

Universidade do Estado do Rio de Janeiro Centro de Tecnologia e Ciências Instituto Politécnico

Caio Lucas dos Santos Souza

Rastreamento de ondas oceânicas em múltiplos vídeos

Nova Friburgo 2017

Caio Lucas dos Santos Souza

## Rastreamento de ondas oceânicas em múltiplos vídeos

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Ricardo Fabbri Coorientador: Prof. Dr. Francisco Duarte Moura Neto

Nova Friburgo 2017

# CATALOGAÇÃO NA FONTE UERJ / REDE SIRIUS / BIBLIOTECA CTC/E

S729	Souza, Caio Lucas dos Santos. Rastreamento de ondas oceânicas em múltiplos vídeos / Caio Lucas dos Santos Souza 2017. 128 f.: il.
	Orientadores: Ricardo Fabbri e Francisco Duarte Moura Neto. Dissertação (mestrado) – Universidade do Estado do Rio de Janeiro, Instituto Politécnico.
	1. Ondas – Métodos de simulação - Teses. 2. Fotogrametria – Modelos matemáticos – Teses. 3. Imagens digitais - Teses. 4. Curvas – Rastreamento – Teses. I. Fabbri, Ricardo. II. Moura Neto, Francisco Duarte. III. Universidade do Estado do Rio de Janeiro. Instituto Politécnico. IV. Título. CDU 528.7:514.752.2

Bibliotecária Cleide Sancho CRB7/5843

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Caio Lucas dos Santos Souza

## Rastreamento de ondas oceânicas em múltiplos vídeos

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional, da Universidade do Estado do Rio de Janeiro.

Aprovada em 28 de Agosto de 2017. Banca Examinadora:

> Prof. Dr. Ricardo Fabbri (Orientador) Instituto Politécnico – UERJ

Prof. Dr. Francisco Duarte Moura Neto (Coorientador) Instituto Politécnico – UERJ

Nelson Violante de Carvalho Universidade Federal do Rio de Janeiro – UFRJ

Lívia Flavia Carletti Jatobá Institituto Politécnico – UERJ

Victor de Amorim d'Ávila Instituto Politécnico – UERJ

God grant me the serenity to accept the things I cannot change, Courage to change the things I can change, And wisdom to know the difference. Reinhold Niebuhr

### RESUMO

SOUZA, Caio Lucas dos Santos Rastreamento de ondas oceânicas em múltiplos vídeos. 2017. 128 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2017.

O rastreamento de padrões na superfície da água em vídeos e imagens é um desafio atual na visão computacional. A interface água-ar impõe diversas dificuldades como falta de texturização suficiente, transparência, reflexões especulares, dentre outras. As abordagens atuais gerais de rastreamento e reconstrução se mostram pouco eficientes nesse cenário, enquanto abordagens mais eficientes tornam-se excessivamente específicas e controladas. Neste trabalho é proposta uma abordagem diferente, baseada na representação de informação por curvas e suas características, ao invés das mais comumente utilizadas por nuvens de pontos e grades regulares. Para isto, são explorados um detector de bordas subpíxel e um vinculador de bordas multi-estágios partindo de agrupamentos locais até considerações globais, sempre respeitando a coerência geométrica para a formação de curvas. O último passo é o aprendizado de máquina aplicado à extração de curvas semântica e geometricamente verídicas. Embora nos experimentos seja utilizado um conjunto de dados de vídeos de um tanque de ondas, esta abordagem pode ser estendida a ondas oceânicas e também a outras cenas, já que os métodos utilizados são bastante gerais. Assim, este trabalho pode ser utilizado diretamente ou como um módulo em outros sistemas.

Palavras-chave: Rastreamento de curvas. Múltiplas vistas. Ondas aquáticas.

Aprendizado de máquina.

## ABSTRACT

SOUZA, Caio Lucas dos Santos *Multiview ocean wave tracking*. 2017. 128 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2017.

Tracking patterns on the water surface in videos and images is a challenge in present computer vision. The water-air interface imposes several difficulties such as lack of sufficient texturing, transparency, specular reflections, among others. Current general approaches for tracking and reconstruction are not very efficient for the surfaces in this particular environment; while there are more efficient approaches, they become overly specific and controlled. In this work a different approach is proposed, based on curves and their characteristics to represent the desired information, instead of the usual representations based on clouds of points and regular grids. For this, a sub-pixel edge detector and a multi-stage edge linker are exploited. Performing local to global groupings and considerations, the methodology always respects the geometric coherence in the formation of curves. The last step includes machine learning tools applied to the extraction of curves semantically and geometrically veridical. Although a set of videos from a wave tank is used in the experiments, this approach can be extended to ocean waves and also to other scenes, since the methods used are very general. Thus, this work can be used directly or as a module in other systems.

Keywords: Curve tracking. Multiview. Water waves. Machine learning.

## LISTA DE FIGURAS

Figura 1	- Tanque de ondas na COPPE/UFRJ	16
Figura 2	- Artefatos de compressão, <i>blocking</i>	17
Figura 3	- Fontes físicas de curvas em imagens.	18
Figura 4	- Técnica estéreo mais difundida na modelagem de ondas	24
Figura 5	- Exemplo de construção das curvas representadas por um $\mathit{curve\ bundle}$ .	27
Figura 6	- Requerimentos na extração de curvas	29
Figura 7	- Visão geral dos principais estágios na extração genérica de curvas	30
Figura 8	- Vinculador simbólico de <i>edgels</i>	34
Figura 9	- Detector de elementos de borda subpíxel	35
Figura 10	- Vários <i>edgels</i> nomeados de X a Z e de A a D, e seus feixes de curva	
	multi coloridos.	36
Figura 11	- Redução das possíveis curvas passando por, um, dois e três $\mathit{edgels.}$	37
Figura 12	- Árvore de hipóteses	38
Figura 13	- Quatro estágios para a extração de fragmentos de curva	39
Figura 14	- A aplicação de um classificador nas possíveis conexões entre as curvas.	41
Figura 15	- Os dois estágios que utilizam o aprendizado de máquina	44
Figura 16	- Feixe de curvas de arco circulares	48
Figura 17	- Etapas na construção de <i>curvelets</i>	48
Figura 18	- Restrição do feixe de curvas conforme a adição de novos <i>edgels</i>	49
Figura 19	- Representação de um feixe de curvas no espaço de parâmetros	50
Figura 20	- Resultado das interseções dos espaços de curva de cada <i>edgel.</i>	51
Figura 21	- Etapas na extração de fragmentos de curva não-ambíguos	52
Figura 22	- $CFG$ Fragmentos de curvas não-ambíguas	53
Figura 23	- $\boldsymbol{CFTG}$ Fragmentos de curvas não-ambíguas em cores sólidas e classes	
	de equivalência das possíveis junções tracejadas	54
Figura 24	- $CFTG$ após remoção das ambiguidades	54
Figura 25	- Dimensões do tanque de ondas LabOceano COPPE/UFRJ $\ .\ .\ .$ .	71
Figura 26	- Quadros representativos extraídos de cada video	72
Figura 27	- Cortes da área de interesse	73
Figura 28	- Imagens das áreas selecionadas	73
Figura 29	- Interface gráfica para visualização e edição de curvas	75
Figura 30	- Quebra de uma curva na interface gráfica.	75
Figura 31	- Junção de duas curvas na interface gráfica	76
Figura 32	- Remoção de uma curva na interface gráfica	76
Figura 33	- Curvas extraídas automaticamente do quadro 140 da câmera 1 e refe	
	rências.	77

Figura $34$	- Auto-gamma — histograma das intensidades dos píxeis	80
Figura 35	- Mapeamento feito pela função <i>auto-gamma</i>	81
Figura 36	- O método de terceira ordem aplicado a uma imagem em detalhes	83
Figura 37	- Comparação visual do método de terceira ordem e o método gaussiano.	84
Figura 38	- Comparação visual do método de terceira ordem para diferentes valores	
	$\det \sigma$	84
Figura 39	- Filtro por tamanho aplicado com diferentes parâmetros num mesmo	
	quadro.	85
Figura 40	- Resultado do filtro por tamanho aplicado nas vistas das quatro câmeras.	86
Figura 41	- Marcações de ruídos na Figura 40 (a).	87
Figura 42	- Filtro de 160 $px$ aplicado em ondas mais quebradas	87
Figura 43	- (Junção) Histogramas das diferenças entre as características fotométri-	
	cas 1D (HSV)	89
Figura 44	- (Junção) Histogramas das características geométricas 1D e 2D	90
Figura 45	- (Junção) Histograma das características geométrica $0{\rm D}$ e fotométrica	
	2D	91
Figura 46	- (Junção) Erros para o classificador <i>merge</i> usando apenas a caracterís-	
	tica em destaque	92
Figura 47	- (Seleção) Histogramas das características fotométricas 1D (HSV). $\ldots$	93
Figura 48	- (Seleção) Histogramas das características geométricas 1D e 2D. $\ldots$ .	94
Figura 49	- (Seleção) Histograma das características geométrica 1D, e fotométrica	
	1D	95
Figura 50	- (Seleção) Erros para o classificador $\mathit{rank}$ usando apenas a característica	
	em destaque.	96
Figura 51	- Erro calculado sobre o conjunto de teste para os classificadores	97
Figura 52	- Comparação entre o erro calculado sobre o conjunto de teste	97
Figura 53	- Histogramas da probabilidade obtida pelos classificadores. $\ldots$ .	98
Figura 54	- Comparação entre cobertura ( <i>recall</i> ) e precisão ( <i>precision</i> )	98
Figura 55	- Comparação entre cobertura (recall) e precisão (precision) sobrepostas.	99
Figura 56	- Medida F para cada extração de curva	100
Figura 57	- Análise de quebra e junção em alguns pontos	101
Figura 58	- Comparação visual entre a extração supervisionada e não-supervisionada,	
	para uma cobertura de 90%, câmera 1	102
Figura 59	- Comparação visual entre a extração supervisionada e não-supervisionada,	
	para uma cobertura de 90%, câmera 2	103
Figura 60	- Comparação visual entre a extração supervisionada e não-supervisionada,	
	para uma cobertura de 90%, câmera 3	104
Figura 61	- Comparação visual entre a extração supervisionada e não-supervisionada,	
	para uma cobertura de 90%, câmera 4.	105

igura 62 - Quatro vistas das câmeras, 1, 2, 3 e 4 nesta ordem, do quadro 32 para
a extração não-supervisionada com filtro de tamanho. 106
igura 63 - Quatro vistas das câmeras, 1, 2, 3 e 4 nesta ordem, do quadro 32 para
a extração supervisionada com quebra, junção e seleção. 107
igura 64 - ModeloKodak KAI 02150, com Kowa Motorized Zoom Lens LMZ 1177-
IR
igura 65 - Disposição das câmeras, vista de cima, com sistemas de coordenadas
indicados, relativos à Figura 25 $\dots \dots \dots$
igura 66 - Correspondências entre pontos e retas, marcadas manualmente 126
igura 67 - Interface gráfica mostrando os pontos utilizados na otimização dos pa-
râmetros das câmeras

# LISTA DE TABELAS

Tabela 1	- Características utilizadas nos classificadores	59
Tabela 2	- Posições das câmeras.	71
Tabela 3	- Tamanhos dos cortes realizados e posições do canto superior esquerdo	
	de onde se iniciam os cortes em cada vídeo. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	72
Tabela 4	- Resumo dos parâmetros de uma câmera em visão computacional	122

# LISTA DE ALGORITMOS

Algoritmo	1 -	- Edgel Sparsity	•	67
Algoritmo	2 -	- Histograma de Textura	•	68

# LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BFS	Breadth First Search
BSDS	Berkeley Segmentation Dataset
CFG	Curve Fragment Graph
CFGD	Curve Fragment Ground-Truth Dataset
CFTG	Curve Fragment Topology Graph
DFS	Depth First Search
ENG	Edgel Neighborhood Graph
ETG	Edgel Topology Graph
gPb	Global Probability Boundary
HSV	Hue Saturation Value
MPI	Message Parsing Interface
RGB	Red Blue Green
SE	Structured Edge
SIFT	Scale Invariant Feature Transform
VXD	Vision Something Development (Library)
VXL	Vision Something Library

# SUMÁRIO

	INTRODUÇÃO	14
1	ABORDAGEM BASEADA EM CURVAS	26
1.1	Terminologia	26
1.1.1	Critérios para avaliação dos fragmentos de curvas	28
1.2	Abordagens recentes	30
1.2.1	Relação com trabalhos anteriores	31
1.2.2	Extração clássica de curvas	31
1.2.3	Literatura recente	31
1.3	De elementos de borda a curvas	32
1.3.1	Estágio I: detecção de bordas	33
1.3.2	Estágio II: agrupamento de bordas em <i>curvelets</i>	35
1.3.3	Estágio III: traçando fragmentos de curvas não-ambíguas	36
1.3.4	Estágio IV: resolução de ambiguidades	36
1.4	Aprendizado da topologia de curvas	38
1.4.1	Estágio V: aprendizado aplicado à junção de fragmentos de curva $\ldots$	40
1.4.2	Estágio VI: extração de curvas baseada em aprendizagem	43
2	RASTREAMENTO DE CURVAS	45
2.1	Detecção de elementos de borda e o método de terceira ordem .	45
2.2	Vinculador simbólico de elementos de borda	47
2.2.1	Clusterização local	47
2.2.2	<u>Clusterização global</u>	50
2.3	Aprendizado de máquina	53
2.3.1	Modelo logístico	55
2.3.2	Classificadores e características	58
2.3.3	Características bg_grad, sat_grad e hue_grad	60
2.3.3.1	Conversão RGB para HSV	60
2.3.3.2	Cálculo da normal a uma curva discreta	60
2.3.3.3	Cálculo das características <b>bg_grad</b> , <b>sat_grad</b> e <b>hue_grad</b>	62
2.3.4	Características <b>abs_k</b> e <b>wigg</b>	62
2.3.4.1	Curvatura	62
2.3.4.2	Curvas Interpolantes	64
2.3.4.3	Cálculo das características $abs_k \in wigg$	66
2.3.5	Cálculo das características len, edge_sparsity e avg_conf	67
2.3.6	Cálculo das características <b>geom_diff</b> e <b>tex_diff</b>	68
3	AQUISIÇÃO DE DADOS	70
3.1	Tanque de ondas e aquisição dos vídeos	70

3.2	Área de interesse	72
3.3	Dados de referência	74
4	EXPERIMENTOS	78
4.1	Sistema e implementação	78
4.2	Pré-processamento das imagens	79
4.3	Detecção de elementos de borda	81
4.4	Extração de contornos não-supervisionada	82
4.5	Aprendizado de máquina aplicado à extração de contornos	88
4.5.1	Junção e quebra de contornos	88
4.5.2	<u>Seleção de contornos</u>	91
4.6	Resultados	94
	CONCLUSÃO	108
	<b>REFERÊNCIAS</b>	109
	<b>APÊNDICE A</b> – Geometria e Calibração do Sistema de Câmeras	117

## INTRODUÇÃO

### Objetivos e motivação

A visão computacional é um ramo da inteligência artificial (IA) que busca a compreensão da informação visual. Esta compreensão se dá através do processamento de imagens e vídeos de forma algorítmica em computadores. Dentro da visão computacional, o presente trabalho é um módulo ou estágio que se encontra na interface de fotogrametria computadorizada, que é a ciência de extrair medidas objetivas a partir de imagens digitais, com elementos de IA de nível superior, como flexibilidade de inspiração humana, raciocínio semântico e qualitativo.

O principal objetivo deste trabalho é o rastreamento de padrões visuais em ondas aquáticas em múltiplos vídeos com a melhor qualidade possível. Essas informações podem ser usadas posteriormente como entrada para sistemas de reconstrução 3D baseados em visão estéreo de múltiplas perspectivas. Tal rastreamento com o objetivo de validar medições é desafiador até mesmo para o estado da arte quando a aquisição de dados não é controlada, ou são utilizados equipamentos de baixo custo.

A reconstrução 3D automática de cenas gerais obtidas através de câmeras convencionais (múltiplas perspectivas), sem aquisição controlada, é um dos objetivos da fotogrametria em visão computacional. Este objetivo é ambicioso até mesmo para padrões modernos, que necessitam de restrições para um bom funcionamento como, por exemplo, câmeras de alta precisão, sensores infra-vermelho, marcações, etc. Essas restrições tornam a aquisição cara e específica, ao passo que a falta de restrições tornam os desafios ainda mais difíceis de serem superados. Estes desafios referem-se às escolhas em larga escala de formas representação e técnicas apropriadas. A escolha dessas devem lidar com:

- a presença simultânea de materiais extremamente diferentes (por exemplo, brilhosos, refrativos e não-lambertianos);
- modelos geométricos apresentando descontinuidades, texturização e deformações em diferentes escalas;
- tipos de região (por exemplo, regiões texturizadas ou não);
- condições de iluminação desconhecidas, sombras e tons (*shades*);
- grandes diferenças de pontos de vista;
- complexidade do plano de fundo;
- número arbitrário de objetos e objetos não-rígidos;

• parâmetros de câmera desconhecidos.

Para explorar parte destes desafios, foi escolhida como aplicação específica o rastreamento de padrões na interface água-ar em escala humana representando idealmente a superfície do oceano ou a superfície de ondas aquáticas em tanques. Para este trabalho, foram utilizados dados do *LabOceano* na COPPE/UFRJ, o tanque de água mais profundo do mundo, Figura 1. Esses dados foram produzidos especificamente para esta pesquisa, em projeto patrocinado pela PETROBRAS, em colaboração direta com o Prof. Nelson Violante-Carvalho da COPPE/UFRJ.

O tipo de imagem na Figura 1 está além da capacidade dos sistemas fotogramétricos atuais, exigindo uma abordagem nova e não convencional para o rastreamento (e também para a reconstrução em 3D). Essas imagens não possuem cantos angulados ou características pontuais que possam ser descritos facilmente por nuvens de pontos, além de não conterem informação confiável em regiões lisas da imagem, que refletem e refratam fortemente, gerando reflexos especulares e transparência.

Além disso, verifica-se que as regiões suaves apresentam fortes artefatos de compressão, Figura 2, conhecidos como *blocking* por formarem blocos de cor sólida em partes com pouca variação de tonalidade. Estes artefatos dificultam a reconstrução da geometria das ondas em regiões lisas. Portanto, no presente trabalho, explora-se a representação de características baseadas em curvas, possuindo padrões próprios e consistência geométrica nas descontinuidades de brilho da imagem. Estas curvas podem ser usadas como entrada em aplicações de reconstrução 3D, para que estas imagens desafiadoras sejam reconstruídas.

Na obtenção destas curvas são explorados inicialmente um detector de borda subpíxel *bottom-up*, um clusterizador de bordas e um extrator de fragmentos de curva geometricamente consistentes. Em seguida, explora-se uma abordagem de aprendizado de máquina para treinar a topologia global dessas curvas. O aprendizado é baseado em dados de treinamento, com marcação humana, cuja finalidade é aperfeiçoar a consistência das curvas em aplicações específicas. Como o objetivo maior é construir sistemas de visão por computador e não soluções de fotogrametria excessivamente específicas, tenta-se empregar técnicas de engenharia bastante gerais que possam ser adaptadas a outros domínios. Escolhe-se aqui explorar as vantagens dos vídeos convencionais, os quais possuem baixo custo de aquisição e alta taxa de dados (*throughput*). Assim, equipamentos mais especializado são empregados apenas para a geração de medidas confiáveis (*ground-truth*) para validação posterior.

Finalmente, as motivações para esta nova abordagem podem ser mais bem definidas:

• falta de métodos verdadeiramente 3D e que funcionem bem com a aquisição de dados a baixo custo sujeita a poucas restrições;



Figura 1 - Tanque de ondas na COPPE/UFRJ

Legenda: A abordagem proposta rastreia padrões visuais difíceis no tanque de ondas de água mais profundo do mundo na COPPE/UFRJ (topo), em quatro vídeos com direções diferentes, uma das quais aqui representada (imagem inferior), para uso em fotogrametria 3D.

Fonte: O autor e LabOceano, UFRJ, 2017.



Figura 2 - Artefatos de compressão, blocking

Legenda: [Esquerda] Quadro de um dos vídeos.

[Direita] Detalhe do quadro com mudança nos níveis da imagem para melhor visualização dos artefatos de compressão *blocking*, blocos de cor sólida comum em vídeos.

Fonte: O autor, 2017.

- obtenção de resultados mais significativos com a determinação de curvas (discretas) que carregam pontos ordenados e com direção, contendo assim, mais informação do que as tradicionais nuvens de pontos;
- possível extensão para outros cenários sem mudanças significativas no processo geral;
- possível utilização como passo preliminar de outros sistemas uma grande vantagem é a filtragem da informação relevante, com alta precisão — alimentando, assim, sistemas que possuem mais restrições quanto às suas entradas.

Estes pontos são abordados e discutidos em mais detalhe ao longo do texto.

### Rastreamento de curvas e aprendizado de máquina

A extração dos contornos de uma imagem é um problema fundamental e antigo em visão computacional. No início dos anos 2000 houve uma diminuição do interesse na determinação de curvas, com uma mudança de paradigma para o uso de abordagens baseadas em pontos-chave (*SIFT*), porém mais recentemente viu-se um avivamento na busca de curvas em imagens, especialmente para a utilização no reconhecimento de objetos, Ferrari et al. (2008), Lin et al. (2009), Payet e Todorovic (2010), Zhu et al. (2008), Hu, Wu e Zhu (2011), e em aplicações de múltiplas perspectivas, Payet e Todorovic (2011), Fabbri e Kimia (2010), Usumezbas, Fabbri e Kimia (2016a), Fabbri e Kimia (2016), Fabbri, Giblin e Kimia (2012), Nurutdinova e Fitzgibbon (2015), Kuang, Oskarsson e Åström (2014).

Em geral, uma imagem contem inúmeras curvas ou bordas, decorrentes de uma



Figura 3 - Fontes físicas de curvas em imagens.

Legenda: Curvas presentes em imagens decorrem de uma variedade de fontes físicas. Fonte: Adaptação de (TAMRAKAR, 2008).

série de fatores físicos, produto direto de descontinuidades na intensidade ou na cor do sinal de uma imagem.

A Figura 3 ilustra os tipos de bordas que podem ser encontrados numa imagem genérica. Neste trabalho é chamado *edgel chain*, "contorno" ou "fragmento de curva" o agrupamento ordenado de elementos de borda (*edgels* — pontos 2D com direção e intensidade).

Neste trabalho, explora-se uma recente abordagem de múltiplos estágios para extração de curvas em imagens, Guo et al. (2014a), Guo, Narayanan e Kimia (2017), onde o espaço de busca de fragmentos de curva é iterativamente reduzido removendo-se candidatos indesejados usando restrições geométricas locais, mas sem afetar a cobertura (sinal desejado), até um ponto onde a aplicação de uma função objetivo global se torna apropriada.

O progresso de fases locais para globais é descrito por: (i) detecção de elementos de borda, (ii) construção de hipóteses para a formação de pequenas curvas locais, (iii) extração de fragmentos de curva não ambíguos, (iv) resolução de ambiguidades para gerar uma conjunto de fragmentos de curvas, (v) união de fragmentos de curva baseado em aprendizado de máquina de sinais (*features*) fotométricos e geométricos incluindo sinais de esparsidade de elementos de borda e (vi) a aplicação de aprendizado de máquina para uma seleção final de fragmentos de curva. Esses estágios podem ser agrupados em

detecção de bordas (i), formação de curvas (ii a iv) e aprendizado de máquina (v e vi). Os métodos utilizados são descritos no Capítulo 2 enquanto detalhes da aplicação destes estão no Capítulo 4.

Os fragmentos de curva obtidos originalmente com o aprendizado genérico aplicado a imagens genéricas proposto por Guo et al. (2014a), Guo, Narayanan e Kimia (2017) foram avaliados nesses trabalhos da seguinte forma:

- foi calculada uma medida de estabilidade dos fragmentos de curva quando as imagens sofrem transformações visuais, como, por exemplo, mudanças de ponto de vista, iluminação, e adição de ruído. Esta medida é um fator crítico para que fragmentos de curva sejam úteis em processos visuais subsequentes, algo frequentemente ignorado nas avaliações de outros métodos;
- foi realizada uma comparação mais tradicional de anotação humana em imagens genéricas, utilizando um novo conjunto de dados anotados, *Curve Fragment Ground-Truth Dataset* (CFGD), e outro conjunto padrão em segmentação, *Berkeley Segmentation Dataset* (BSDS).

Em ambos os métodos de avaliação, os resultados obtidos por Guo et al. (2014a), Guo, Narayanan e Kimia (2017) são mais significativos que algoritmos do estado da arte quando empregados em domínios não especializados. Portanto, decidiu-se escolher esta abordagem descrita por Tamrakar e Kimia (2007), Tamrakar (2008), Guo et al. (2014a), Guo, Narayanan e Kimia (2017).

O processo, quando utilizado em imagens genéricas tem bons resultados, mas é bastante ambíguo, pois não é específico para imagens de ondas aquáticas ou qualquer outra aplicação com domínio restrito. Encontra-se então uma oportunidade de explorar a construção de uma especialização desse sistema, para sequências desafiadoras de múltiplos vídeos de ondas aquáticas. A especialização é feita de tal maneira que a lógica geral dos processos continua genérica e pode ser adaptada para outras aplicações. Sendo assim, a especialização é feita principalmente em termos dos dados utilizados no aprendizado. Isso efetivamente fornece um idioma para expressar o sistema (em termos de características e dados de aprendizado). Porém vale destacar que, em geral, detalhes específicos devem ser projetados para cada aplicação.

### Curvas como base para reconstrução 3D

Nesta seção são revisados sistemas de reconstrução 3D comumente utilizados tanto em cenas gerais como específicas para o cenário de ondas aquáticas, e então motiva-se o uso de curvas na reconstrução, calibração e fotogrametria destas. Este sistema maior e mais abrangente, se fizer o uso do rastreador proposto neste trabalho como um de seus estágios, impõe restrições e requisitos no planejamento deste último, os quais são tomados como necessários neste trabalho.

Sistemas funcionais e completos que resolvam todos os desafios encontrados na reconstrução 3D ainda estão acima da tecnologia atual, porém tem havido progresso significativo nos últimos anos com abordagens que se encaixam em três classes gerais, dependendo de quais estruturas são utilizadas para a correlação entre imagens, associadas a pontos isolados, curvas ou partes de superfícies.

A grande maioria dos métodos de reconstrução baseados em múltiplas vistas se utilizam da correlação de pontos de interesse isolados, produzindo nuvens de pontos em 3D sem qualquer organização. A abordagem com pontos de interesse tem tido muito sucesso na reconstrução de cenas bem texturizadas em sistemas de foto-turismo em 3D e reconstruções em larga escala como Heinly et al. (2015), Pollefeys et al. (2004), Agarwal et al. (2009), Diskin e Asari (2015). Embora tenham sua utilidade, tais métodos falham em representar superfícies lisas sem textura, dada a falta de pontos de interesse nessas regiões, ou não permitirem mudanças drásticas entre as vistas.

Tais limitações dificultam a aplicação dessas técnicas, especialmente em ambientes construídos pelo homem, Simoes et al. (2012), por exemplo, em objetos como carros, Shinozuka e Saito (2014), em superfícies não-lambertianas como ondas do mar, variação na aparência causada pelo clima, Baatz et al. (2012), e em linhas de base ampla (*baseline*), Moreels e Perona (2007), estas três últimas presentes neste trabalho.

Outra abordagem faz a correspondência entre os padrões de intensidade de vistas utilizando várias vistas estéreo, produzindo nuvens de pontos mais densas ou malhas. Esta abordagem alcança reconstruções detalhadas de objetos quando a aquisição de imagens é feita por um extenso conjunto de câmeras precisamente calibradas, Furukawa e Ponce (2007), Habbecke e Kobbelt (2007), Esteban e Schmitt (2004), Goesele et al. (2007), Seitz et al. (2006), Calakli et al. (2012), Restrepo, Ulusoy e Mundy (2014). Para cenas gerais e complexas, contendo vários tipos de objetos e superfícies, e com diferentes propriedades, estas abordagens se mostram promissoras. Entretanto, imagens com áreas homogêneas, texturas repetidas, e áreas onde a intensidade é dependente da vista, se tornam desafiadoras para esses métodos. Todos esses desafios estão presentes na superfície da água.

Um número bem menor de técnicas se baseiam na correlação de curvas entre vistas para reconstrução 3D. Sistemas baseados em retas, (veja Lebeda, Hadfield e Bowden (2014), Zhang (2013), Fathi, Dai e Lourakis (2015) para avaliações recentes), e com caracterização geral por curvas, Teney e Piater (2012), Litvinov, Yu e Lhuillier (2012), Fabbri e Kimia (2010), Fabbri, Giblin e Kimia (2012), Pötsch e Pinz (2011), Berthilsson, Åström e Heyden (2001), Fabbri e Kimia (2005), foram propostos, mas alguns pecam, por exemplo, na generalidade, necessitando de modelos específicos de curva, Espino et al. (2012).

O sistema de esboço de curvas 3D, Fabbri (2010), Fabbri e Kimia (2010), e de

desenho 3D, Usumezbas, Fabbri e Kimia (2016b), Usumezbas, Fabbri e Kimia (2017), Usumezbas, Fabbri e Kimia (2016a), são o estado da arte em se tratando da reconstrução de curvas com geometria arbitrária. O trabalho atual tem o intuito de servir a estes sistemas futuramente. Veja também Fabbri, Giblin e Kimia (2012), Fabbri e Kimia (2016), Fabbri e Kimia (2005).

As várias representações para técnicas com múltiplas vistas, como documentado acima, emergem da riqueza geométrica e semântica das estruturas 3D, Zia, Stark e Schindler (2015), Feng et al. (2014). Por exemplo, um prédio contém paredes, janelas, portas, teto, chaminés, etc. Essas estruturas podem ser descritas por nuvens ou malhas de pontos, que, quando renderizadas como superfície texturizadas ou com cor sólida, são visualmente atrativas. Porém, esse tipo de representação não consegue reter informação semântica: quais áreas representam uma janela ou uma parede? quais paredes são adjacentes? Este tipo de detalhamento na representação requer fronteiras explícitas para cada parte, como cristas e junções onde estas fronteiras se encontram.

A mesma prerrogativa emerge se os objetos numa cena forem definidos apenas por suas estruturas de curvas. A representação de um prédio por suas cristas tem um apelo visual forte à sua estrutura, mas falha nos mesmos pontos de separação semântica anterior. Ambas, curvas e superfícies, são importantes e necessárias, por exemplo, em aplicações como robótica, Carlson, Vuong e Johansson (2014), e planejamento urbano e projeto industrial, Yee (2012), Leggitt (2015). Imagens de água são desafiadoras nesses quesitos, pois não há uma semântica bem definida ou topologia de curvas, pois seus padrões podem variar desmesuradamente.

Assim, o rastreador aqui apresentado almeja ser um módulo num sistema maior de reconstrução 3D por meio de curvas, estendendo-se à recuperação de superfícies (quando possível). Tais curvas reconstruídas em 3D podem ser utilizadas independentemente em aplicações como no rápido reconhecimento de cenas 3D, Pötsch e Pinz (2011), na transmissão eficiente destas cenas, no entendimento e modelagem sobre a análise nas junções, Mattingly et al. (2015), na renderização não foto realista consistente de vídeos, Chen e Klette (2014), dentre outras, Rao, Chung e Hutchinson (2012), Kowdle et al. (2012), Wang, Choi e Medioni (2014).

Em geral, curvas são atrativas por possuírem boa localização, maior invariância a mudanças de iluminação do que pontos de interesse, estabilidade com respeito a uma grande variedade de configurações estéreo, e tipicamente mais densas do que pontos de interesse. Além disso, a reflectância ou cristas proporcionam condições de contorno para a reconstrução de superfícies, enquanto oclusão de contorno entre vistas conduz a superfícies, Cipolla e Giblin (2000), Liu et al. (2008), Taubin et al. (2008). Estudos recentes reforçam a noção que curvas contêm muita informação da imagem, Koenderink, Doorn e Wagemans (2013), Zucker (2014), Kunsberg e Zucker (2014), Cole et al. (2009). Ademais, curvas são estruturalmente ricas por sua geometria diferencial, um fato explorado em sistemas recentes, Zucker (2014), Abuhashim e Sukkarieh (2012), Fabbri, Giblin e Kimia (2012), Fabbri e Kimia (2010), e em estudos de percepção, Fleming, Holtmann-Rice e Bülthoff (2011), Zucker (2014).

### Revisão das técnicas de fotogrametria de ondas aquáticas

A medição de padrões na superfície aquática apresenta os desafios anteriormente mencionados às técnicas gerais de visão computacional. Técnicas especializadas alcançaram sucesso em configurações que minimizam parte desses desafios, mas pecam na generalidade fora dessas configurações. Estas técnicas assumem algumas das seguintes restrições:

- o mar em um estado contendo textura suficiente, por exemplo, agitado;
- uma superfície que não é totalmente 3D, chamada de 2.5D em visão computacional. Como consequência, esses sistemas não conseguem lidar com auto-oclusão, o que é particularmente útil no rastreamento das cristas de onda ou filmagens com ângulos inclinados;
- condições de iluminação tal qual a superfície da água seja predominantemente difusa, e possa ser bem modelada por modelos lambertianos de iluminação. Implicações dessa abordagem são buracos ou interpolações irrealistas quando medidas pouco confiáveis são utilizadas.

Estas restrições podem ser superadas com o uso de curvas e suas características, como é exposto neste trabalho. A seguir são apresentadas, brevemente, algumas das técnicas mais utilizadas nos campos de oceanografia física e engenharia oceânica.

### Técnicas de fluxo óptico

Algumas técnicas tentam examinar o fluxo óptico e informações similares dos padrões de deformação na superfície da água a partir de um único vídeo, Rapizo et al. (2015). As restrições na aquisição devem ser projetadas tais que se possa assumir uma constância no brilho, entre outras hipóteses. Uma desvantagem dessas técnicas é que a informação de um único vídeo pode ser muito ambígua, conduzindo a sistemas rígidos e soluções especializadas, com fortes hipóteses nas condições de aquisição e câmeras pré-calibradas. Entretanto, o entendimento de um só vídeo é mais simples em certos cenários, e pode ser muito útil quando o sistema aqui proposto executa uma análise espaço-temporal. Um trabalho inicial, Amorim et al. (2013), explorou como a dinâmica do estado do mar nas praias pode ser predita e representada com aprendizado não-supervisionado através de mapas de difusão num único vídeo. Os autores apresentam ideias interessantes de como o aprendizado de máquina pode ser utilizado posteriormente na atualização do modelo e na compressão de vídeo baseadas em informações anteriores.

### Técnicas estéreo: reconstrução 2.5D

O trabalho mais utilizado no campo de oceanografia física é o de Benetazzo e Yezzi, Benetazzo (2006), Benetazzo et al. (2012), Gallego et al. (2011), Fedele et al. (2013), Fedele et al. (2012), Benetazzo et al. (2016), Gallego et al. (2013), Bergamasco et al. (2016). Esse sistema é composto por duas câmeras sincronizadas como uma cabeça estéreo posicionada em plataformas oceânicas e também em tanques de água, e apresenta vários níveis de sucesso.

Sua ideia principal é representar a superfície a ser reconstruída como o gráfico de uma função z = f(x, y), chamada de superfície 2.5D na visão computacional. Essa superfície é definida dentro de uma área de interesse e otimizada utilizando duas técnicas. A primeira, tradicionalmente de estéreo denso assumindo uma superfície lambertiana, com regiões de píxeis a serem correlacionadas ao longo de linhas epipolares (ver Apêndice A). A segunda, a otimização dessas superfícies utilizando métodos variacionais, em termos de suavidade e fidelidade de informação. Ambos produzem resultados similares ao ilustrado na Figura 4.

Como é discutido neste trabalho, essas técnicas não são adequadas para o conjunto de dados utilizado aqui, sofrendo das já mencionadas desvantagens. O conjunto de dados pode apresentar pouquíssima textura fora das cristas, e modelos variacionais com z = f(x, y) podem otimizar alucinações nas partes suaves. A abordagem proposta foca em processar apenas as partes contendo informação confiável.

### Organização desta dissertação

Este trabalho está organizado da seguinte forma: apresentação e discussão da abordagem proposta, revisando a literatura atual e motivando a nova abordagem, no Capítulo 1. O detalhamento dos métodos utilizados na extração de curvas e principalmente no aprendizado supervisionado aplicado, tendo foco apenas nos métodos, feito no Capítulo 2. A maior parte desse assunto é baseado em Guo et al. (2014a), Guo, Narayanan e Kimia (2017), Tamrakar (2008), com adoção de esclarecimentos substanciais e comentários sobre as questões relativas aos objetivos específicos deste trabalho. A aquisição e geração do Figura 4 - Técnica estéreo mais difundida na modelagem de ondas



Legenda: Técnica estéreo mais difundida na modelagem de ondas pequenas com múltiplas câmeras. Esta pode sofrer de suavização extrema ou do surgimento de buracos na reconstrução quando os dados não obedecem as hipóteses assumidas. É, portanto, inapropriada para geometrias 3D complexas.

Fonte: (GALLEGO et al., 2011; GALLEGO et al., 2013).

conjunto de dados utilizado são vistas detalhadamente no Capítulo 3, contendo os vídeos, seu processamento e a descrição das anotações manuais feitas nas referências. Finalmente os experimentos realizados são tratados no Capítulo 4, seguido das conclusões. O processo de calibração das câmeras de vídeo, para uso futuro em processamento 3D, que é a aplicação final visada por este trabalho, está descrito no Apêndice A, o qual complementa o Capítulo 3.