no limite.

3.5 Geoestatística e CRF

Tanto as abordagem de CRF, quanto de Geoestatística (GS) tentam minimizar uma função que tenta impor uma certa homogeneidade espacial. Ou seja, *superpixeis* próximos tendem a ser da mesma classe. A diferença é que o modelo de Geoestatística é baseado em uma amostragem o que torna o problema da inferência mais simples. O modelo de GS é também análogo a um CRF densamente conectado (KOLTUN; VLADLEN, 2011), mas com amostragens mais esparsas.

A abordagem de GS pode ser vista como uma representação mais esparsa do CRF porém, com medidas estatísticas mais ricas. Não obstante, a computação da matriz de pesos W para a Eq. 3.14 pode ser considerado como a minimização de uma função de energia, usando uma soma ponderada.

Como uma forma de comparar ambos os métodos, a Figura 17 mostra o modelo GS como um modelo gráfico probabilístico. O vértice central, em verde claro, é o caso atual sendo calculado. Os vértices em verde escuro são aqueles amostrados. Cada vertice em verde escuro contribui para a distribuição do vertice central dependendo das probabilidades de transição estimadas da Fig. 15. Em vermelho são representados os fatores unários de cada quadrado em azul é a contribuição desses mesmos (fatores locais).

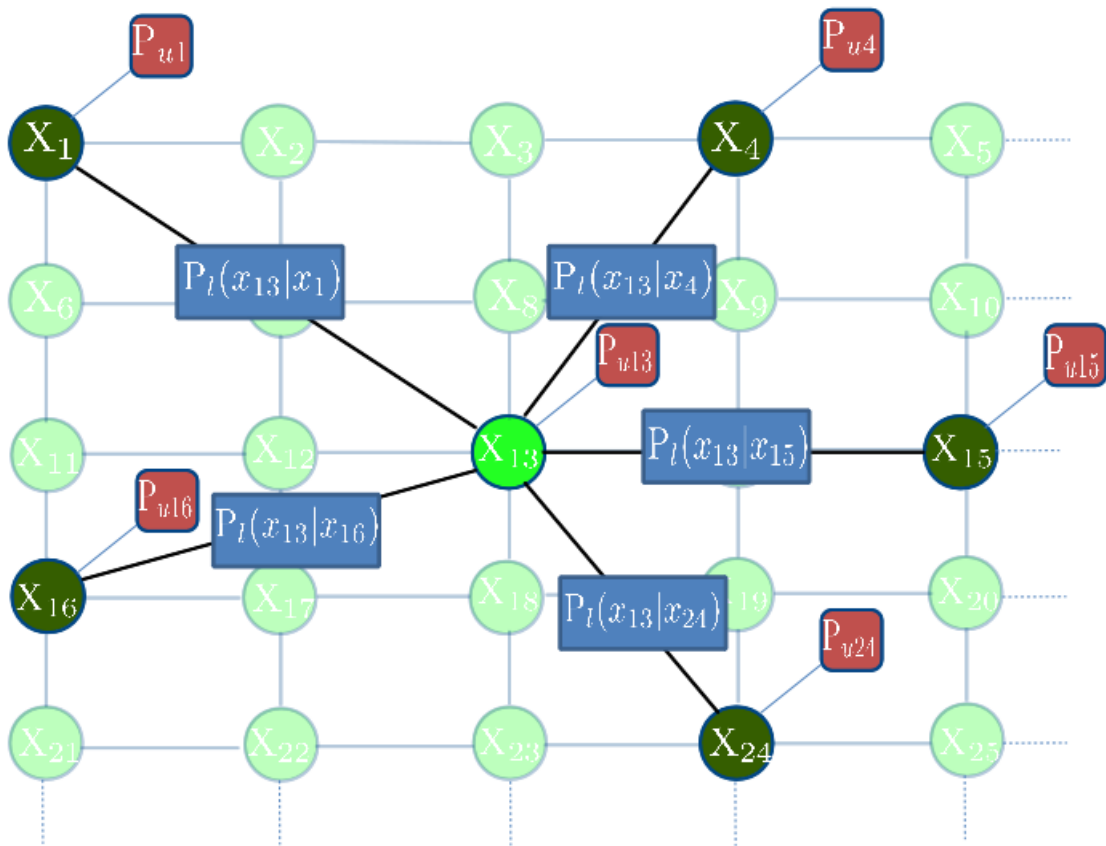


Figura 17 – Representação gráfica do modelo de Geoestatística (GS). Os fatores locais são representados em azul e usam a estatística de probabilidade de transição computada pela Eq. 3.3. Diferentemente do que no modelo da Fig. 7, vizinhos de diferentes distâncias também contribuem para calcular a distribuição de cada posição.

3.6 Sumário

Neste Capítulo apresentou-se um novo método para adição de contexto na classificação. O método foi inspirado nas técnicas de modelagem da variabilidade espacial usada em Geoestatística.

Foi feita, por fim, uma comparação do método proposto com o CRF. Acredita-se que o método apresentado neste capítulo tende a se comportar melhor que o *CRF* quando existem menos dados de treinamento, e os mesmos dados não possuem padrões bem definidos, como no caso do ambiente subaquático. Isso pode ser atingido visto que o método proposto estima padrões de forma para as classes. Sendo assim as relações de correlação espacial, são também estimadas com base em um modelo para as classes. O método proposto será testado e avaliado no Capítulo 6.

4 Classificação de Imagens do Assolho Oceânico

Neste Capítulo é apresentado o domínio de aplicação no qual será aplicado o método de adição de contexto proposto no Capítulo 3.

Como apresentado na introdução, o conhecimento sobre as espécies presentes no fundo do mar, especialmente os recifes de corais, é de fundamental importância para os especialistas na área.

Ao se fazer monitoramento do assolho oceânico, assim como para o caso do sensoriamento remoto, é interessante ser capaz de rotular automaticamente cada pixel das imagens e assim ser capaz de medir propriedades relevantes. O objetivo desta classificação é fazer os chamados *mapas temáticos*. Tais mapas são a representação final de um mapa classificado de uma imagem, feito de forma visualmente interpretável. Mapas temáticos agregam grandes conjuntos de imagens em mosaicos representando áreas de grande extensão.

Considerando o ambiente subaquático, suas propriedades fotométricas demandam um tratamento especial para contornar a degradação da imagem. Esses desafios próprios do meio não são comumente endereçados na literatura. Tais propriedades causam problemas como bordas confusas entre objetos, variação na qualidade da imagem, etc.

Neste capítulo primeiramente é formalizada as propriedades do meio subaquático, o que será útil também para capítulos posteriores. Depois, é apresentada uma visão geral dos principais sistemas utilizados para classificação de mosaicos do assolho oceânico. Entre os sistemas apresentados, um em especial será detalhado, o qual será utilizado como um estudo de caso para adição de contexto.

4.1 Propriedades de Imagens Subaquáticas

De forma a obter imagens capturadas em ambiente subaquático com uma melhor qualidade visual, é fundamental o entendimento de sua formação, levando em conta os aspectos específicos que ocorrem no meio subaquático.

Um modelo de formação de imagens busca descrever os caminhos pelos quais a luz passa, desde a fonte até a sua captura, onde é formada a imagem. A Figura 18 ilustra este processo de propagação. Em meios participativos, a irradiação, ou seja, a quantidade de energia luminosa em um pixel, pode ser obtida pelo somatório de três componentes as quais chegam por caminhos distintos. A componente direta, a qual contém a luz sem es-

palhamento que veio diretamente do objeto. Muitas vezes, informações que vinham de um único ponto são espalhadas entre seus pontos vizinhos causando um efeito de borramento na imagem. Este fenômeno é chamado espalhamento dianteiro (*forward scattering*), representado pela componente *forward scattering*. O *forward-scattering* faz com que as informações visuais da cena fiquem espalhadas, causando um efeito de borramento.

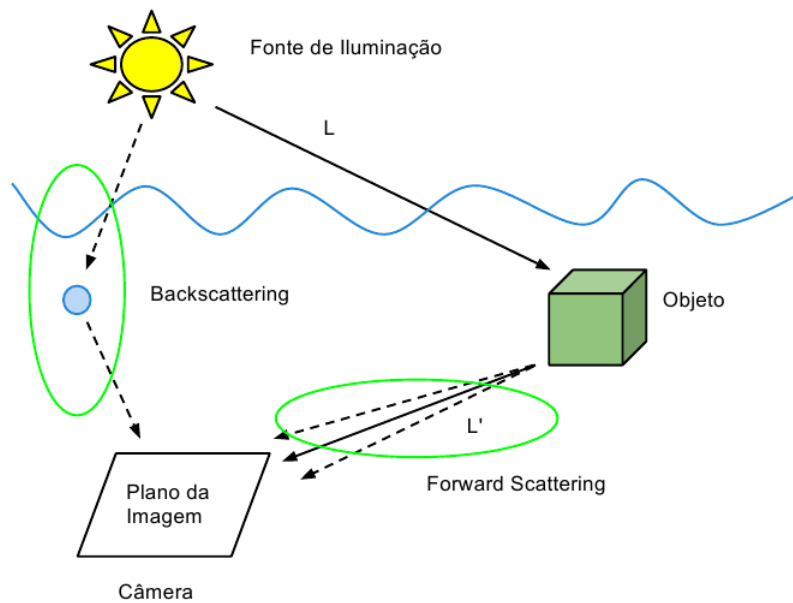


Figura 18 – Três trajetórias da luz até o plano da imagem. O componente direto, contendo a informação direta da cena. O *forward-scattering*, contendo informação da cena espalhada. Por fim, o *backscattering* contendo informações de fora da cena.

Por último, tem-se a componente de *backscattering*, a qual luz chega no plano da imagem a partir de um ponto que não faz parte da cena observada. Isso acontece devido à alguma partícula flutuante que desvia a trajetória da luz para o plano da imagem. O *backscattering* se comporta tal como um ruído aditivo.

Para calcular cada uma das componentes, algumas simplificações devem ser consideradas. Tais simplificações visam tornar o modelo mais simples e tratável computacionalmente, ressaltando somente alguns aspectos principais na formação da imagem.

Primeiramente, se assume o objeto como tendo sua reflectividade uniforme. Assume-se uma iluminação completa e uniforme da cena. Por fim, pode-se descartar os parâmetros da câmera e considerar a captura da luz como sendo também uniforme.

Normalmente o efeito causado pelo *forward-scattering* tende ser desprezado, por contribuir com uma participação menor que o *backscattering* na formação da imagem (TREIBITZ; SCHECHNER, 2006).

A descrição final do modelo é dada pela equação de Koschmieder (KOSCHMIEDER, 1924), bastante utilizada para a propagação da luz na névoa. Sendo assim, a

formação de um ponto (x, y) na imagem é dado por:

$$I(x, y) = J(x, y) e^{-cz(x,y)} + B_{\infty}(1 - e^{-cz(x,y)}), \quad (4.1)$$

Sendo $J(x, y)$ a imagem sem degradação e $z(x, y)$ uma função da distância para cada ponto na imagem. Essa equação pode ser interpretada da seguinte forma: quanto mais distante estiver o objeto maior será o componente *backscattering*, menos da cena real irá existir na imagem.

Sabe-se que, devido as propriedades do meio subaquático, existe uma diferença significativa entre a absorção e espalhamento dos comprimentos de onda (DUNTLEY, 1963). Desta forma, o modelo pode ser estendido de forma incorporar diferentes comprimentos de onda. A equação 4.1 modela a quantidade de luminosidade capturada relativa a um determinado pixel. Porém é possível adequá-la para diferentes comprimentos de onda, ou no caso do padrão RGB de representação, dividi-la em três canais conforme a equação 4.2,

$$I^{\lambda}(x, y) = J^{\lambda}(x, y) e^{-c_{\lambda}z(x,y)} + B_{\infty}^{\lambda}(1 - e^{-c_{\lambda}z(x,y)}), \quad \lambda \in \{R, G, B\} \quad (4.2)$$

A Figura 19 apresenta uma típica imagem com alto nível de *turbidez*. *Turbidez* é uma propriedade comum no meio aquático que esta relacionada com a quantidade de luz que é absorvida ou espalhada ao invés de ser transmitida em uma linha reta (OMAR; MATJAFRI, 2009).



Figura 19 – Imagem de exemplo para as degradações do ambiente subaquático. É possível ver que existe uma variação conforme a distância e uma perda significativa da informação de cor.

É interessante observar que a degradação não afeta uniformemente a imagem. Existem níveis de degradação mais altos de acordo com a distância. Além disso, o comprimento de onda vermelho tende a se perder rapidamente, tendo a cor verde nesse caso como predominante.

Por fim, vale notar que, fenômenos adicionais também acontecem. Um exemplo é o efeito da "neve submarina", a qual causa aparecimento de pequenos pontos brancos na imagem. Por estes e outros fatos é relevante constatar que o meio subaquático já tem uma alta presença de ruído (BAZEILLE et al., 2006).

4.2 Classificação Autônoma de Imagens do fundo Oceânico

A Figura 20 mostra uma adaptação do que é usado pela maioria dos *frameworks* em visão computacional para criação de mapas temáticos de mosaicos em ambientes subaquáticos (SHIHAVUDDIN et al., 2013).



Figura 20 – A sequência utilizada para classificação de imagens em meio subaquático.

Para classificar os objetos de uma imagem é necessário passar por diversas etapas. A seguir são listadas as etapas apresentando algumas das técnicas usadas na literatura:

- **Pré-processamento:** etapa fundamental em ambientes subaquáticos. Normalmente é onde correções de cor (PIZARRO et al., 2008) e contraste (JOHNSON-ROBERSON; KUMAR; WILLAMS, 2007) são aplicadas para atenuar a degradação e ressaltar aspectos importantes das imagens subaquáticas.
- **Segmentação:** Nesta etapa a imagem é super-segmentada em regiões com propriedades similares. Tal etapa pode ser evitada, para o caso onde ocorre uma seleção manual do que ser classificado.
- **Extração de Descritores:** é onde as características relevantes para cada segmento são extraídas e representadas. Diversas abordagens são utilizadas, um exemplo seria o uso de descritores locais e *bag-of-words* por Pizarro, Eustice e Singh (2004). Os descritores de textura e cor são bastante utilizados no meio subaquático (BEIJBOM et al., 2012) (STOKES; DEANE, 2009), (MARCOS; SORIANO; SALOMA, 2005).
- **Classificação:** é onde se realiza o treinamento do classificador e classificação para os testes. Diversos classificadores são utilizados como o SVM (PIZARRO; EUSTICE; SINGH, 2004) ou o LDA (MARCOS; SORIANO; SALOMA, 2005).
- **Pós-processamento:** é onde informações adicionais são utilizadas para refinar o resultado da classificação. Em (SHIHAVUDDIN et al., 2013) é feito um simples sistema de votação para verificar a consistência da vizinhança. No caso, até onde se sabe, não ocorreram outras aplicações de técnicas mais elaboradas para adição de contexto.

Nesta seção é especificado em detalhe cada etapa apresentada elucidando o que foi utilizado por [Shihavuddin et al. \(2013\)](#) para geração de mapas temáticos. Tal método foi escolhido como base para aplicação de técnicas para adição de contexto. O mesmo foi escolhido devido a alta taxa de acerto na classificação quando comparados com diversos métodos do estado da arte ([SHIHAVUDDIN et al., 2013](#)). Os algoritmos usados em cada uma das etapas da Figura 20 são elucidados a seguir.

4.2.1 Pré-Processamento

O processo de pré-processamento almeja deixar a imagem o mais próxima possível da cena em qual a mesma foi capturada. Isso é feito tanto no escopo radiométrico quanto geométrico. Ou seja, o objectivo é tornar, as estruturas geométricas, seu brilho e cor o mais próximos possível da cena ([GONZALEZ; WOODS, 2006](#)).

Para lidar com o processamento embaixo d'água, primeiramente, precisa-se considerar todos os princípios básicos de propagação da luz nesse meio os quais foram colocados na Seção 4.1. ([SCHETTINI; CORCHS, 2010](#))

Seguindo a ideia de que a qualidade visual subjetiva é importante, pode-se melhorar a qualidade de imagens subaquáticas utilizando técnicas que abordam diretamente os efeitos degradantes apontados. Esta seção apresenta as alternativas existentes para corrigir cada um dos tipos de degradação.

4.2.1.1 Contraste

Observa-se pela Equação 4.1 que o processo de degradação da imagem em ambiente subaquático não é uniforme ao longo da imagem. O mesmo depende da distância de cada ponto a câmera.

Nesse contexto, o [Shihavuddin et al. \(2013\)](#) faz o uso do *CLAHE* (*Contrast Limited Adaptive Histogram Equalization*) ([ZUIDERVELD, 1994](#)) para correção de contraste. Tal método faz uma construção de histograma diferente para cada segmento da imagem e aplica uma equalização de histograma somente nesse segmento. Além disso, o método coloca um limite de intensidade máxima, redistribuindo todas as intensidades que ficam acima deste limite.

4.2.1.2 Correção de Cor

Como mostrado na seção 4.1, existe uma não uniformidade na absorção de cada comprimento de onda no ambiente subaquático. Isso causa que boa parte da informação cromática da cena seja perdida.

De forma a obter cores mais próximas de realidade existe a necessidade de estimar tais diferenças de absorção. Uma das formas de resolver isso é considerar que é possí-

vel obter as diferenças de absorção considerando essas diferenças como uma questão de estimativa da fonte de luz. Colocado de tal forma, o problema, se torna basicamente a aplicação de algoritmos de balanceamento de branco, os quais podem ser uma simples normalização.

O método de (SHIHAVUDDIN et al., 2013) aplicou a premissa de que o ponto de maior intensidade da imagem foi causado por reflexão perfeita. Desta forma a iluminação pode ser estimada achando o ponto de maior intensidade da imagem. Sendo assim, para tornar a cor da imagem balanceada, o ganho para cada pixel pode ser dado como:

$$\begin{aligned} R_{ganho} &= 1/R_{max} \\ G_{ganho} &= 1/G_{max} \\ B_{ganho} &= 1/B_{max} \end{aligned} \quad (4.3)$$

4.2.2 Segmentação

Diversos desafios em classificação de imagens colocam o desafio atual como classificar os objetos pixel a pixel (FULKERSON; VEDALDI; SOATTO, 2009).

O caso da aplicação em sensoriamento remoto, claramente se beneficia deste fato, onde cada pixel da imagem é relevante. A questão é que, devido ao custo computacional, e ao fato que somente um pixel não possui grande significado semântico para efetuar a classificação e extrair os descritores, muitas vezes a abordagem de usar segmentos da imagem, ajuda a melhorar a consistência.

Existe e a tendência de muitos autores fazer uma pré-segmentação, a qual aparentemente não esta relacionada com a classificação final. Porém, tal segmentação ajuda a garantir que cada parte da imagem sendo classificada tenha uma homogeneidade local. Tal segmentação é chamada de segmentação em *superpixels*.

Para o caso da abordagem de Shihavuddin et al. (2013), os *superpixels* são utilizados como estrutura de interação. A imagem é definida como um conjunto de *superpixels* a serem classificados.

Diversos algoritmos existem para a criação de superpixels. Porém, Shihavuddin et al. (2013) selecionou aquele que tende a manter uma estrutura o mais regular possível. No caso foi utilizado os *Turbopixels* (LEVINSHTEIN et al., 2009).

4.2.3 Descritores

Para descrever a imagem foi utilizado majoritariamente descritores de textura. Textura pode ser definida como a variação dos dados visuais em escalas menores que a escala observada (PETROU; GARCÍA-SEVILLA, 2006).

O assoalho submarino é tipicamente texturizado. Observou-se diversos bancos de dados de corais, e outras estruturas encontradas no meio, e percebe-se que existem sempre oscilações na estrutura dos objetos em diferentes escalas. Tal fenômeno caracteriza a existência da textura. Além da tendência existente na literatura em usar textura (SHIHAVUDDIN et al., 2013).

Com isso em vista, (SHIHAVUDDIN et al., 2013) utiliza três descritores como descritores de texturas: *Gabor Filter*, *Grey Level Co-occurrence Matrix (GLCM)* e *Completed Local Binary Pattern (CLBP)*.

Os *Gabor Filters* são um grupo de Wavelets 2D que tomam forma de uma gaussiana 2D modulada no espaço 2D (PORTER; CANAGARAJAH, 1997). Basicamente são uma representação da variação de frequência em um segmento da imagem.

O *GLCM* (HARALICK; SHANMUGAM; DINSTEN, 1973), utiliza a representação de padrões de variações espaciais dos segmentos da imagem em uma matriz, que representa a variação de intensidade dos pixels em diferentes ângulos e distâncias. Diversos indicadores são computados a partir dessas matrizes como a média de variações ou a entropia.

O *CLBP* (GUO; ZHANG, 2010), é um descritor de textura invariante a rotação o qual retrata, principalmente, a variação de sinais de um pixel central para com pixels ao redor em uma determinada posição.

A utilização de cor é complexa dado a perda de cor não uniforme entre os comprimentos de onda como mostrado na Seção 4.1. Porém ainda é possível utilizar um descritor de cor que possui propriedades importantes como robustez a variações fotométricas causadas por sombras, sombreamento e também mudanças geométricas como escala e alteração de ponto de vista. (SHIHAVUDDIN et al., 2013) utiliza o trabalho de (WEIJER; SCHMID, 2006) que aproximou tais propriedades.

Ao final, ao utilizar múltiplos descritores, se tem uma representação da imagem com uma grande quantidade de dimensões e muitas vezes com um padrão pouco evidente. Para resolver isso é aplicado normalizações e modificações nos descritores. Por exemplo, os descritores podem ser manipulados de forma que os mesmos sejam o mais próximos a se tornaram linearmente separáveis. Essa modificação é fundamental para se melhorar a qualidade da classificação. Por fim, os descritores são normalizados de forma a que todos os descritores estejam numa escala compatível.

4.2.4 Treinamento e Classificação

O treinamento foi feito utilizando três classificadores distintos de forma mutuamente exclusiva. Foram utilizados o *Support Vector Machine (SVM)*, o *K-nearest-neighbors* e o *PDWMD* proposto por (STOKES; DEANE, 2009). Cada um destes classi-

ficadores foi utilizado dependendo das características dos dados.

Dado que o aprendizado foi feito, novos dados podem ser classificados. É feito um mapa temático baseado na segmentação em *superpixels*. Sendo que cada *superpixel* é classificado individualmente.

4.3 Conclusões

Neste capítulo apresentou-se o cenário onde vai ser feito o estudo desta dissertação. Também se apresentou alguns métodos os quais já fizeram classificação de imagens do bentos.

No capítulo 6 serão apresentados os resultados de aplicação do método de (SHIHAVUDDIN et al., 2013) e será feito o estudo sobre a incorporação de contexto para esse método.

5 Testes e Resultados 1: Detecção de Pontos de Interesse em Ambiente Subaquático

Uma das principais contribuições desta dissertação foi a criação de um experimento para analisar e compreender o comportamento dos detectores de pontos de interesse quando utilizados em ambiente subaquático.

Como apresentado no Capítulo 1, diversos detectores foram desenvolvidos para serem invariantes a uma serie de fenômenos. A ideia é que o mesmo ponto de interesse possa ser encontrado independentemente de diversas circunstâncias da cena.

Porém, existem fenômenos adicionais que atuam sobre a cena no ambiente subaquático os quais devem ser considerados. Quando a luz se propaga neste meio, ela é absorvida e espalhada pelos diferentes coeficientes de refração encontrados nas partículas presentes no meio. Isso espalha a informação capturada e cria o efeito de "enevoado" na imagem. Tais fenômenos foram descritos mais detalhadamente no Capítulo 4, Seção 4.1.

Um estudo feito por *Garcia e Gracias* (GARCIA; GRACIAS, 2011), comparou os detectores de pontos de interesse mais populares na literatura. Eles encontraram que estruturas do tipo blob, obtidos por métodos baseados em *Hessian* (BEAUDET, 1978), por exemplo, são melhores detectadas tanto para o caso de métodos invariantes a escala como os de única escala. A justificativa é que a turbidez da água tende a suavizar quinas e borrar regiões definidas, fazendo com que métodos como *Harris* (HARRIS; STEPHENS, 1988) ou *Harris-Laplace* (MIKOLAJCZYK; SCHMID, 2004) sejam menos propícios para o ambiente. Entretanto, eles avaliaram somente algumas estruturas em uma única cena. É do interesse desta dissertação melhorar este estudo. Neste contexto, alguns principais objetivos são buscados.

Foi proposto um novo *dataset* no qual é possível utilizar diferentes estruturas submarinas obtidas através da impressão de fotos subaquáticas. Estas estruturas foram refotografadas dentro de um tanque de água onde imagens com a degradação controlada foram produzidas. Isso é uma melhoria a tentativas anteriores em termos de diversidade de elementos visuais. Considerando que a degradação causada por imagens com baixa e alta turbidez não é linear, uma contribuição é dividir a análise em diferentes intervalos de turbidez.

Foram testados detectores de pontos de interesse, considerando diferentes abordagens, com respeito a sua robustez a degradação causada pela turbidez. Foi focado investigar o problema de que detectores invariantes a escala tendem a ter baixa performance (GARCIA; GRACIAS, 2011). Isto é feito através da análise de diferentes espaços de es-

cala. Finalmente, foi indicado o melhor detector invariante para imagens subaquáticas como sendo o *DoG* (LOWE, 2004).

Este Capítulo está organizado da seguinte maneira. A seção 5.1 apresenta a descrição completa do experimento a ser realizado. Tal seção mostra todos os detalhes da experimentação necessários para que o mesmo seja bem sucedido. Também explica todas as considerações feitas para se ter dados aceitáveis. Por fim, a Seção 5.3 mostra os resultados obtidos para tal experimento, e apresenta uma discussão sobre os resultados encontrados.

5.1 Descrição do experimento

Na literatura, poucos são os trabalhos que analisam o comportamento dos detectores de pontos de interesse em ambiente subaquático. Nesta seção, descreve-se todo o processo de realização do experimento para que ele seja completamente reproduzível.

Neste experimento foram capturadas diversas imagens em uma cena onde a única modificação entra as cenas é a degradação causada pela turbidez. O objetivo fundamental do experimento é tentar obter o máximo de isolamento desta degradação possível. Para tal, a câmera utilizada deve estar estática e a iluminação deve ser controlada.

5.1.1 Cena Montada

Construí-se uma cena onde as imagens foram colocadas. A Figura 21 mostra a especificação da cena.

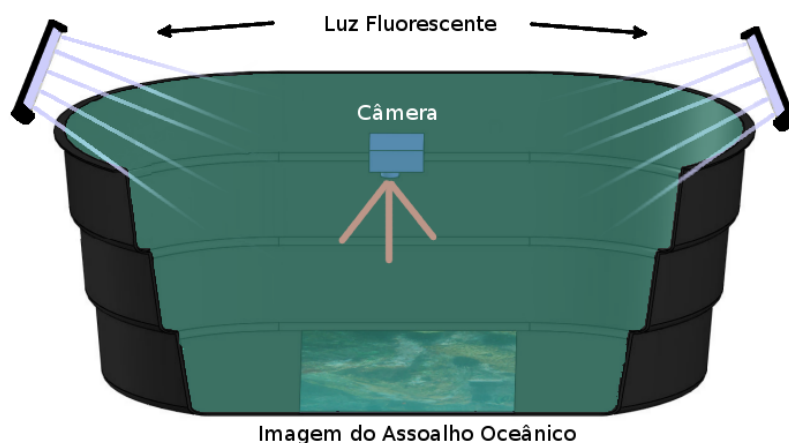


Figura 21 – A cena criada para avaliar os algoritmos de avaliação de features. Ela é composta por lâmpadas fluorescentes e uma camera fotografando fotos impressas do assoalho do oceano.

Na cena montada existe uma fotografia a ser capturada por uma câmera posicionada a uma distância perpendicular de 0.58cm . A fotografia está posicionada em uma

caixa de água de mil litros. Duas luminárias usando lâmpadas fluorescentes brancas foram posicionadas perto do tanque.

Três fotografias diferentes foram utilizadas, representando o fundo do mar capturado nas Bahamas em condições próximas ao ideal de turbidez (ZVULONI et al., 2009). As diferentes cenas contêm os mais variados tipos de textura que podem ser encontradas no ambiente subaquático e também objetos feitos pelo homem. As fotografias foram impressas usando um "ploter" a laser usando uma mídia de vinil adesivo fosco e a prova d'água. Cada fotografia foi impressa num tamanho de 91cm X 60 cm e possuem 4928x3264 pixels de resolução. O diferencial desta deste *dataset* é que ele contém verdadeiras estruturas do assoalho oceano, e ainda algumas estruturas feitas pelo homem, sendo o seu principal problema, a perda de resolução devido a impressão e a refotografia. Isso cria uma perda de resolução de 20 *pixels/mm*² para 4 *pixels/mm*² e adição de algumas pequenas imperfeições devido a erros de impressão.

A Figura mostra as imagens que foram impressas, nomea-se cada uma das imagens como P_1 , P_2 e P_3 .

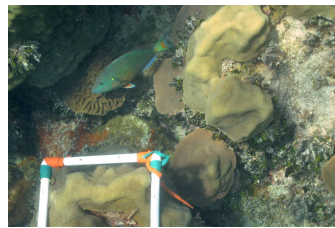
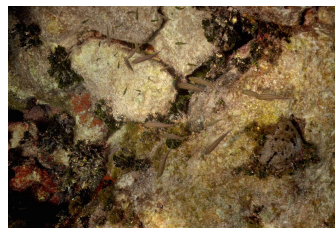
(a) P_1 (b) P_2 (c) P_3

Figura 22 – As imagens utilizadas no teste. As três imagens foram capturadas nas Bahamas em condições de turbidez próximas do ideal em uma resolução de 4928x3264 pixels

A câmera utilizada para a captura foi uma Gopro Hero 3 Black edition. Cada imagem foi capturada em uma resolução de 12 mega pixels(3000x4000).

Imagem (T_i)	Quantidade de Leite Integral	Leite Adicionado
T1	5 ml	5 ml
T2	10 ml	5 ml
T3	15 ml	5 ml
T4	20 ml	5 ml
T5	25 ml	5 ml
T6	30 ml	5 ml
T7	36 ml	6 ml
T8	42 ml	6 ml
T9	50 ml	8 ml
T10	58 ml	8 ml
T11	66 ml	8 ml
T12	74 ml	8 ml
T13	82 ml	8 ml
T14	90 ml	8 ml
T15	100 ml	10 ml
T16	110 ml	10 ml
T17	120 ml	10 ml
T18	130 ml	10 ml
T19	190 ml	60 ml

Tabela 1 – A quantidade de leite adicionada para cada nível de turbidez simulado.

5.1.2 Procedimento

Foi decidido simular principalmente o efeito do fenômeno de *backscattering*. Sabe-se que os motivos que levam a degradação de uma imagem capturada em meio subaquático são complexos (DUNTLEY, 1963). Porém, neste experimento tentou-se isolar o principal fenômeno que causa a degradação na imagem. Um estudo feito por Narasimhan et al. (2006) mostra que uma solução de água e leite integral apresenta um alto grau de *backscattering*, apontado por alguns como a principal fonte de degradação da imagem (TREIBITZ; SCHECHNER, 2006). Isso é causado pelo maior tamanho das partículas do leite integral que fazem que o ângulo de refração seja maior, aumentando o *backscattering*.

Foi decidido dividir o experimento em 3 ensaios, cada um contendo uma imagem diferente.

Cada ensaio foi capturado com 19 níveis de turbidez diferentes, cada um contendo uma determinada quantidade de leite. Chamou-se cada nível de turbidez de $T_1 \dots T_{19}$. Considera-se T_0 como o nível de turbidez com a imagem limpa. A Tabela 1 mostra os níveis de turbidez e suas respectivas quantidades de leite (Em uma caixa com aproximadamente 1000 litros de água).

Para capturar as imagens, a câmera foi setada para capturar uma foto a cada 10 segundos. Para cada nível de turbidez foi escolhido um grupo de fotos com o menor nível de perturbação. Como explicado no Capítulo 4, Seção 4.1, o meio subaquático é

composto por uma certa quantidade de ruído. Em um ambiente controlado, como o que foi feito é possível, tendo uma seleção de fotos em um mesmo nível de turbidez T_i , reduzir o ruído extraindo mediana entre as imagens iguais (GARCIA; GRACIAS, 2011). Desta forma busca-se reduzir a degradação na imagem por ruídos que podem ter diversas causas como erro no sensor da câmera, partículas bloqueando totalmente a passagem da luz, etc. Tenta-se de certa forma isolar significativamente a degradação por turbidez (IDT), como o principal fenômeno da cena.

A Figura 23 apresenta as imagens geradas pelo experimento. Neste caso fez-se a distinção entre diversos intervalos de turbidez. Na Figura 23 é mostrado um nível de turbidez por intervalo, para cada imagem.

5.2 Avaliando a degradação causada pela turbidez

Medir a quantidade de degradação é fundamental neste experimento de forma a comparar os detectores somente relativo a este fenômeno. A degradação causada pela turbidez é dependente da quantidade de partículas em suspensão na água, e também os tipos de partículas em suspensão. Além disso, a quantidade de iluminação e a maneira como a cena é iluminada é também fundamental para determinação da degradação causada pela turbidez.

Este conceito difere do conceito de turbidez que esta relacionado somente com a quantidade de sedimentos flutuantes (SSC) na água os quais espalham a luz. A degradação causada pela turbidez difere pois ela não esta relacionado somente as partículas presentes na água e sim a degradação que o *SSC* causa na cena, levando em conta os parâmetros da câmera e o volume de água iluminado.

Uma forma de medir a turbidez é usando um turbidímetro nefelômetro, o qual mede a turbidez pela quantidade de luz espalhada ao emitir um feixe de laser numa porção da água.

Esta alternativas não é capaz de estimar a degradação causada pela turbidez, que é também dependente da cena. Com essas considerações, Garcia e Gracias (2011) propuseram a utilização de uma variação *Structural Similarity Index* (WANG et al., 2004), para avaliar a degradação, chamado *Structural Degradation Index* (SDI). Essa abordagem avalia a degradação pela perda de informação estrutural, o que de fato esta relacionado com a turbidez. Porém, a mesma não tenta isolar a medição do fenômeno de absorção e espalhamento como principais causadores da degradação.

Neste trabalho utiliza-se a métrica proposta por (GARCIA; GRACIAS, 2011), porém normalizada em função da imagem completamente turva pelo leite. Tal métrica é capaz de medir a porcentagem de degradação em função da imagem onde teve sua

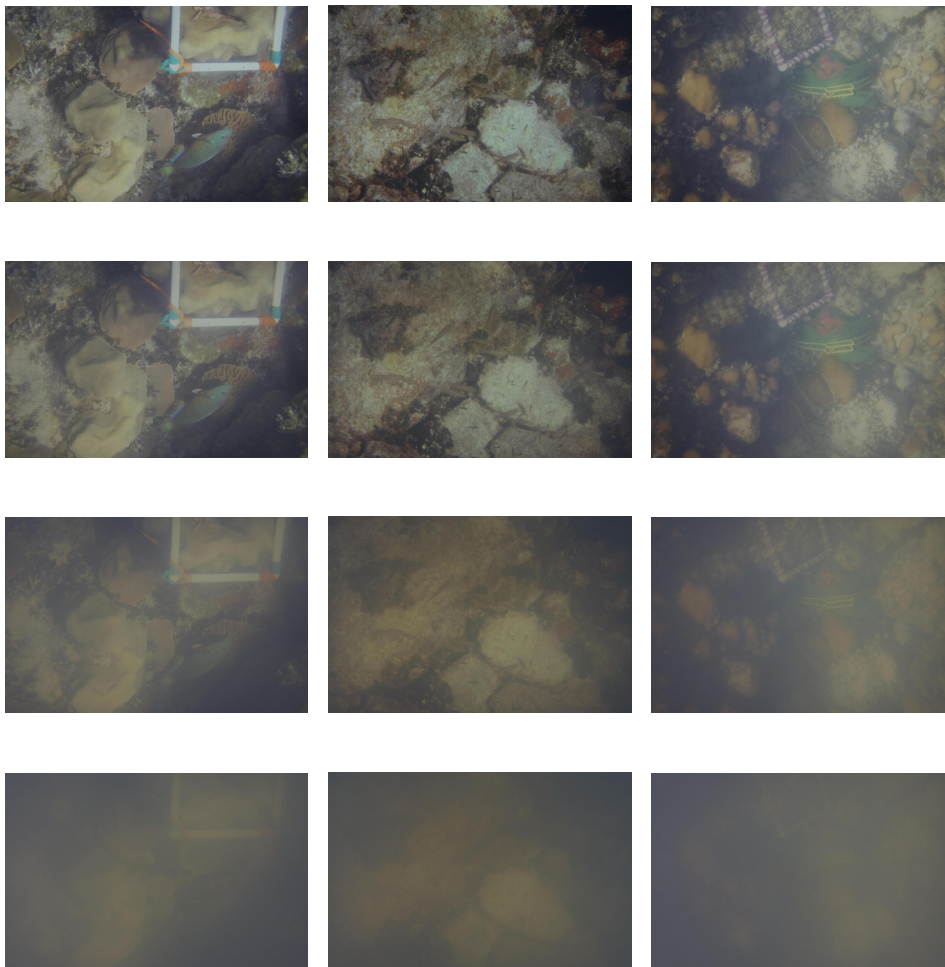


Figura 23 – As imagens capturadas sob diferentes níveis de degradação devido a turbidez, controlado pela adição de leite. Foram fotografadas três fotos impressas diferentes, P_1 (primeira coluna), P_2 (segunda coluna) e P_3 (terceira coluna). Na primeira linha foi mostrada a imagem limpa (sem leite) para cada foto capturada. A segunda linha apresenta o intervalo de *Baixa Turbidez* com por volta de 15ml de leite (T_4). O intervalo de *Média Turbidez* é mostrado na segunda linha e contém por volta de 50 ml de leite (T_{10}). Finalmente, na última (quarta) linha é mostrado o intervalo com *Alta turbidez* tendo por volta de 100 ml de leite (T_{16}). Quantidade de leite setada para uma caixa com 1000 litros de água.

informação visual inicial completamente eliminada. O método é explicado na secção 5.3.1

5.3 Resultados

Nesta seção são mostradas as comparações entre os detectores. Foram comparados os seguintes detectores, previamente definidos no Capítulo 1. Para única escala, *Harris* (HARRIS; STEPHENS, 1988), *Hessian* (BEAUDET, 1978) e *Laplacian* (TUYTELAARS; MIKOLAJCZYK, 2008). Com múltiplas escalas avaliou-se *Fast Hessian* do SURF (BAY et al., 2008) e *Difference of Gaussians* do SIFT (LOWE, 2004). Foram avaliados também

outros detectores com propriedades relevantes. Os três *kernels* baseados em difusão anisotrópica do detector *KAZE* (ALCANTARILLA; BARTOLI; DAVISON, 2012) e o gerado pelo centro e arredores usando estruturas poligonais *CenSurE* (AGRAWAL; KONOLIGE; BLAS, 2008) usando tanto um polígono convexo de seis lados e um polígono estrelado de também seis lados.

5.3.1 Procedimento de Avaliação

Os resultados são avaliados quanto ao critério de repetibilidade descrito em (SCHMID; MOHR; BAUCKHAGE, 2000). Tal critério indica a porcentagem dos pontos de interesse que se repetiram, ou seja, ainda foram encontrados após a aplicação da transformação.

Primeiramente computa-se $N = 1000$ pontos de interesse para cada detector na imagem com a turbidez T_0 e para todos os níveis $T_1 \dots T_{19}$. Os N pontos de interesse selecionados são os N melhores pontos de interesse segundo o critério do detector, no caso *Hessian* ou *Harris*. Na imagem com turbidez T_0 é selecionada cada ponto-chave e é testado se esse ponto é resistente na presença de turbidez. Para esse ponto-chave ser resistente é necessário que o mesmo seja encontrado nas imagens turvas sem sofrer um deslocamento maior que um fator de $e = 5$ pixels. Esse valor é determinado de forma a escolher somente os melhores pontos de interesse. Subsequentemente, para determinar a repetibilidade de um certo detector, o número de pontos-chaves encontrados em cada imagem túrbida são contados. Considerando essa questão, a repetibilidade quanto ao degradação por turbidez (R) é calculada como:

$$R = \frac{N_i}{N_0} \quad (5.1)$$

Onde N_0 é o número de pontos de interesse na imagem limpa (capturada em T_0) e N_i é a imagem com a degradação estimada.

Para medir a degradação causada pela turbidez, foi usado uma versão diferente do SDI (WANG et al., 2004) (GARCIA; GRACIAS, 2011). O índice SDI não responde com os mesmos valores para as mesmas quantidades de turbidez. Por esta razão, foi utilizado uma versão normalizada do SDI. Considerando a imagem T_{19} como sendo totalmente degradada, pode-se medir o SDI como uma porcentagem da degradação máxima, o que facilita a comparação:

$$NSDI_i = SDI_i / SDI_N \quad (5.2)$$

Onde SDI_N é o índice de degradação da imagem T_{19} .

5.3.2 Comparação

A Figura 24, mostra os gráficos com os valores de repetibilidade para as três fotos impressas (P_1, P_2, P_3) testando múltiplos detectores. No eixo x é mostrado o índice $NSDI$ e a quantidade de leite adicionada.

Da Figura 24, são mostrados as análises para três intervalos diferentes de degradação causada por turbidez baseado no $NSDI$. Desde 0 a 0.25 de $NSDI$ foi considerado como um ambiente de *Baixa Turbidez* (Fig. 23 segunda linha). Nestes casos a maioria da informação estrutural é mantida e o *backscattering* é mínimo. No intervalo de 0.25 até 0.75 foi considerado como imagens de parte de um intervalo de *Média Turbidez* (Fig. 23 terceira linha). Nestes níveis, a informação estrutural é parcialmente mantidas, mas as bordas passam a ser mal definidas. Ao final, desde 0.75 até 1, em *Alta Turbidez* (Fig. 23 quarta linha), quase nenhuma informação estrutural é mantida. Nestes níveis, os detectores podem somente fazer uso de algumas poucas pistas visuais que ainda resistiram a turbidez.

Para todos os intervalos de turbidez, é possível separar claramente os detectores analisados em quatro grupos.

Os detectores baseados em única escala (Azul Fig. 24) obtiveram os melhores resultados em todos os intervalos de turbidez. Comparado com outras comparações de detectores de pontos de interesse (GIL et al., 2010) (CRISTINACCE; COOTES, 2006), a superioridade dos detectores *não* invariantes a escala em comparação a aqueles que *são* invariantes a escala, em situações onde a escala não varia, é mais expressiva. O detector *Harris* foi melhor para o caso de P_1 (Fig. 24a) até um nível médio de turbidez. Após isso o mesmo teve um decaimento maior, quando as estruturas começaram a se perder.

O detector baseado em espaços de escala com difusão anisotrópica, obteve os piores resultados (Vermelho Fig. 24). Isso é o oposto do que é mostrado em cenas fora d'água (ALCANTARILLA; BARTOLI; DAVISON, 2012). O algoritmo *KAZE* necessita calcular as respostas das bordas antes de obter o espaço de escala. Isso é mais difícil em ambientes subaquáticos devido a suas propriedades naturais. Porém, no intervalo de *Baixa Turbidez*, *KAZE* foi capaz de obter uma repetibilidade maior que o *FastHessian* e o *DoG*, dado que as bordas ainda estão bem definidas. Para o caso de *Média Turbidez*, a taxa de acerto cai rapidamente, chegando a zero em *Alta Turbidez*.

Os melhores resultados para níveis de media e alta turbidez, foram, de fato, obtidos pelo detector *DoG* (LOWE, 2004). O detector *Fast Hessian*, possui uma aproximação mais brusca do espaço de escala a qual tende a produzir artefatos. Por isso, tratou-se do pior resultado dentre os analisados.

CenSurPoly e o *CenSurStar* apresentaram resultados similares ao *Kaze* tendo piores resultados para níveis mais altos de turbidez.

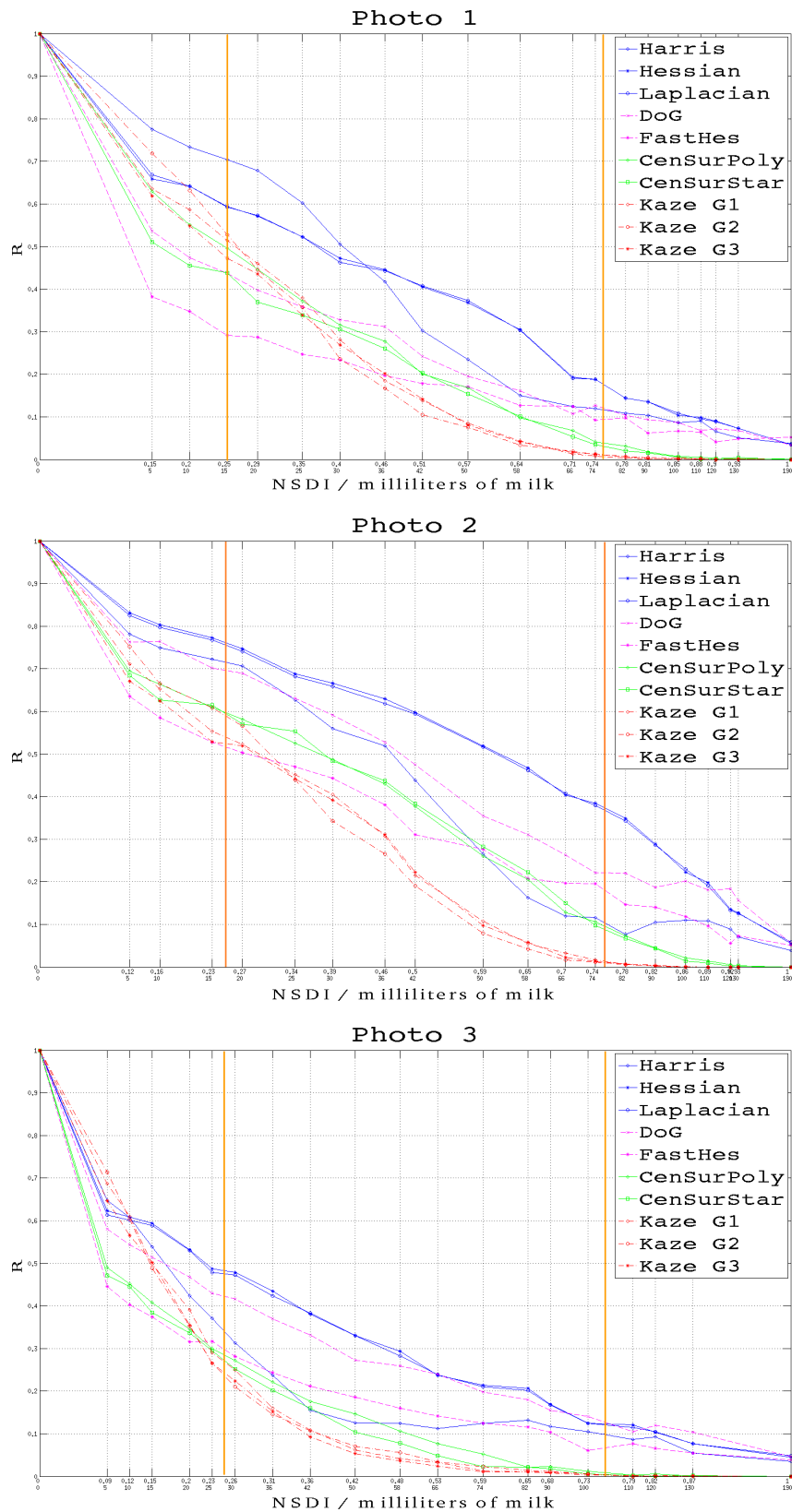


Figura 24 – Repetibilidade (Taxa de Acerto) contra o índice de degradação estrutural normalizado (*NSDI*). As linhas em laranja indicam os intervalos de degradação. *Baixa Turbidez* 0 até 0.25; *Média Turbidez*, 0.25 até 0.75, e *Alta Turbidez* de 0.75 até 1.

A Figura 25 mostra a comparação de um determinado nível de espaço de escala gerado por um *kernel* gaussiano, um *kernel* baseado em caixas (*FastHessian*), um *kernel* baseado em polígonos estrelados (*CenSurE*) e um gerado pelo filtro anisotrópico (*KAZE*) g_2 da Eq. 1.12. Tais *kernels* são aplicados em múltiplos níveis de turbidez, sendo que cada linha da figura apresenta um nível de turbidez diferente.

É possível perceber que a informação estrutural se mantém mais para o polígono estrelado. O que justifica o seu estudo, principalmente para o caso de maior turbidez. Já o *KAZE* também possui um comportamento interessante, porém muito da informação tende a se perder com a turbidez para um mesmo nível de escala.

Como mostrado no Capítulo 4, Seção 4.1, existe um comportamento de borramento regido por um certo fenômeno. É possível que funções, como as utilizadas pelo *CenSurE* e o *KAZE*, as quais tendem a não seguir o comportamento do borramento causado pelas propriedades do meio subaquático, tendam a manter as estruturas geométricas, e, ao encontrar pontos que possuem máximo sobre escala, encontrem regiões em que ainda existe informação visual provida pela imagem.

5.4 Conclusões finais

Este capítulo apresentou a avaliação a invariância a degradação em ambientes subaquáticos para detectores de pontos de interesse mais utilizados na literatura. Foi proposto um novo dataset, completamente aberto, usando fotos impressas reais as quais tinham uma quantidade controlada de turbidez.

Foi concluído que para, imagens subaquáticas, métodos de única escala tem uma repetibilidade consideravelmente melhor que abordagens de multipla escala. Entre *Harris* (HARRIS; STEPHENS, 1988) e *Hessian* (BEAUDET, 1978), *Hessian* obteve resultados melhores principalmente para níveis mais altos de turbidez e em imagens onde há pouca informação estrutural.

Considerando múltipla escala, foram avaliados novos detectores os quais não usam os espaços de escala Gaussianos. Foi proposto que nestes espaços diferentes, como os center surround ou os baseados em difusão anisotrópica, a difusão não acontece com a mesma estrutura que o fenômeno de degradação da turbidez, assim então produzindo melhores resultados, em alguns níveis de turbidez.

Os melhores resultados para múltipla escala foram obtidos pelo *DoG* (LOWE, 2004) Também mostrou-se que o *KAZE* (ALCANTARILLA; BARTOLI; DAVISON, 2012) apresenta resultados relevantes mas tende a perder precisão em níveis mais altos de turbidez.

Finalmente, a avaliação proposta mostra que um espaço de escala não Gaussiano

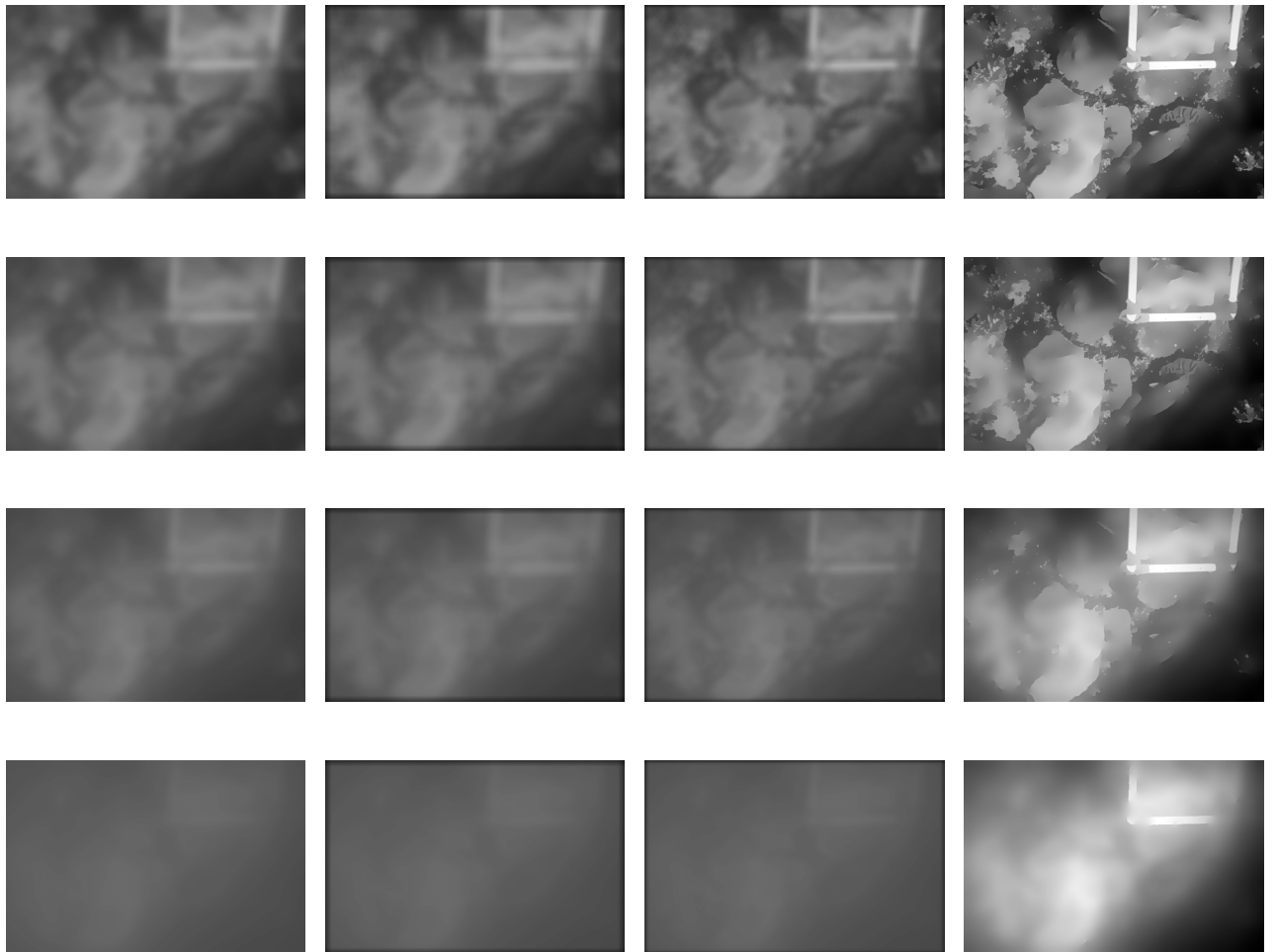


Figura 25 – Comparação entre a geração de *um* nível do *kernel* do espaço de escala usado por quatro detectores diferentes. O *kernel* foi aplicado em níveis de turbidez diferentes para a imagem P_1 . Sendo que a primeira linha é a imagem limpa (T_0), a segunda linha é uma imagem com baixo nível de degradação (T_4), a terceira linha apresenta uma imagem com médio nível de degradação (T_{10}), a quarta linha apresenta imagens do nível de degradação alto (T_{16}). Para cada caso é mostrado o resultado de filtro equivalente a a aproximadamente um *kernel* gaussiano de $\sigma = 59.0$. Primeira Coluna: Gaussiano puro. Segunda Coluna: Borramento aproximado em caixas. Terceira Coluna: Difusão utilizando um polígono estrelar de seis pontas. Quarta Coluna: *kernel* anisotrópico g_2 do *KAZE*. É possível ver de certa forma estruturas mais definidas para o esquema de difusão usado pelo CenSurE (AGRAWAL; KONOLIGE; BLAS, 2008).

pode também produzir melhores resultados. Como trabalho futuro, buscar-se-á explorar que espaços de escala que consideram a degradação causada pela turbidez de forma a obter melhores resultados de repetibilidade.

6 Testes e Resultados 2: Contexto em Classificação Subaquática

Aqui são apresentados os resultados de aplicação do método proposto baseado em Geoestatística (no Cap. 3), comparado com outros métodos, com e sem a incorporação do contexto.

O método será aplicado em mosaicos de imagens do assoalho oceânico, para obtenção de *mapas temáticos*. As imagens resultantes são a representação final de um mosaico, feito de forma visualmente interpretável, contendo a classificação realizada de forma pixel-a-pixel.

O capítulo apresenta os *datasets*, compostos por mosaicos, utilizados como caso de teste para a classificação e também as configurações utilizadas para os testes e, por fim, os resultados da classificação dos mosaicos são mostrados.

6.1 *Datasets* Utilizados

Para avaliação dos resultados obtidos foi proposto utilizar dois *datasets* distintos de mosaicos de recifes de corais. Cada *dataset* é composto por um mosaico obtido pela junção de centenas de imagens coletadas por especialistas da Universidade de Miami.

O *dataset Redsea* contém imagens do Mar Vermelho, capturadas em águas bastante rasas perto da cidade de Eilat, como parte de uma pesquisa por ecologistas de recifes de corais (ZVULONI et al., 2009). Para a classificação, foram considerado cinco classes: *Urchin*, *Branching Coral*, *Brain Coral*, *Favid Coral* e o *Background*. Os mosaicos utilizados foram capturados a uma resolução de 1.1 pixels/mm^2 .

O segundo *dataset* chamado *Marker*, foi capturado nas Bahamas. Foi feita a divisão em quatro classes para classificação. *General Corals*, *Sea Gorgons*, *Sand* e *Background*. Os mosaicos utilizados foram capturado em uma resolução de 2.2 pixels/mm^2 .

6.2 Descrição do Geral do Sistema

Nesta seção é descrito uma versão geral do sistema, tanto para a classificação em nível unário, quanto os tipos de classificação integrando contexto.

6.2.1 Pré-Processamento

Tanto para os dados do *dataset Redsea* quanto para o caso do *dataset Marker*, a qualidade visual da imagem é bastante satisfatória, contendo baixa presença de degradação devido a turbidez. O principal tipo de degradação encontrado é a variação de cor ao longo do *datasets* existentes durante a captura. Para resolver esta questão foi utilizado *CLAHE* (ZUIDERVELD, 1994) e uma normalização de cor em ambos os *datasets*.

6.2.2 Segmentação e Descrição

Os *datasets* foram segmentados como *superpixels* baseados em *TurboPixels* (LEVINSHTEIN et al., 2009). O tamanho do pixel foi escolhido como que para caber em uma janela de tamanho de aproximadamente 32x32 pixels.

Para cada superpixel, a combinação entre três descritores de textura foi utilizada. Filtros de Gabor, CLBP e GLCM. Um *kernel mapping* é feito depois para tornar os descritores mais linearmente separáveis. O resultado é também por fim, normalizado.

6.2.3 Classificação

No caso, para todos os testes, foi utilizado um SVM configurado com um *kernel* linear.

6.2.4 Adição de Contexto

A adição de contexto é apresentada feita de duas formas distintas: utilizando os *Conditional Random Fields* (*CRF*) e utilizando o modelo de Geoestatística, o quais foram explicados nos Capítulos 2 e 3.

Para o *CRF* utilizado o algoritmo de *Loopy Belief Propagation* (LBP) para realizar a inferência estatística, dado a sua baixa taxa de erros e alta performance (WEISS, 2000).

6.3 Treinamento

Aqui é descrito como foi realizado o treinamento das partes do sistema onde o treinamento é necessário.

6.3.1 Treinamento do Classificador

O treinamento unário diz respeito ao treinamento da função de discriminação do classificador.

Tanto para o *dataset Redsea* quanto para o *Marker* foram feitas diversas amostras dos mosaicos existentes para o treinamento do classificador de cada uma das classes. A Figura 26, mostra exemplos de segmentos usados para o treinamento do classificador para os dois *datasets* utilizados. Foram feitas amostras de por volta de 200 segmentos para cada classe sendo cada uma tendo 64x64 pixels. As amostras foram indicadas por especialistas da universidade de Miami.

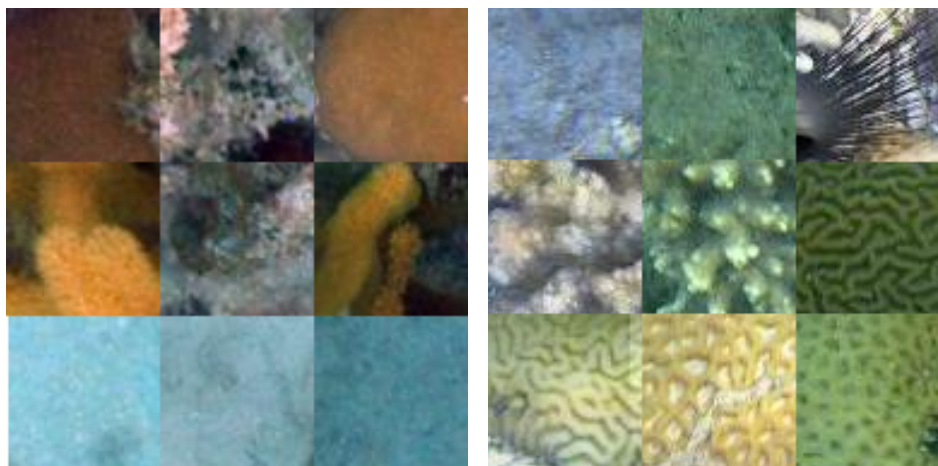


Figura 26 – Partes manualmente segmentadas utilizadas para treinamento do classificador. A esquerda são mostrados exemplos de nove amostras usadas para treinar o *dataset Redsea*. A direita são apresentadas nove amostras do *dataset Marker*.

6.3.2 Treinamento Unário

Para gerar a curva de confiança, usada para gerar os distribuição de probabilidades unária tanto para o CRF, quanto para o modelo de Geoestatística, foram também utilizadas amostras do mosaico de treinamento da Figura 26.

As Figuras 27 e 28 mostram as curvas de confiança obtidas para cada um dos dois datasets em cada uma das classes. O processo de geração das curvas é descrito no Capítulo 3 Seção 3.2.

Pelos gráficos das Figuras 27 e 28 percebe-se que determinadas classes se adaptam melhor que outras a uma curva de confiança, como a classe *Sea Gorgon* (Fig. 28c) do *dataset Marker*. Para esta classe é possível saber quais distâncias do classificador que existe uma grande probabilidade de se acertar a classe, enquanto para outras o modelo não se adaptou tão adequadamente (Fig. 29b).

Entretanto, o principal erro em adaptação da curva se da na classe *Background* para os dois casos (Fig. 27a e Fig. 29a). Isso se da devido a alta variabilidade intra-classe inerente a classe *Background*, a qual contém todos os tipos de objetos que não são de interesse para classificação.

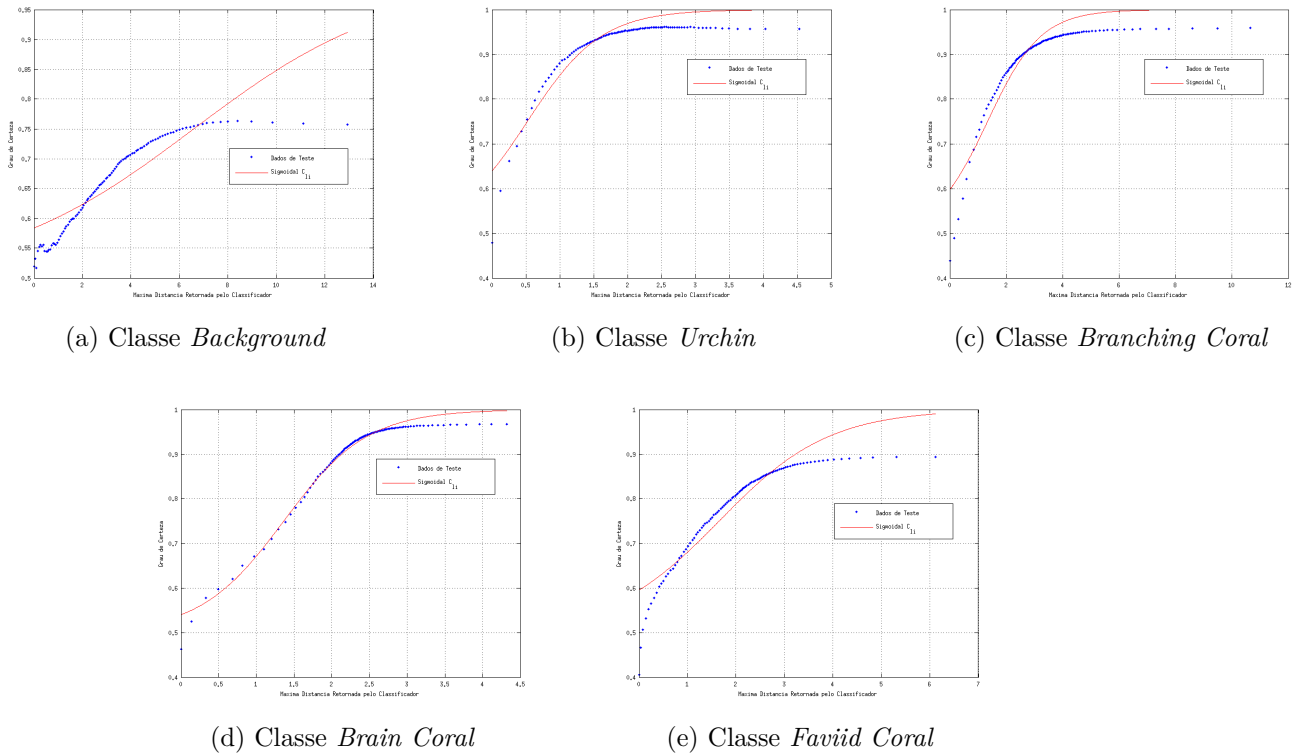


Figura 27 – Curvas de confiança geradas no treinamento unário de cada classe para o *dataset Redsea*. A curva de confiança C_i treinada para cada uma das classes é mostrada, se bem como o grau de confiança obtido.

6.3.3 Treinamento Potenciais Locais

Para treinamento dos potenciais locais foi utilizado cada *dataset* na sua totalidade.

Para o caso do CRF foi utilizado o algoritmo de treinamento descrito no Capítulo 2, Seção 2.2.2.2 .

As Tabelas 2 e 3 mostram a matriz de covariância obtida para cada um dos dois *datasets*. Tal matriz está relacionada a uma indicação de determinada classe estar próxima a outra.

Classes	Background	Urchin	Branching Coral	Brain Coral	Faviid Coral
Background	1.9115	0.8599	0.8559	0.9094	0.8400
Urchin	0.7844	0.9679	0.9458	0.9670	1.0424
Branching Coral	0.7658	0.9897	1.3870	1.0219	0.9745
Brain Coral	0.8767	0.9605	0.9461	1.6384	0.8972
Faviid Coral	0.7612	0.9427	0.9111	0.9353	1.6972

Tabela 2 – Matriz de covariância que mostra as relações de proximidade entre as classes. Tais medidas são fatores que indicam correlação e não distribuições de probabilidade. Este resultado é normalizado ao final.

Os resultados do treinamento dos vetores de transição, necessários para a simulação

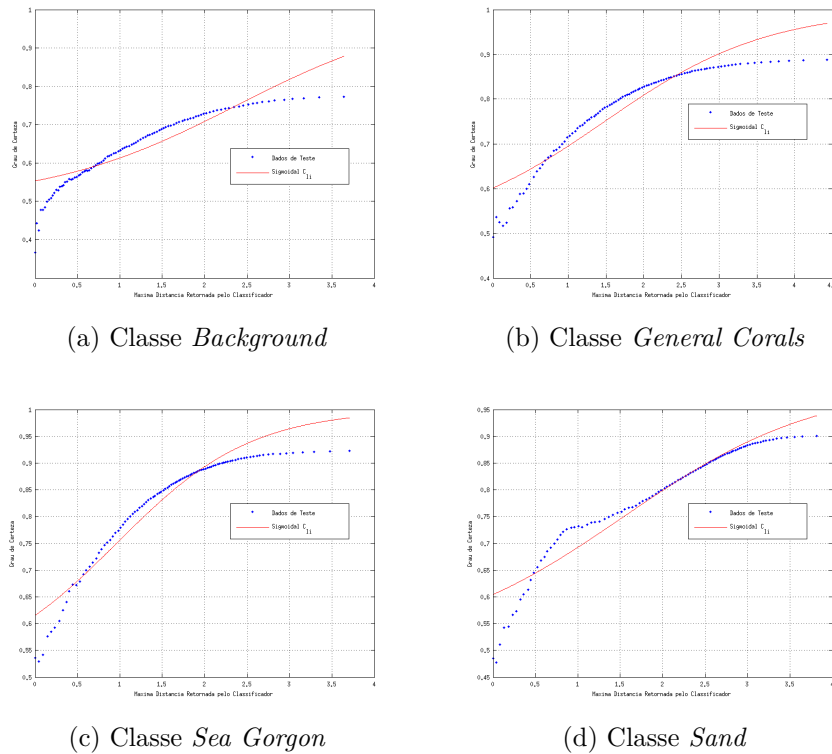


Figura 28 – Curvas de confiança geradas no treinamento unário de cada classe para o *dataset Marker*. A curva de confiança C_{l_i} treinada para cada uma das classes é apresentada, bem como o grau de confiança obtido.

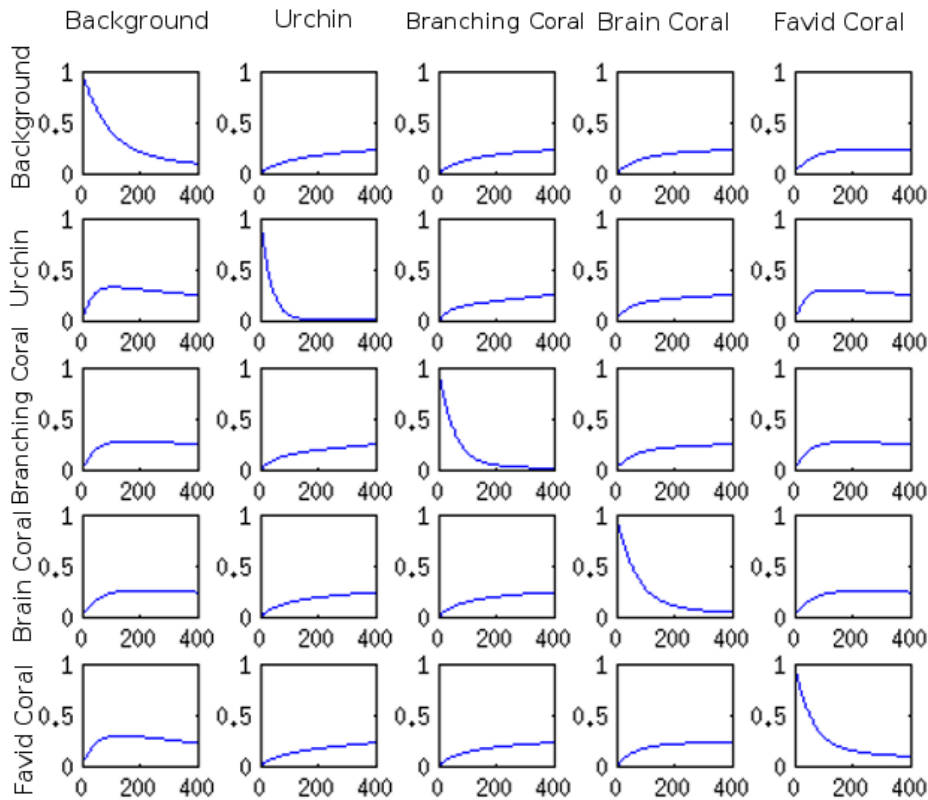
Classes	<i>Background</i>	<i>General Coral</i>	<i>Sea Gorgon</i>	<i>Sand</i>
<i>Background</i>	1.8831	0.8967	0.8899	0.9086
<i>General Coral</i>	0.9010	0.9544	0.9507	0.9028
<i>Sea Gorgon</i>	0.8905	0.9516	0.9738	0.8930
<i>Sand</i>	0.8950	0.8990	0.8906	1.8773

Tabela 3 – Matriz de covariância que mostra as relações de proximidade entre as classes. Tais medidas são fatores que indicam correlação e não distribuições de probabilidade. Este resultado é normalizado ao final.

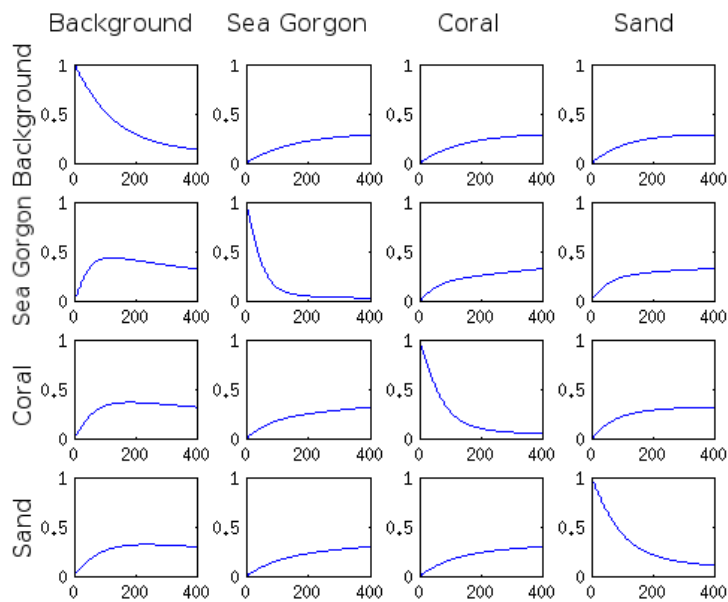
do método de Geoestatística, são mostrados na Figura 29 para ambos os *datasets* testados.

Para ambos os *datasets* analisados se observa a classe *background* como sendo predominante nas estatísticas medidas em ambos os treinamentos. No caso do treinamento dos vetores de transição (Geoestatística), também foi vista uma tendência de outras classes em transitar para o *Background* (Fig. 29). Algo que, para o treinamento dos potenciais locais do CRF, indicou principalmente uma tendência do *Background* ter proximidade consigo próprio.

Para o *dataset Redsea*, nas relações locais treinadas pelo CRF se observa algumas tendências:



(a) Dataset Redsea



(b) Dataset Marker

Figura 29 – Vetores de transição obtidos na etapa de treinamento para o método de Geoestatística do Capítulo 3. Os vetores indicam a probabilidade de uma classe transitar para outra a uma determinada distância. O eixo x apresenta a distância em pixels. O eixo y dos gráficos apresenta as probabilidades de transição. Pode-se observar, por exemplo, uma certa tendência na classe *Urchin* em transitar para categoria de *background*.

- As classes *Urchin* tem uma grande possibilidade de estar próxima a classe *Faviid Coral*;
- Cada classe tem uma forte tendência de estar próxima a si própria, o que enfatiza a pouca variabilidade de classes em espaços pequenos;
- Existe algumas tendências assimétricas treinadas, como a grande tendência da classe *Faviid Coral* estar próxima da classe *Urchin*, mas não ao contrário.

As transições assimétricas, ou seja, uma dada classe A estar próxima a classe B mas não B próxima da A , não são incentivadas pelos potenciais treinados pelo método de Geoestatística.

Considerando as relações treinadas pelo método de Geoestatística, existe uma tendência forte principalmente de transição da classe *Urchin* para a classe *Faviid Coral* e a classe *Background*.

Para o *dataset Marker*, nenhuma outra tendência de proximidade foi obtida para o CRF, fora a tendência de *background* estar próximo de si mesmo. As mesmas tendências são observadas para o treinamento dos vetores de transição para o caso da Geoestatística.

6.4 Sistemas Testados

Quatro sistemas são testados quanto a sua taxa de acerto em relação a classificação de mosaicos de imagens. Isso foi feito principalmente de forma a avaliar a consideração de contexto, juntamente com a nova proposta apresentada no Capítulo 3

Inicialmente foi avaliado o sistema *Unário*, proposto por [Shihavuddin et al. \(2013\)](#) onde somente as informações unárias são consideradas, ou seja, dada a definição de classificação considerando uma segmentação em regiões ([SHIHAVUDDIN et al., 2013](#)). Somente a descrição da própria região foi usada para classificação, o sistema é detalhado no Cap. 4 .

Após foi testado e analisado o sistema *Unário* porém baseado em distribuição de probabilidades. Em tal sistema foi feita a classificação apenas considerando a parcela unária do sistema com base no modelo em Geoestatística proposto no Cap. 3. A classificação de um segmento foi escolhida como o rótulo com máxima a probabilidade.

Apresenta-se também o sistema, *GS*, baseado em *Geoestatística* proposto no Cap. 3. A classificação de cada segmento (*Superpixel*) é dada pela Eq. 3.1, do Cap. 3.

Por fim, apresenta-se os resultados do sistema *CRF* o qual é uma implementação dos *Conditional Random Fields* , tal qual explicada no Cap 2.

Todos os sistemas foram implementados em Matlab, para o *CRF*, foi utilizada a biblioteca *UGM* para inferência estatística (SCHMIDT et al., 2009).

6.5 Computação do Mapa Temático

No *dataset Redsea*, um mosaico de 3256x2939 pixels foi utilizado para testes. Já para *dataset Marker*, foi utilizado um mosaico de 2592x3963.

As Figuras 30 e 31 mostram os mapas temáticos completos computados para ambos os datasets.

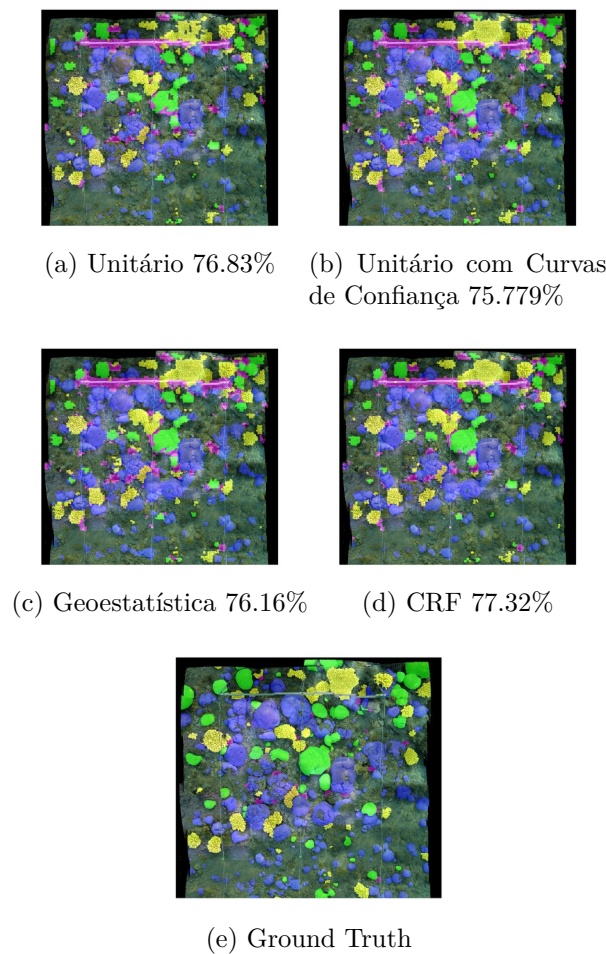


Figura 30 – Mapa temático dos Mosaicos para o *dataset Redsea*. As figuras mostram a porcentagem de acerto relativa ao *GroundTruth*. As classes são representadas pelas seguintes cores: Verde *Brain Coral*; Amarelo *Branchin Coral*; Azul *Faviid Coral*; Magenta *Urchin* e sem cor é o *background*. Os seguintes resultados são mostrados. (30a) classificação Unária. (30b) mostra a classificação Unária baseada nas curvas de confiança. (30c) classificação com adição de contexto baseada em Geoestatística. (30d) classificação com adição de contexto utilizando *CRF*.

Observa-se que num caso geral o *CRF* é o método que obtém os melhores resultados. O método de Geoestatística é capaz de melhorar um pouco, porém depende muito

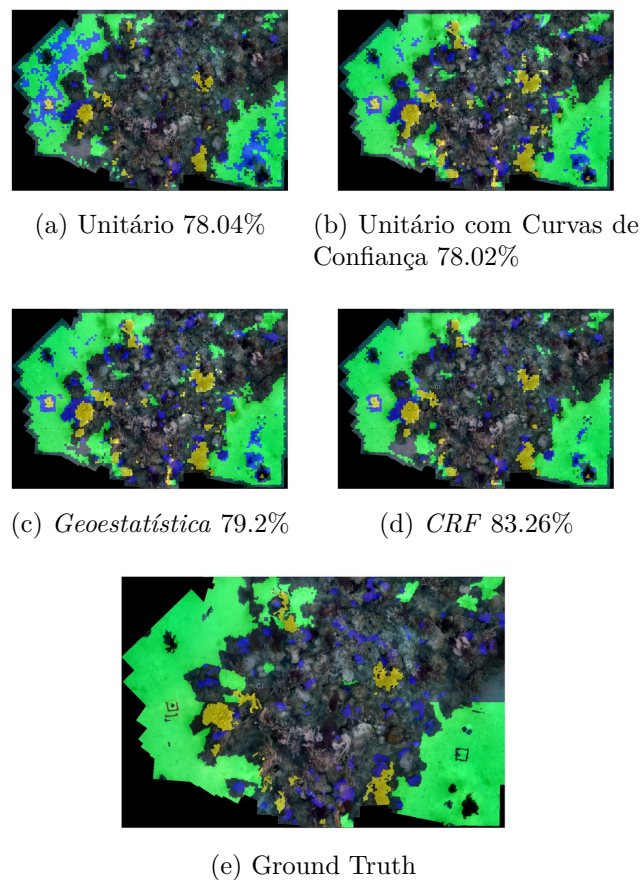


Figura 31 – Mapa temático dos Mosaicos para o *dataset Marker*. As figuras mostram a porcentagem de acerto relativa ao *GroundTruth*. As classes são representadas pelas seguintes cores: Verde *Sand*; Amarelo *Sea Gorgon*; Azul *Corals* e sem cor é o background. Os seguintes resultados são mostrados. (31a) classificação Unária. (31b) mostra a classificação Unária baseada nas curvas de confiança. (31c) classificação com adição de contexto baseada em *Geoestatística*. (31d) classificação com adição de contexto utilizando *CRF*.

de um bom treinamento da distribuição de probabilidades de cada segmento.

Para o caso do *dataset Marker*, a adição de contexto foi mais eficaz para ambos os casos. Isso ocorre dado que muitas posições geraram resultados com distribuição unária uniforme, ou seja *sem* uma classe com alta probabilidade. O que contribuiu para adição de contexto foi, que tais regiões, estavam cercadas por locais onde existia uma classe predominante.

Ao se observar a configuração do *dataset Redsea* se percebe uma tendência espacial em se ter "ilhas" de classes envolvidas pela classe *background*. Dado que a classe *background* tem uma alta variabilidade intra-classe, é bastante complicado se ter uma tendência forte para uma classe na distribuição unária. Isso dificulta a proliferação da informação de contexto na região.

De forma a analisar melhor as diferenças entre o *CRF* e o método de *Geoestatística*, é mostrado na Figura 32 duas áreas diferentes do mosaico do *Redsea* para mostrar algumas

vantagens da abordagem com base em GS. É apresentada a área original da imagem com a classificação mostrada em cores. Na primeira linha, pode se perceber o grau de acerto maior para o CRF (Fig. 32c). O CRF tende a melhorar significativamente a suavização local das estruturas classificadas. Ou seja, impõe que áreas pequenas devam ter menos variações de classes. Por esta razão o CRF teve uma boa classificação especialmente para o caso da classe representada em azul (Faviid Corals).

Por outro lado, na segunda linha, GS (Fig. 32f) obteve um grau de acerto maior. Dado que o CRF (Fig. 32g) impõe mais suavidade local, isso tende a eliminar classes menores (Fig. 32g). Este caso é evitado pela GS pelo fato de que a abordagem baseada em Geostatística usa estatísticas medidas em longas distâncias e assim o tamanho da classe é considerado. Na terceira linha da Figura 32, são mostrados resultados similares para o *dataset Marker*.

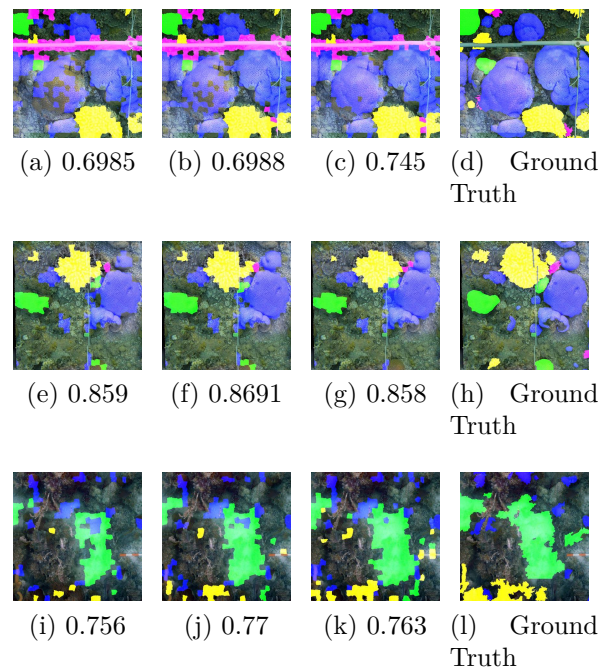


Figura 32 – Resultados de classificação para os datasets *Marker* e os *datasets Redsea*. A primeira coluna apresenta a classificação unitária. A segunda coluna apresenta os resultados de Geoestatística. A terceira coluna apresenta os resultados para o CRF. Por fim, a última coluna apresenta o *GroundTruth*. Foi utilizada como peso para o potencial local w_l como sendo 0.4 para ambas as abordagens. Na primeira coluna foi possível perceber um resultado melhor para o CRF devido a uma maior suavização local. Na segunda linha, o método de Geoestatística obteve melhores resultados devido a suas medidas estatísticas de longa distância. Na última linha é mostrado os resultados para o *dataset Marker*, onde ambas as abordagens tiveram melhores resultados para esse caso.

Também os algoritmos foram testados para múltiplos segmentos diferentes extraído dos mosaicos. Foram recortadas amostras quadradas aleatórias de diferentes tamanhos. Para cada tamanho recortado, a tabela 4 mostra a taxa de acerto média do método

Tamanho do Segmento	500	1100	1700	2300	Média
Unitary	81.1%	78.0%	79%	80.2%	79.7%
GS	81.2%	78.3%	79%	80.2%	79.8%
CRF	80.5%	78.9%	79.3%	79.7%	79.8%
Voting	78%	78%	78.4%	80.2%	79.2%

Tabela 4 – Resultados para a taxa de acerto de diferentes segmentos para o *dataset Redsea*. Foram testados diversos segmentos quadrados amostrados aleatoriamente nos mosaicos. O tamanho do segmento é especificado pelo lado do quadrado

aplicado em 20 segmentos aleatórios. Também foi testado um método simples de votação onde um superpixel é modificado caso a classe de todos os vizinhos seja diferente.

No *dataset Redsea*, para todas as abordagens, não foi percebido mais do que ganhos marginais quando comparados com a versão unitária. O método de votação também obteve resultados similares.

6.6 Conclusões

Conclui-se que o uso de estatísticas mais ricas, inspiradas pelos conceitos de Geo-estatísticas, é benéfico e pode conduzir a melhores resultados que o CRF tradicional em alguns casos.

As melhorias obtidas, foram, no entanto, de pequena magnitude, o que está alinhado com o que é discutido em (LUCCHI et al., 2011). Os resultados utilizando contexto normalmente não melhoram mais do que a suavidade local dos resultados, ou seja, não mais do que evitam grande variação de classes em uma pequena área. Porém ainda é possível obter melhorias significativas, para alguns *datasets* como no caso do *Marker*.

7 Conclusões Finais

Considerando o problema de estender a utilização de métodos de visão computacional para o cenário subaquático, esta dissertação apresentou o estudo e tratamento para alguns dos principais problemas existentes no meio.

Foi feito um estudo sobre duas áreas distintas relevantes para o problema: A detecção de pontos de interesse e o uso da informação de contexto na classificação de imagens.

Nas seções que seguem serão apresentadas as contribuições sobre as duas áreas distintas analisadas, como também as limitações das propostas.

7.1 Detectores de Pontos de Interesse em Imagens Subaquáticas Turvas

No contexto subaquático, foi feito um estudo sobre como se comportam os múltiplos detectores de pontos de interesse sobre a presença da turbidez, fenômeno o qual se faz presente no meio subaquático.

7.1.1 Contribuições Obtidas

As principais contribuições obtidas foram:

- A proposta de um dataset novo contendo imagens reais do assoalho oceânico porém com a turbidez controlada.
- Uma análise geral da repetibilidade dos detectores em meios túrbidos, dividindo a análise em intervalos de turbidez distintos.
- Dentre os detectores estudados, foi apontado o *DoG* como o mais robusto detector para ambientes com presença de turbidez. Tal detector contém também invariância a escala.
- Foi concluída a possibilidade do uso de espaços não gaussianos para geração de espaço de escala em meios subaquáticos túrbidos.

7.1.2 Limitações e Trabalhos Futuros

O estudo não foi capaz de propor um método para medir de fato a degradação causada pela turbidez. A medida utilizada é capaz de verificar a degradação estrutural o que não necessariamente está associada a turbidez.

Um outro ponto a ser tratado diz respeito a uma análise mais criteriosa com respeito a invariância a outras transformações como rotação ou escala, *juntamente* com a robustez à degradação causada pela turbidez.

7.2 Adição de Contexto Baseado em Geoestatística

Foi proposto um novo método para adicionar informação espacial na classificação de imagens. O método se baseou nos estudos da área de Geoestatística. Tal método foi aplicado em imagens de mosaicos de recifes de corais. O mesmo foi comparado com as versões sem a utilização de contexto e com o modelo dos *Conditional Random Fields* (*CRF*).

Apresentou-se que a adição de contexto pode, em alguns casos, ser benéfica para a classificação de imagens subaquáticas. Obtendo-se um ganho de até 5% a mais em taxa de acerto.

7.2.1 Contribuições Obtidas

O trabalho apresentou um novo método para adição de contexto em imagens subaquáticas. O uso de medidas estatísticas mais ricas, como as baseadas em Geoestatística, mostrou-se útil em algumas situações para adição de informação de contexto.

Também essa dissertação serve como uma conexão entre duas áreas distintas: a Geoestatística e os Modelos Probabilístico Gráficos (MPGs). Acredita-se que através desta intersecção, as aplicações que fazem uso de Geoestatística podem também se beneficiar dos MPGs.

7.2.2 Limitações e Trabalhos Futuros

Coloca-se que a abordagem apresentada foi aplicada para um cenário subaquático específico. Uma direção seria a aplicação em *dataset* com classes mais genéricas, os da *Pascal Visual Object Classes Challenge* (VOC) ([EVERINGHAM et al., 2010](#)).

Os resultados para o método de Geoestatística foram satisfatórios porém ficaram abaixo em taxa de acerto quando comparados ao *CRF*. Apesar da tendência de se usar os métodos de Geoestatística para casos onde há pouca quantidade de informações ([CARLE; FOGG, 1996](#)).

Existe ainda uma necessidade maior de alteração no modelo original de Geoestatística visando uma melhor adaptação para o caso de imagens subaquáticas.

Referências

- ABFALG, J. et al. Multi-represented classification based on confidence estimation. In: *Advances in Knowledge Discovery and Data Mining*. [S.l.]: Springer, 2007. p. 23–34. Citado 2 vezes nas páginas 51 e 52.
- AGRAWAL, M.; KONOLIGE, K.; BLAS, M. R. Censure: Center surround extremas for realtime feature detection and matching. In: *Computer Vision–ECCV 2008*. [S.l.]: Springer, 2008. p. 102–115. Citado 4 vezes nas páginas 14, 35, 77 e 81.
- AGTERBERG, F. Mathematical geology. In: *General Geology*. Springer US, 1988, (Encyclopedia of Earth Science). p. 573–582. ISBN 978-0-442-22499-8. Disponível em: <http://dx.doi.org/10.1007/0-387-30844-X_76>. Citado na página 55.
- ALCANTARILLA, P. F.; BARTOLI, A.; DAVISON, A. J. Kaze features. In: *Computer Vision–ECCV 2012*. [S.l.]: Springer, 2012. p. 214–227. Citado 5 vezes nas páginas 35, 36, 77, 78 e 80.
- AULINAS, J. et al. Feature extraction for underwater visual slam. In: IEEE. *OCEANS, 2011 IEEE-Spain*. [S.l.], 2011. p. 1–7. Citado na página 24.
- BAR, M. Visual objects in context. *Nature Reviews Neuroscience*, Nature Publishing Group, v. 5, n. 8, p. 617–629, 2004. Citado 2 vezes nas páginas 25 e 41.
- BAY, H. et al. Speeded-up robust features (surf). *Computer vision and image understanding*, Elsevier, v. 110, n. 3, p. 346–359, 2008. Citado 2 vezes nas páginas 33 e 76.
- BAZEILLE, S. et al. Automatic underwater image pre-processing. In: *CMM'06*. [S.l.: s.n.], 2006. p. xx. Citado na página 65.
- BEALL, C. et al. 3d reconstruction of underwater structures. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. [S.l.: s.n.], 2010. p. 4418–4423. Citado 2 vezes nas páginas 24 e 27.
- BEATTIE, C.; MILLS, B.; MAYO, V. Development drilling of the tawila field, yemen, based on three-dimensional reservoir modeling and simulation. In: *SPE annual technical conference*. [S.l.: s.n.], 1998. p. 715–725. Citado na página 49.
- BEAUDET, P. R. Rotationally invariant image operators. In: *Proceedings of the 4th International Joint Conference on Pattern Recognition*. Kyoto, Japan: [s.n.], 1978. p. 579–583. Citado 4 vezes nas páginas 30, 71, 76 e 80.
- BEIJBOM, O. et al. Automated annotation of coral reef survey images. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 1170–1177. Citado na página 66.
- BIEDERMAN, I.; MEZZANOTTE, R. J.; RABINOWITZ, J. C. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, Elsevier, v. 14, n. 2, p. 143–177, 1982. Citado na página 39.

- BOIX, X. et al. Harmony potentials. *International journal of computer vision*, Springer, v. 96, n. 1, p. 83–102, 2012. Citado 2 vezes nas páginas 46 e 47.
- BOYKOV, Y. Y.; JOLLY, M.-P. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In: IEEE. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. [S.l.], 2001. v. 1, p. 105–112. Citado na página 45.
- CARBONETTO, P.; FREITAS, N. de; BARNARD, K. A statistical model for general contextual object recognition. In: *Computer Vision-ECCV 2004*. [S.l.]: Springer, 2004. p. 350–362. Citado 2 vezes nas páginas 43 e 45.
- CARLE, S. F.; FOGG, G. E. Transition probability-based indicator geostatistics. *Mathematical Geology*, Springer, v. 28, n. 4, p. 453–476, 1996. Citado 4 vezes nas páginas 49, 56, 57 e 96.
- CARLE, S. F. et al. Conditional simulation of hydrofacies architecture: a transition probability/markov approach. *Hydrogeologic models of sedimentary aquifers, concepts in hydrogeology and environmental geology*, v. 1, p. 147–170, 1998. Citado 2 vezes nas páginas 54 e 55.
- CORKE, P. et al. Experiments with underwater robot localization and tracking. In: *Robotics and Automation, 2007 IEEE International Conference on*. [S.l.: s.n.], 2007. p. 4556–4561. ISSN 1050-4729. Citado na página 27.
- CRISTIANINI, N.; SHAWE-TAYLOR, J. *An introduction to support vector machines and other kernel-based learning methods*. [S.l.]: Cambridge university press, 2000. Citado na página 51.
- CRISTINACCE, D.; COOTES, T. F. Feature detection and tracking with constrained local models. In: CITESEER. *BMVC*. [S.l.], 2006. v. 2, n. 5, p. 6. Citado na página 78.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, JSTOR, p. 1–38, 1977. Citado na página 45.
- DERPANIS, K. G.; LEUNG, E. T.; SIZINTSEV, M. Fast scale-space feature representations by generalized integral images. In: IEEE. *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. [S.l.], 2007. v. 4, p. IV–521. Citado na página 33.
- 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 65 e 74.
- EMERY, X. Properties and limitations of sequential indicator simulation. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 18, n. 6, p. 414–424, 2004. Citado na página 54.
- 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 96.
- FINK, M.; PERONA, P. Mutual boosting for contextual inference. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2003. p. None. Citado na página 42.

FISCHLER, M. A.; ELSCHLAGER, R. A. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, Citeseer, v. 22, n. 1, p. 67–92, 1973. Citado na página 40.

FULKERSON, B.; VEDALDI, A.; SOATTO, S. Class segmentation and object localization with superpixel neighborhoods. In: IEEE. *Computer Vision, 2009 IEEE 12th International Conference on*. [S.l.], 2009. p. 670–677. Citado 4 vezes nas páginas 41, 44, 45 e 68.

GALLEGUILLOS, C.; BELONGIE, S. Context based object categorization: A critical survey. *Computer Vision and Image Understanding*, Elsevier, v. 114, n. 6, p. 712–722, 2010. Citado 3 vezes nas páginas 39, 40 e 41.

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 24, 71, 75 e 77.

GIL, A. et al. A comparative evaluation of interest point detectors and local descriptors for visual slam. *Machine Vision and Applications*, Springer, v. 21, n. 6, p. 905–920, 2010. Citado 2 vezes nas páginas 27 e 78.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X. Citado na página 67.

GUO, Z.; ZHANG, D. A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on*, IEEE, v. 19, n. 6, p. 1657–1663, 2010. Citado na página 69.

HANSON, A. R.; RISEMAN, E. M. VISIONS: A computer system for interpreting scenes. In: HANSON, A. R.; RISEMAN, E. M. (Ed.). *Computer Vision Systems*. New York: Academic Press, 1978. Citado na página 40.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, n. 6, p. 610–621, 1973. Citado na página 69.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: MANCHESTER, UK. *Alvey vision conference*. [S.l.], 1988. v. 15, p. 50. Citado 5 vezes nas páginas 29, 30, 71, 76 e 80.

JOHNSON-ROBERSON, M.; KUMAR, S.; WILLAMS, S. Segmentation and classification of coral for oceanographic surveys: a semi-supervised machine learning approach. In: IEEE. *OCEANS 2006-Asia Pacific*. [S.l.], 2007. p. 1–6. Citado na página 66.

KOLTUN; VLADLEN. Efficient inference in fully connected crfs with gaussian edge potentials. In: . [S.l.: s.n.], 2011. Citado 2 vezes nas páginas 46 e 60.

KOSCHMIEDER, H. *Theorie der horizontalen Sichtweite*. [S.l.]: Keim Nemnich, 1924. Citado na página 64.

KRUPPA, H.; SCHIELE, B. *Using Local Context to Improve Face Detection*. 2003. Citado na página 42.

- KUMAR, S.; HEBERT, M. A hierarchical field framework for unified context-based classification. In: IEEE. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. [S.l.], 2005. v. 2, p. 1284–1291. Citado na página 40.
- LEVINSHTEIN, A. et al. Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 31, n. 12, p. 2290–2297, 2009. Citado 3 vezes nas páginas 49, 68 e 84.
- LINDEBERG, T. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, Taylor & Francis, v. 21, p. 225–270, 1994. Citado 2 vezes nas páginas 31 e 32.
- LINDEBERG, T. *On the axiomatic foundations of linear scale-space*. [S.l.]: Springer, 1997. Citado na página 32.
- LINDEBERG, T. Feature detection with automatic scale selection. *International journal of computer vision*, Springer, v. 30, n. 2, p. 79–116, 1998. Citado na página 32.
- LINDEBERG, T.; EKLUNDH, J.-O. On the computation of a scale-space primal sketch. *Journal of Visual Communication and Image Representation*, v. 2, n. 1, p. 55 – 78, 1991. ISSN 1047-3203. Citado na página 29.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, n. 2, p. 91–110, 2004. Citado 8 vezes nas páginas 11, 33, 34, 45, 72, 76, 78 e 80.
- LUCCHI, A. et al. Are spatial and global constraints really necessary for segmentation? In: IEEE. *Computer Vision (ICCV), 2011 IEEE International Conference on*. [S.l.], 2011. p. 9–16. Citado 3 vezes nas páginas 47, 49 e 93.
- MARCOS, M. S. A.; SORIANO, M.; SALOMA, C. Classification of coral reef images from underwater video using neural networks. *Optics express*, Optical Society of America, v. 13, n. 22, p. 8766–8771, 2005. Citado na página 66.
- MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, SIAM, v. 11, n. 2, p. 431–441, 1963. Citado na página 52.
- MIKOLAJCZYK, K.; SCHMID, C. Scale & affine invariant interest point detectors. *International journal of computer vision*, Springer, v. 60, n. 1, p. 63–86, 2004. Citado 2 vezes nas páginas 32 e 71.
- 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 na página 74.
- NEMETH, R. S. et al. Characterization of deep water reef communities within the marine conservation district, st. thomas, us virgin islands. 2008. Citado na página 23.
- NICOSEVICI, T. et al. Efficient three-dimensional scene modeling and mosaicing. *Journal of Field Robotics*, v. 26, 2009. Citado 2 vezes nas páginas 24 e 27.

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 65.

PADMAVATHI, G.; MUTHUKUMAR, M.; THAKUR, S. K. Kernel principal component analysis feature detection and classification for underwater images. In: IEEE. *Image and Signal Processing (CISP), 2010 3rd International Congress on*. [S.l.], 2010. v. 2, p. 983–988. Citado 2 vezes nas páginas 24 e 27.

PERONA, P.; MALIK, J. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 12, n. 7, p. 629–639, 1990. Citado na página 36.

PETROU, M.; GARCÍA-SEVILLA, P. *Image processing - dealing with texture*. [S.l.]: Wiley, 2006. ISBN 978-0-470-02628-1. Citado na página 68.

PIZARRO, O.; EUSTICE, R.; SINGH, H. Large area 3d reconstructions from underwater surveys. In: *OCEANS '04. MTTs/IEEE TECHNO-OCEAN '04*. [S.l.: s.n.], 2004. v. 2, p. 678–687 Vol.2. Citado na página 66.

PIZARRO, O. et al. Towards image-based marine habitat classification. In: IEEE. *OCEANS 2008*. [S.l.], 2008. p. 1–7. Citado na página 66.

PLATT, J. C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: CITESEER. *Advances in large margin classifiers*. [S.l.], 1999. Citado na página 52.

PORTER, R.; CANAGARAJAH, N. Robust rotation-invariant texture classification: wavelet, gabor filter and gmrf based schemes. In: IET. *Vision, Image and Signal Processing, IEE Proceedings-*. [S.l.], 1997. v. 144, n. 3, p. 180–188. Citado na página 69.

PURKIS, S.; VLASWINKEL, B.; GRACIAS, N. Vertical-to-lateral transitions among cretaceous carbonate facies: A means to 3-d framework construction via markov analysis. *Journal of Sedimentary Research*, SEPM, v. 82, n. 4, p. 232–243, 2012. Citado na página 49.

RABINOVICH, A. et al. Objects in context. In: *Proceedings of the International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2007. Citado na página 40.

SCHETTINI, R.; CORCHS, S. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, Hindawi Publishing Corp., v. 2010, p. 14, 2010. Citado na página 67.

SCHMID, C.; MOHR, R.; BAUCKHAGE, C. Evaluation of interest point detectors. *International Journal of computer vision*, Springer, v. 37, 2000. Citado 2 vezes nas páginas 30 e 77.

SCHMIDT, M. W. et al. Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. In: *International Conference on Artificial Intelligence and Statistics*. [S.l.: s.n.], 2009. p. None. Citado na página 90.

- SHIHAVUDDIN, A. et al. Image-based coral reef classification and thematic mapping. *Remote Sensing*, v. 5, n. 4, p. 1809–1841, 2013. ISSN 2072-4292. Disponível em: <http://www.mdpi.com/2072-4292/5/4/1809>. Citado 7 vezes nas páginas 25, 66, 67, 68, 69, 70 e 89.
- SHOTTON, J. et al. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, Springer, v. 81, n. 1, p. 2–23, 2009. Citado 2 vezes nas páginas 44 e 45.
- SIVIC, J.; ZISSERMAN, A. Video google: Efficient visual search of videos. In: *Toward Category-Level Object Recognition*. [S.l.]: Springer, 2006. p. 127–144. Citado na página 45.
- SOJKA, E. A new approach to detecting the corners in digital images. In: IEEE. *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. [S.l.], 2003. v. 3, p. III–445. Citado 2 vezes nas páginas 11 e 31.
- 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 2 vezes nas páginas 66 e 69.
- SUTTON, C.; MCCALLUM, A. An introduction to conditional random fields for relational learning. *Introduction to statistical relational learning*, MIT press, p. 93–128, 2006. Citado 2 vezes nas páginas 42 e 43.
- TORRALBA, A. Contextual priming for object detection. *International journal of computer vision*, Springer, v. 53, n. 2, p. 169–191, 2003. Citado na página 40.
- TORRALBA, A.; MURPHY, K. P.; FREEMAN, W. T. Contextual models for object detection using boosted random fields. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2004. p. 1401–1408. Citado 2 vezes nas páginas 40 e 41.
- TREIBITZ, T.; SCHECHNER, Y. Y. Instant 3descatter. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*. [S.l.: s.n.], 2006. (CVPR '06), p. 1861–1868. Citado 2 vezes nas páginas 64 e 74.
- TUYTELAARS, T.; MIKOLAJCZYK, K. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, Now Publishers Inc., v. 3, n. 3, p. 177–280, 2008. Citado 4 vezes nas páginas 27, 28, 30 e 76.
- WANG, Z. et al. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, IEEE, v. 13, n. 4, p. 600–612, 2004. Citado 2 vezes nas páginas 75 e 77.
- WEICKERT, J.; ROMENY, B. T. H.; VIERGEVER, M. A. Efficient and reliable schemes for nonlinear diffusion filtering. *Image Processing, IEEE Transactions on*, IEEE, v. 7, n. 3, p. 398–410, 1998. Citado na página 36.
- WEIJER, J. V. D.; SCHMID, C. Coloring local feature extraction. In: *Computer Vision–ECCV 2006*. [S.l.]: Springer, 2006. p. 334–348. Citado na página 69.

- WEISS, Y. Correctness of local probability propagation in graphical models with loops. *Neural computation*, MIT Press, v. 12, n. 1, p. 1–41, 2000. Citado 2 vezes nas páginas [45](#) e [84](#).
- ZUIDERVELD, K. Contrast limited adaptive histogram equalization. In: ACADEMIC PRESS PROFESSIONAL, INC. *Graphics gems IV*. [S.l.], 1994. p. 474–485. Citado 2 vezes nas páginas [67](#) e [84](#).
- ZVULONI, A. et al. Spatio-temporal transmission patterns of black-band disease in a coral community. *PLoS One*, Public Library of Science, v. 4, n. 4, p. e4993, 2009. Citado 2 vezes nas páginas [73](#) e [83](#).