



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Sorocaba

PROGRAMA DE PÓS-GRADUAÇÃO em

Ciências Ambientais

BRUNO BURINI ROBLES ARINE

Detecção automática de baleias do gênero *Eubalaena* em imagens aéreas

Projeto de conclusão da disciplina de Processamento Digital de Imagens do Programa de Pós-graduação em Ciências Ambientais, 2º semestre de 2015.

Prof. Dr. Antonio Cesar Germano Martins

Sorocaba
2015

RESUMO

A espécie baleia-franca-do-atlântico-norte (*Eubalenae glacialis*) é, atualmente, representada por não mais que 500 indivíduos em habitat. Desta forma, obter informações acerca da saúde e condições de cada indivíduo é de caráter urgente para que entidades de preservação, em companhia de a comunidade científica, sejam capazes de protegê-los da extinção. Para rastrear e monitorar a população, baleias-francas são fotografadas por sondagem aérea e manualmente comparadas com um catálogo online de foto-identificação. O processo de identificação leva tempo, dificultando o rastreamento individual para amostragem biológica, gravação acústica e diversas avaliações médicas relevantes. Atualmente, há uma quantidade limitada de profissionais capazes de identificar visualmente cada indivíduo presente nas fotos quando estes emergem. O presente trabalho propõe um método automático de localização de baleias-francas em imagens aéreas. Desta forma, a composição de material de treino para uma potencial rede neural de reconhecimento de indivíduos, com a região de maior interesse de cada foto, se torna menos laboriosa e com interferência minimizada de um operador humano.

Palavras-chave: visão computacional, análise de textura, processamento, imagens digitais, aprendizado de máquina, baleias-francas.

Sumário

1. Introdução e justificativa	4
1.1. Rastreamento e monitoramento dos indivíduos	5
1.2. Automatização através de aprendizado de máquina	6
2. Objetivos	7
3. Metodologia	9
3.1. Pré-processamento	11
Escolha dos canais	11
Remoção de reflexos e sombras nas ondas	12
Correção de vinheta	14
Escolha do canal menos conturbado	16
3.2. Segmentação	18
Método por detecção de bordas	18
Método por limiarização	19
Método por agrupamento de superpixels	22
Método por similaridade de histogramas	24
3.3. Pós-processamento	25
4. Resultados e discussão	25
4.1. Desempenho	25
4.2. Comparação com outros trabalhos	28
5. Conclusões	29
6. Referências	30
7. Anexo	32
7.1. Exemplos de detecção de cada algoritmo	32

1 INTRODUÇÃO E JUSTIFICATIVA

As baleias-francas são compostas por três espécies pertencentes ao gênero *Eubalaena*: a baleia-franca-do-atlântico-norte (*E. glacialis*), a baleia-franca-do-pacífico-norte (*E. japonica*) e a baleia-franca-austral (*E. australis*). As baleias-francas têm corpos cilíndricos com tribunas arqueadas, respiráculos em forma de V e pele cinza escuro ou preta. A característica mais distinta da baleia-franca são as manchas ásperas da pele na região da cabeça, possuindo uma cor desbotada devido ao parasitismo por piolhos-de-baleia (KRAUS et al., 1986). Estas espécies costumam pesar em média 80 toneladas e medir 18 metros (MACDONALD, 2009).

Todas as três espécies são migratórias, movendo-se sazonalmente para se alimentarem ou darem a luz. As águas quentes equatoriais formam uma barreira que isola as espécies predominantes no norte e no sul. No Hemisfério Norte, as baleias-francas tendem a evitar águas abertas, preferindo permanecer próximas a penínsulas, baías ou plataformas continentais, uma vez que estas áreas oferecem maior abrigo e abundância de alimento. No Hemisfério Sul, as baleias-francas se alimentam longe da costa no verão, mas uma grande parte da população permanece em águas costeiras no inverno (BRILLANT et al., 2015).



Figura 1: Baleia-franca checando de perto um mergulhador (Foto: Bran Skerry).

As baleias-francas foram, durante muito tempo, alvo preferencial de baleeiros devido a sua natureza dócil, movimentos lentos, tendência em permanecerem próximos à costa e seu alto teor de gordura, o que as faz flutuar quando mortas e que proporciona alto rendimento no processamento de óleo de baleia.

Hoje, as baleias-francas do Atlântico Norte e do Pacífico Norte estão entre as baleias mais ameaçadas de extinção (KRAUS et al., 2005). As populações ocidentais de ambos os oceanos contam com apenas algumas centenas de indivíduos, sendo a população ao norte do Pacífico oriental considerada criticamente ameaçada, contando com menos de 50 indivíduos (ZERBINI et al., 2015).

Embora as leis tenham se tornado mais severas e programas de proteção tenham sido implementados, o homem continua a ser, de longe, a maior ameaça para estas espécies. Atualmente, as duas principais causas de morte das baleias-francas são colisões com navios e emaranhamento em redes de pesca. (FUGIWARA & CASWELL, 2001) (LAIST et al., 2014)

1.1 Rastreamento e monitoramento dos indivíduos

Os métodos tradicionais pelos quais a abundância da população cetácea é estimada dependem da contagem de baleias a partir de encontros visuais com aeronaves e navios, ou

contagens em avistadouros em terra (EBERHARDT et al., 1979). Em estudos breves de amostragem (em torno de uma semana), baleias-francas são rastreadas e monitoradas através de telemetria via satélite, sobretudo na banda do infravermelho, onde mais se destacam (MATE et al., 1997), algumas vezes complementada pela banda “costal”, representada pelo azul mais escuro característico de águas profundas (FRETWELL et al., 2014).

Todavia, este método de rastreamento possui sérias limitações. Uma vez que as baleias se espalhem e saiam do campo do sensor do satélite, torna-se impossível localizá-las individualmente de novo. Zerbini et al. (2015) menciona que alguns poucos indivíduos possuem transmissores acoplados no corpo, permitindo que sejam rastreados por um período mais prolongado (na ordem de alguns meses). Para estudos de longa duração, alguns anos de monitoramento são necessários e, para tal fim, o rastreamento físico se torna impraticável. Técnicas de sensoriamento remoto, sobretudo havendo a possibilidade de utilização de satélites de alta resolução, surgem como uma alternativa promissora para tal tipo de estudo.

1.2 Automatização através de aprendizado de máquina

O rastreamento automático de baleias de diversas espécies não é um tema particularmente restrito. Métodos não visuais são comumente empregados, como os modelos de Gillespie (2004) e Yack et al. (2010), que utilizam métodos estatísticos e redes neurais respectivamente para a identificação de uma determinada espécie de baleias através de espectros acústicos. Contudo, programas para abordagens visuais existem, por exemplo, para acompanhar movimentos migratórios de certos grupos através de câmeras térmicas (NOAA, 2015). Para rastreamento individual, conta-se com monitoramento dos oceanos por satélites como o WorldView2, que possui poder de resolução de 50 cm e 8 bandas espectrais, encompassando desde o espectro visível até várias frações do infravermelho. Tal equipamento se mostrou, como já foi experienciado, consideravelmente útil para o estudo de animais marítimos remotamente (FRETWELL et al., 2014).

O processo de identificação individual, contudo, exige um alto grau de especialização de profissionais da área de oceanografia. Os detalhes que diferenciam um indivíduo de outro são altamente sutis, dependendo da catalogação destes detalhes em um almanaque de referência que, na grande maioria das vezes, arca com um grande tempo de espera entre o momento de avistamento *in situ* e a publicação da entrada atualizada no almanaque.

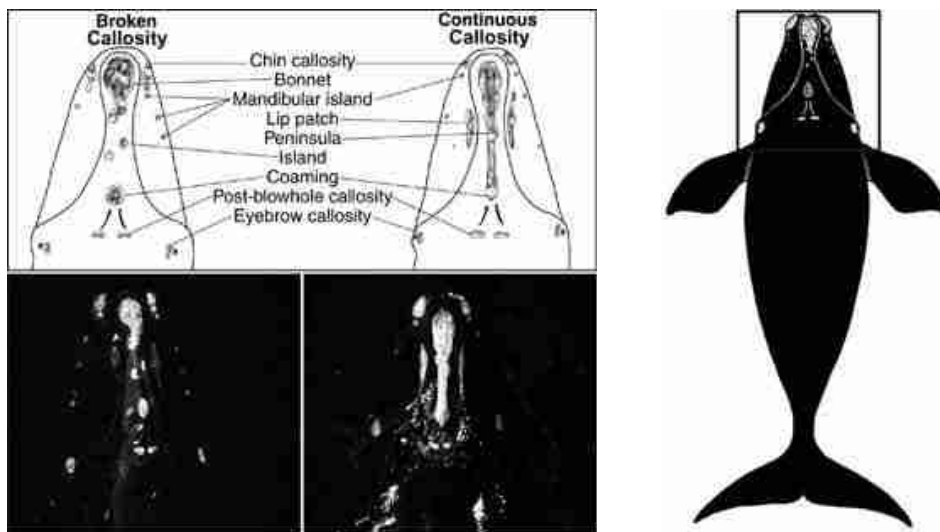


Figura 2: Exemplo de página de catálogo de identificação de baleias-francas (“Catálogo No. 1233 and Catalog No. 1027”, NEW ENGLAND AQUARIUM, 2015)

2 OBJETIVOS

Tendo em vista o processo exaustivo de identificação manual de indivíduos das baleias-francas, um classificador automático seria de extrema valia para programas de proteção ambiental voltados a esta espécie. Classificadores automáticos baseados em aprendizado de máquina se provaram empiricamente eficazes nas mais diversas tarefas de reconhecimento. A utilização destes para a identificação individual de baleias parece, não somente lógico, como o método mais apropriado para a tarefa, dependendo também de um cuidadoso processo de pré-processamento das imagens.

Para todo modelo de aprendizado de máquina, independente da tarefa para a qual ele foi designado, o grau de desempenho está diretamente relacionado com a quantidade e qualidade do material de treino. A quantidade, obviamente, é dada pelo tamanho da população que compõe o material. A qualidade refere-se à relevância dos parâmetros que são fornecidos e a proporção de sinal-ruído das amostras (representado pelo número de falsos positivos detectados).

Logo, neste trabalho desejou-se avaliar diferentes técnicas de processamento digital de imagens para automatizar a localização das baleias nas fotos disponibilizadas pela plataforma Kaggle como alternativa à segmentação manual e laboriosa das baleias em todas as fotos disponibilizadas para o estudo. A delimitação da região de interesse nas fotos para as proximidades da baleia otimiza os resultados de um futuro algoritmo de busca pela cabeça; uma vez que não seja necessário varrer a foto inteira, diminui-se o tempo de execução e as chances de lidar com falsos positivos.

3 METODOLOGIA

Os algoritmos foram escritos na linguagem Python 2.7.6 utilizando as bibliotecas Scikit-Image, OpenCV e NumPy. Ambas são amplamente utilizadas para pesquisa e produção (OLIPHANT, 2007) (VAN DER VALT et al., 2014). A biblioteca Scikit-Image é composta de inúmeras funções relacionadas ao processamento de imagens (histogramas, máscaras de convolução, operações morfológicas, estatísticas de ordem n, detecção de formas etc.), implementadas a partir de artigos científicos abertos ao público. A biblioteca OpenCV é originalmente compilada em linguagem C, com ganchos para utilização dentro do ambiente Python. Esta biblioteca tem ganhado destaque entre pesquisadores de visão computacional por seu alto desempenho e quantidade de funções disponíveis. A biblioteca NumPy é um pacote genericamente voltado para computação científica de alto desempenho com funções matemáticas diversas, compiladas a partir de FORTRAN e C, mas com ganchos em Python e otimizadas para operações vetoriais.

As imagens utilizadas durante o processo de pesquisa dos métodos mais adequados e da elaboração geral do algoritmo foram compostas por um subconjunto de 100 imagens disponibilizadas pelo site Kaggle (2015) em conjunto com o NOAA (National Oceanic and Atmospheric Administration), onde mais de 8 gigabytes de imagens são disponibilizadas

gratuitamente em decorrência de um concurso proposto para a elaboração de algoritmos capazes de identificar individualmente cada espécime contida nas fotos. Todas as fotos foram reduzidas para um tamanho aproximado de 800 pixéis de altura por 1000 pixeis de largura.

O esquema geral do projeto proposto neste trabalho é mostrado no fluxograma da Figura 3.

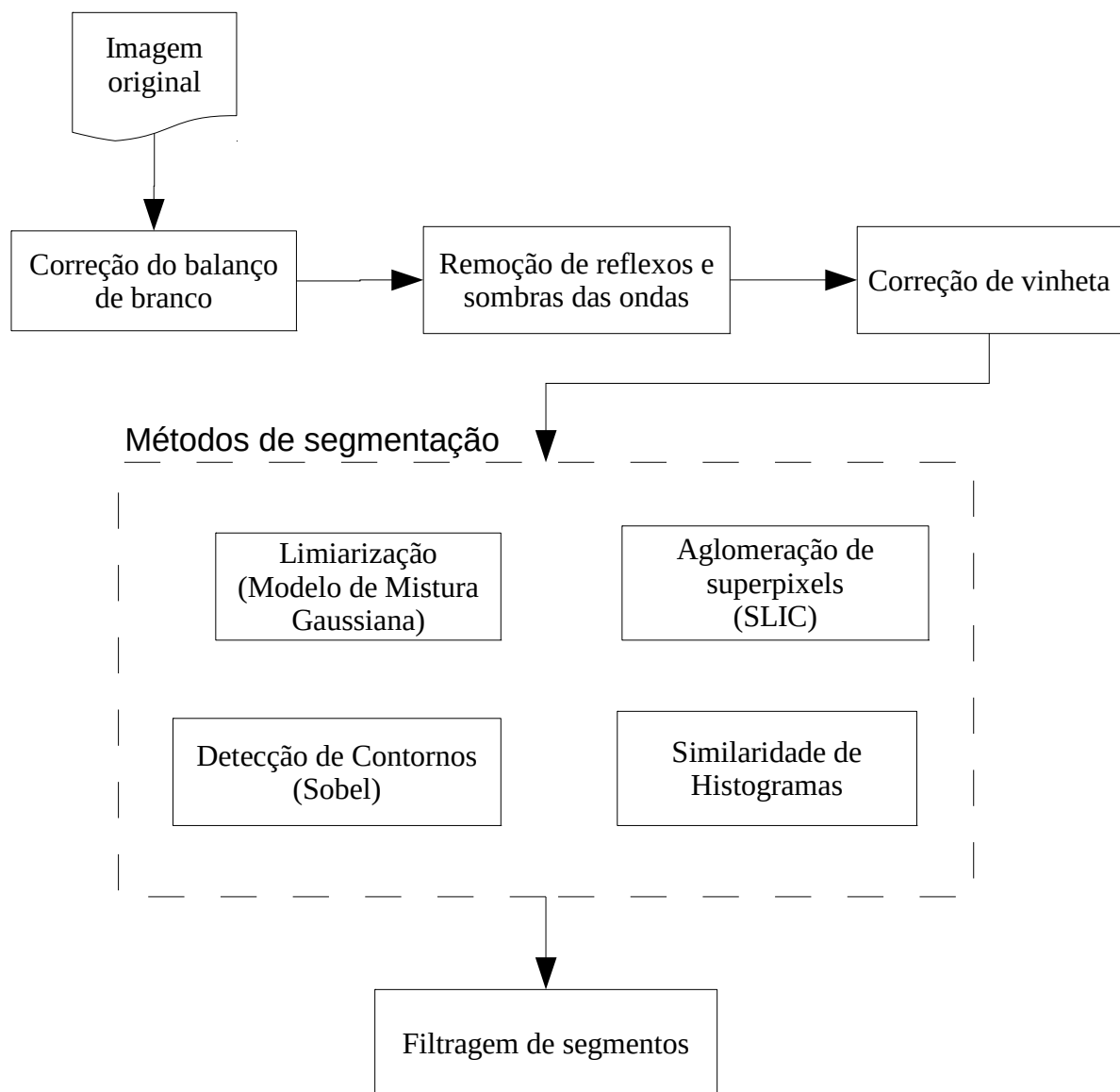


Figura 3: Esquema geral do funcionamento do programa.

Foram avaliados quatro maneiras diferentes de localizar as baleias nas imagens: busca de contornos, limiarização, aglomeração de superpixels e busca por similaridade de

histogramas. O tempo de execução de cada algoritmo foi medido e levado em consideração nas discussões finais.

3.1 Pré-processamento

3.1.1 Escolha dos canais

O espaço de cores definido pelos canais R, G e B é sensível à variação de iluminação, assim como a presença de sombras e reflexos no ambiente fotografado. As fotos disponíveis para este trabalho, no entanto, foram tiradas em condições distintas de clima, hora do dia, iluminação e ângulo de incidência do sol. Além destes fatores naturais, nota-se a câmera provavelmente não é a mesma, já que algumas imagens possuem resoluções diferentes entre si.

Para garantir que haja um padrão no brilho e matiz geral das imagens, preferiu-se dar prioridade a espaços de cores que sejam invariantes à maioria desses fatores. A tabela 1 relaciona a qualidade de invariância de alguns desses espaços de cores.

Canal	R	G	B	I	V	L	a	b	S	Rn	Gn	H
Iluminação	-	-	-	-	-	-	+/-	+/-	+	+	+	+
Sombras	-	-	-	-	-	-	+/-	+/-	+	+	+	+
Brilho intenso	-	-	-	-	-	-	-	-	-	-	-	+

Tabela 1: Propriedades de invariância dos canais individuais em ordem crescente de invariância. O sinal de "-" significa totalmente variante, "+/-" parcialmente invariante e "+" totalmente invariante à propriedade correspondente.

Alguns exemplos podem ser vistos nas figuras abaixo. Durante os testes, todos os modelos apresentaram um desempenho satisfatório com os canais H e S do espaço HSV, o canal a do espaço Lab, e o canal Rn do espaço RGB Normalizado.

A este fenômeno se atribuiu a hipótese de que uma vez água do mar sendo intensamente azulada em relação à baleia, estes elementos aparecerão de forma distinta no canal H devido à diferença de matiz. Em fotos tiradas em dias ensolarados, a água exibe uma

coloração consideravelmente mais intensa, produzindo pixels mais saturados e consequentemente com valores mais altos do que os pixels da baleia no canal S. O canal R nem sempre se mostra como a melhor opção para a segmentação da imagem, porém, em certos casos a região da água se encontra quase totalmente enegrecida neste canal, por não haver tons de coloração vermelha na água. Este mesmo efeito ocorre no canal a do espaço Lab, uma vez que este espaço separa o vermelho do azul nos canais a e b respectivamente, havendo ainda a vantagem de variações de iluminação permanecerem isoladas no canal L.

3.1.2 Remoção de reflexos e sombras nas ondas

Dependendo das condições meteorológicas e da hora do dia, os raios incidentes do sol nas ondulações da água produzem sombras e reflexos que dificultam o processo de segmentação, sobretudo na etapa onde se tenta encontrar as bordas da baleia na imagem (Figura 4).

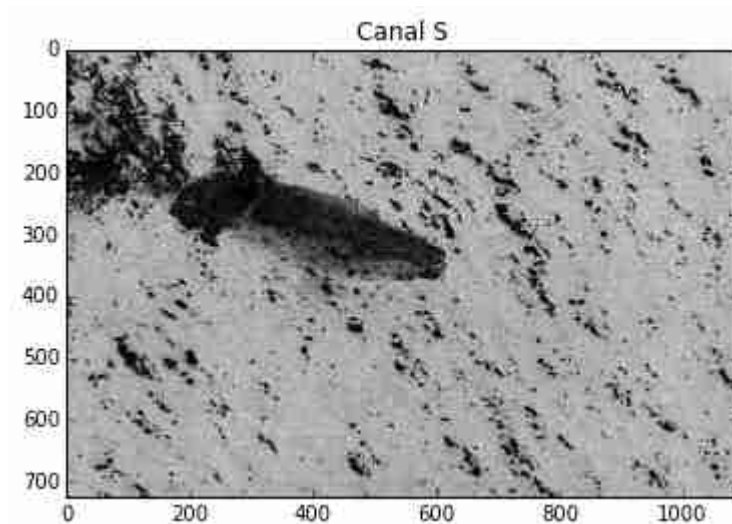


Figura 4: A intensidade da diferença entre os artefatos gerados pela ondulação da água prejudicam a segmentação da imagem.

Nota-se claramente que estes artefatos indesejados na imagem diferem do corpo da baleia em tamanho e frequência; enquanto a há apenas uma baleia por foto ocupando uma região considerável da imagem, as sombras e réstias são numerosas e finas. Uma maneira de

eliminá-las seria através de filtragem de banda no domínio das frequências da imagem, seja pela transformada de Fourier ou wavelet. Um exemplo de como tal processo pareceria é exibido na Figura 5.

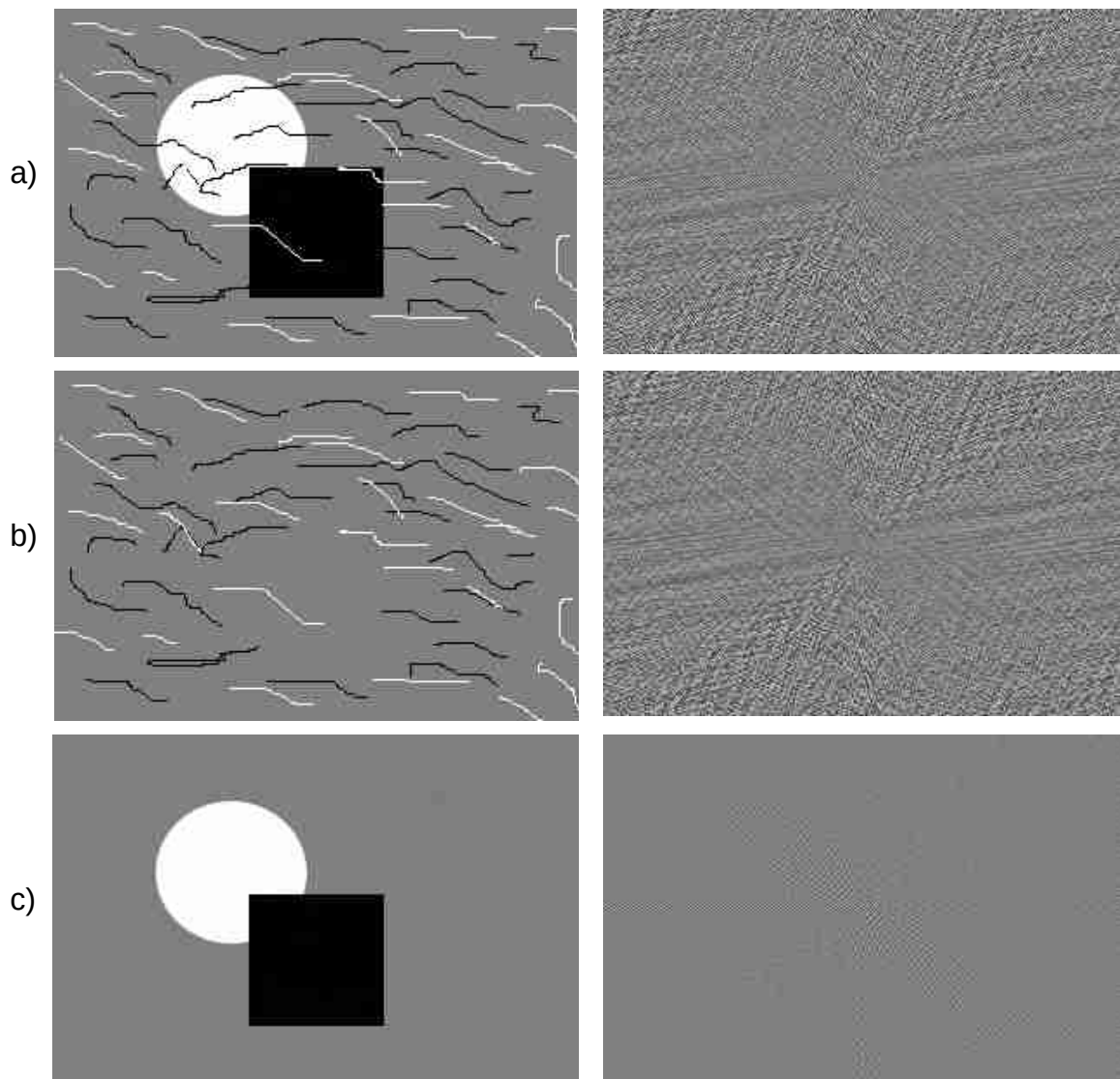


Figura 5: Simulação de transformada de Fourier em 2D em a) uma imagem com formas geométricas e linhas de interferência, b) linhas de interferência somente e c) formas geométricas somente.

Ao se aplicar a transformada rápida de Fourier em uma foto de verdade e obter o espectro de potência (elevando-se ao quadrado a parte real e imaginária do espectro), o resultado não é intuitivo o suficiente para que saibamos qual região do espectro neutralizar para que as interferências desejadas sejam eliminadas (Figura 6).

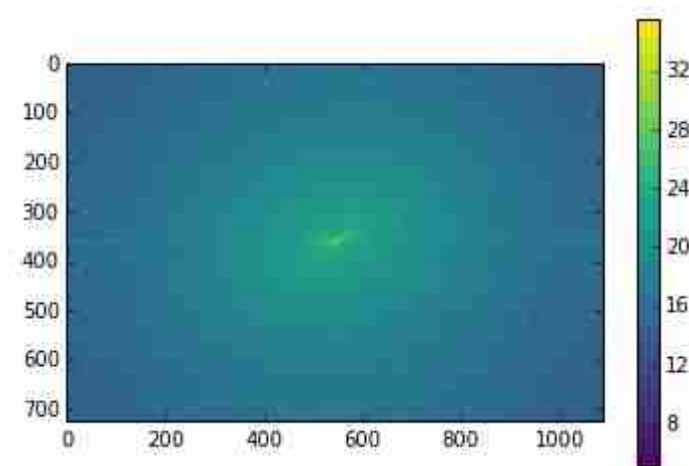


Figura 6: A transformada de Fourier do canal S da imagem não exibe imediatamente pistas de como eliminar as ondulações da água.

Desta forma, para fins de simplicidade, optou-se por aplicar um filtro mediano pelos canais selecionados antes da execução dos processos de segmentação, certificando-se que a máscara utilizada fosse de tamanho maior do que o tamanho das sombras e réstias que se desejaria eliminar.

3.1.3 Correção de vinheta

Dado o ângulo de incidência e natureza do equipamento fotográfico utilizado para adquirir as imagens, em algumas das fotos há um acentuado sombreamento próximo às bordas da imagem. Este sombreamento é conhecido como vinheta, facilmente observado sobretudo no canal V (intensidade) como um degradê escuro. Este fenômeno dificulta o processo de busca por um limiar global de limiarização da imagem, uma vez que o histograma de valores de intensidade passa a arcar com uma região extra, formada pelos pixels do sombreamento, como é possível observar na Figura 8.

Apesar de alguns dos processos terem se mostrado robustos à presença de vinhetas, manteve-se em mente a hipótese de que a remoção das vinhetas não causaria prejuízo a nenhum dos algoritmos.

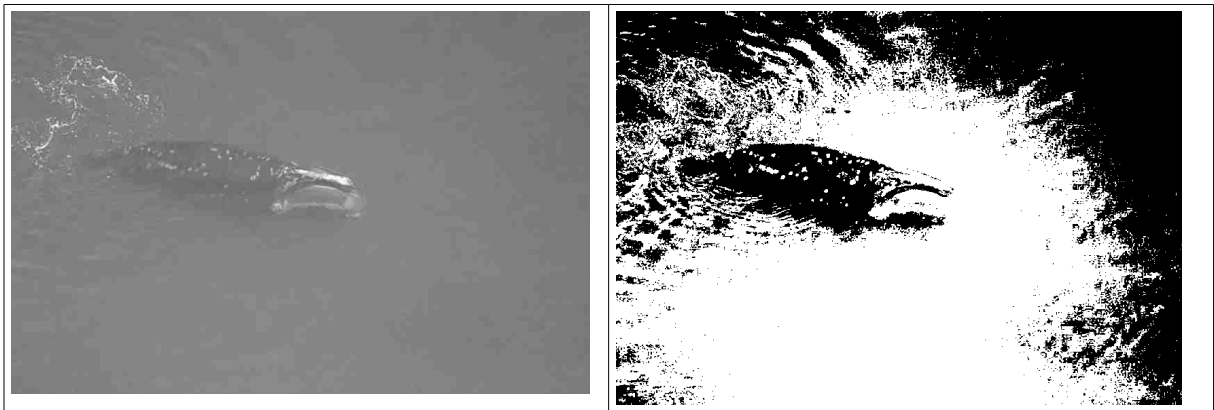


Figura 7: Exemplo de binarização por método de Otsu em imagem com vinheta.

Normalmente, a correção de vinheta é feita através da subtração dos valores dos pixels da imagem original pelos valores dos pixels de uma imagem de um fundo branco obtida com o mesmo equipamento fotográfico e sob as mesmas condições de iluminação. Caso não haja imagens de referência disponíveis, o processo de correção da vinheta, dependendo do objetivo final, se torna amplamente mais complexo. Um dos métodos, por exemplo, envolve a utilização de mistura de modelos gaussianos para determinar a natureza de cada pixel e obter a separação entre a vinheta e a imagem em si (CHEN et al., 2013).

No entanto, para fins deste trabalho, optou-se por uma solução significativamente mais simples. Utilizando o canal V como referência, a imagem foi borrada através de convolução em 2D com uma máscara gaussiana. Imediatamente, percebe-se que a região onde há sombreamento se destaca do restante da imagem, que possui iluminação mais clara (Figura 8b).

Notou-se experimentalmente que este filtro gaussiano produziu um degradê mais sutil pela própria natureza da máscara, ao contrário da máscara quadrada do filtro médio, que produz bandas distintas na imagem.

A correção da imagem se passa inteiramente no canal V, através da fórmula

$$O = I - I \circ G + 128$$

onde I é o canal V e G é a máscara gaussiana de convolução. Um exemplo de resultado pode ser visto na imagem a seguir



Figura 8: a) imagem severamente afetada pelo efeito vinheta, b) imagem original após aplicação do filtro gaussiano, c) subtração.

Percebe-se que a correção não é precisa, mas ameniza consideravelmente o efeito causado pela vinheta durante o processo de limiarização. Como consequência, o contraste da imagem é visivelmente aplainado, embora não de forma significativa, sendo este facilmente normalizado *a posteriori*.

3.1.4 Escolha do canal menos conturbado

Como o algoritmo de limiarização e detecção de contornos utiliza somente um canal como entrada, decidiu-se criar um método para que fosse escolhido automaticamente o canal menos conturbado para a tarefa.

A definição de conturbação de uma imagem é arbitrária e sujeita a diversas interpretações. O que de fato se desejou, avaliando os diferentes canais da imagem, foi encontrar aquele no qual a separação entre elementos de frente (baleia) e fundo (água) fosse menos complexa possível.

Não existe um consenso sobre qual seria a medida mais objetiva que traduzisse o grau de conturbação de uma imagem. Uma primeira tentativa de fazê-lo foi baseando-se em estatísticas de segunda ordem calculados a partir da matriz de coocorrência: objetos fotografados tendem a ter uma correlação pixel a pixel que, por definição, não está presente em ruído branco. Todavia, o ruído branco não necessariamente causa transtorno durante o

processo de segmentação da imagem; sobretudo porque este é trivialmente removido com a aplicação de um filtro médio, por exemplo. Assim, a correlação como medida de conturbação não é apropriado neste caso.

No contexto deste trabalho, percebeu-se que as baleias de algumas imagens estavam ora em meio uma grande turbilhão de espuma devido à movimentação, ora parcialmente submersas; em ambos os casos, mesmo sem a presença de ruídos de fundo na imagem, o algoritmo mostrava um desempenho pouco satisfatório na etapa de segmentação.

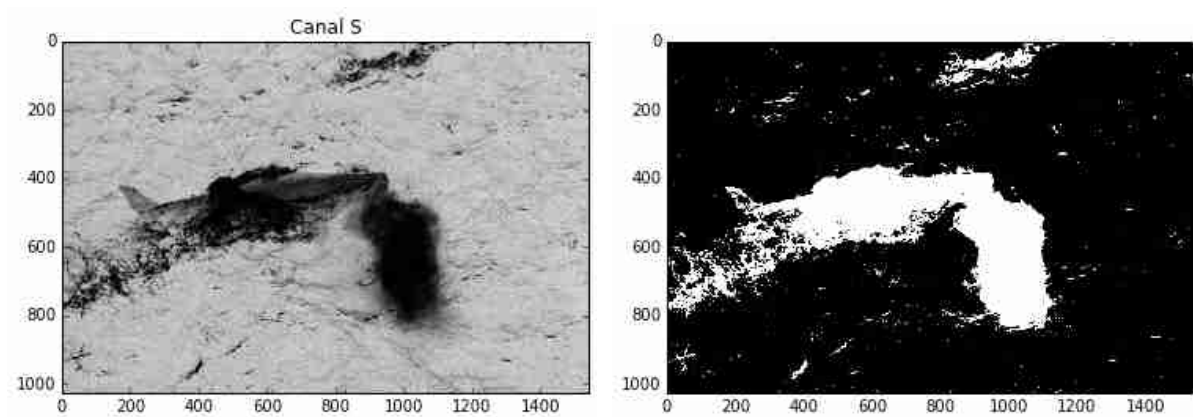


Figura 9: Vapor de água expelido pelo espiráculo da baleia confunde o algoritmo de limiarização.

Em vista deste fato, optou-se por simular uma detecção crua de bordas em cada um dos canais candidatos (como citado anteriormente, canais com maior invariância à diferença de iluminação tiveram maior prioridade) através do cálculo do gradiente da imagem. Uma vez detectadas as bordas, escolhe-se aquela onde a soma dos pixels da imagem for menor. Em termos práticos, isto significa escolher o canal que possui menos interferências.

3.2 Segmentação

3.2.1 Método por detecção de bordas

Após a verificação de qual canal é menos conturbado (dentre H, S e R normalizado), para este método), calcula-se o gradiente da imagem o qual, em seguida, é submetido a binarização através do método de Otsu, como é mostrado na figura 10.

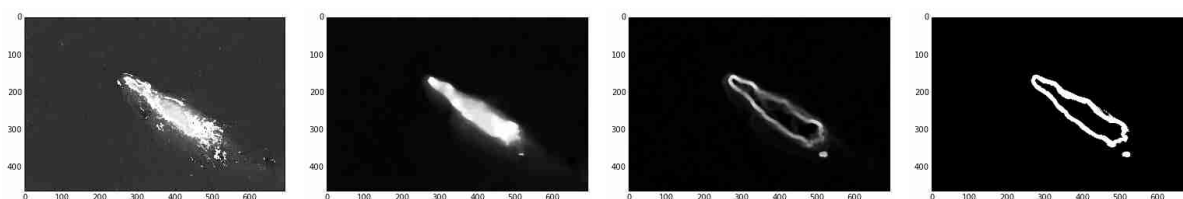


Figura 10: O canal mais apropriado foi selecionado, neste exemplo sendo o canal H. Em seguida ele é filtrado, tem seu gradiente calculado e limiarizado em seguida.

Caso a imagem de referência escolhida tenha sido o canal R normalizado, esta será submetida a um processo diferente de detecção de bordas. Isto se deve ao fato do canal r ser naturalmente mais complexo do que o canal H ou S quando estes estão propícios para a segmentação da baleia, ou seja, quando mostram precisamente a silhueta do animal e nada mais. O mesmo não ocorre com o canal r, por menos conturbado que este canal esteja em relação ao H ou S. Devido a isto, o canal r é tratado novamente com um filtro mediano para remoção secundária de elementos menores e, em seguida, sujeito ao filtro Canny de detecção de bordas. O filtro Canny é visto atualmente como um dos algoritmos mais eficientes para a tarefa.

A imagem resultante conterá diversas bordas provavelmente não contínuas ao longo da imagem. Elas serão formadas em torno de vários elementos da imagem, sobretudo da baleia em si, mas também de pequenos artefatos produzidos por sombras, réstias, espumas, sujeiras e outras interferências. A correção das falhas nas linhas que compõem as bordas dos elementos da imagem são corrigidos através do método morfológico de fechamento da imagem limiarizada.

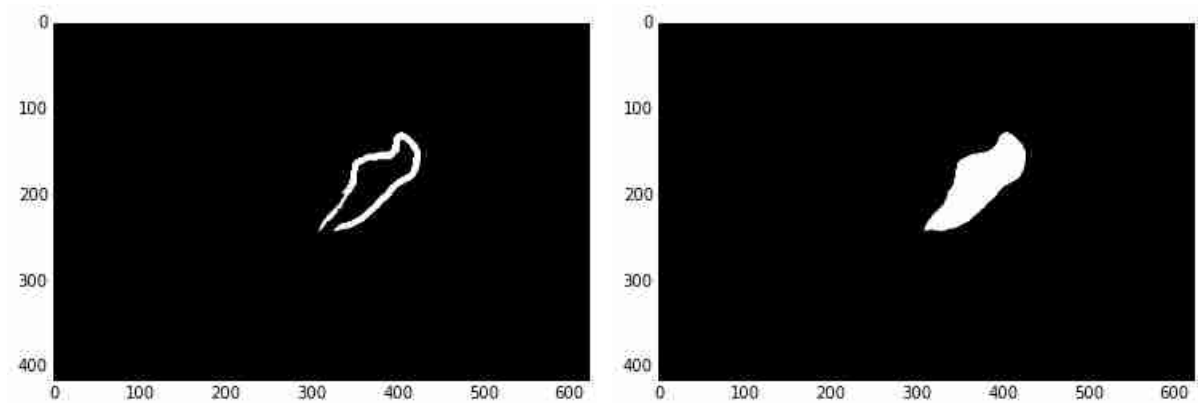


Figura 11: A dilatação seguida de erosão dos elementos da imagem faz com que vãos presentes entre um ponto e outro sejam conectados, permitindo o preenchimento da figura formada.

Após a correção das falhas nos contornos e tornando-os linhas contínuas, o preenchimento das formas se torna trivial e, independente do método utilizado, o resultado final será o mesmo. No caso deste trabalho, o algoritmo de preenchimento consiste em invadir a forma complementar da imagem através do processo de dilatação. Buracos não estão conectados às bordas e, portanto, não são invadidos. O resultado é o subconjunto complementar da região invadida.

3.2.2 Método por limiarização

Ao contrário do método anterior, neste caso a imagem é segmentada diretamente do canal selecionado, sem passar pelo processo de detecção de bordas. Baseando-se na hipótese de haver elementos de fundo e de frente separáveis na imagem apenas pelo valor dos pixels, o método consiste em encontrar o limiar ótimo para que o máximo de separação ocorra. Foram feitos testes com o canal *a* do espaço Lab e *H* do espaço HSV, ambos com ótimas propriedades invariantes. Os resultados mostraram que dentre eles, o canal *a* é o mais apropriado para a separação da baleia do elemento de fundo.

A limiarização se beneficia se transformarmos os valores dos pixels, de inteiros de 8 bits para decimais de 32 bits, e em seguida aplicarmos um filtro de esmaecimento como o

filtro de média ou gaussiano. Este tem o papel de interpolação dos valores discretos dos pixels, melhorando o desempenho do algoritmo do método de limiarização.

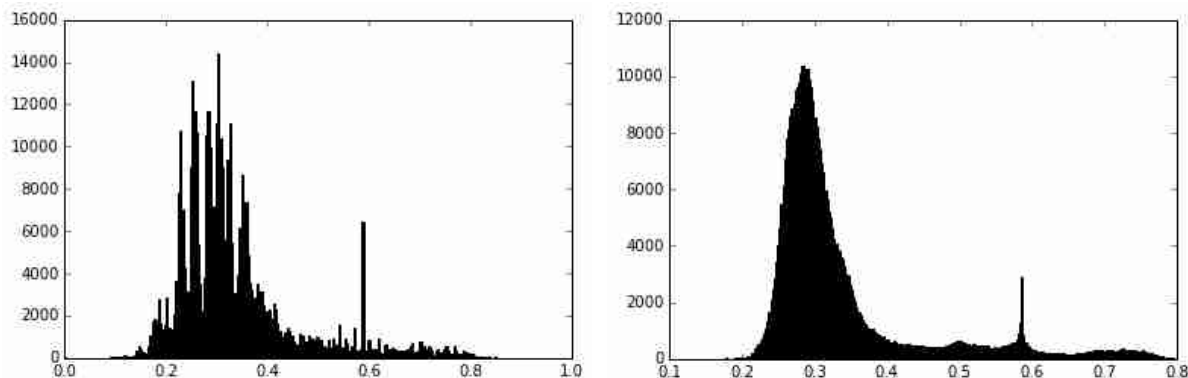


Figura 12: À direita, histograma da imagem original. À esquerda, histograma da imagem após filtro gaussiano com núcleo de tamanho 3 pixels. Nota-se como as formas de picos gaussianos são melhores distintos após o esmaecimento.

O método de Otsu para cálculo do limiar, como mencionado, assume que a imagem contenha duas classes de pixels, cujo histograma assume a forma de uma distribuição bimodal. O algoritmo realiza uma procura exaustiva de tal forma que a variância intraclasse seja minimizada, isto é, a soma ponderada das variâncias das duas classes (OTSU, 1979).

Entretanto, o método de Otsu é incompatível com certas imagens cujos histogramas são distribuições multimodais. Acredita-se que a natureza destes se deva à presença de quatro elementos distintos na imagem: a água, a baleia, as sombras e os reflexos que não puderam ser removidos no pré-processamento, seja porque os valores dos pixels são muito próximos aos da baleia, seja porque o tamanho das impurezas na foto são maiores que a máscara utilizada para filtrá-las.

Este problema pode ser contornado através da generalização do método de Otsu para mais de um limiar, conhecido como Método de Otsu Multiníveis. Liu e Yu (2009) mostraram que o método de Otsu multiníveis é totalmente compatível com métodos de clusterização, sendo alguns deles mais eficientes do ponto de vista computacional. Logo, ao invés do método de Otsu, a segmentação do canal foi feita utilizando um modelo de mistura gaussiano.

Este é um modelo probabilístico, o qual assume que a distribuição dos dados é formada a partir da mistura de um número finito de distribuições gaussianas de parâmetros desconhecidos. Uma vez que o histograma dos valores dos pixels de cada elemento principal da imagem deve seguir aproximadamente uma distribuição normal, o modelo de mistura gaussiana pareceu ser mais apropriado que outros métodos de clusterização, como o método de k-médias, por exemplo, o qual assume que as seções calculadas têm dimensões equivalentes.

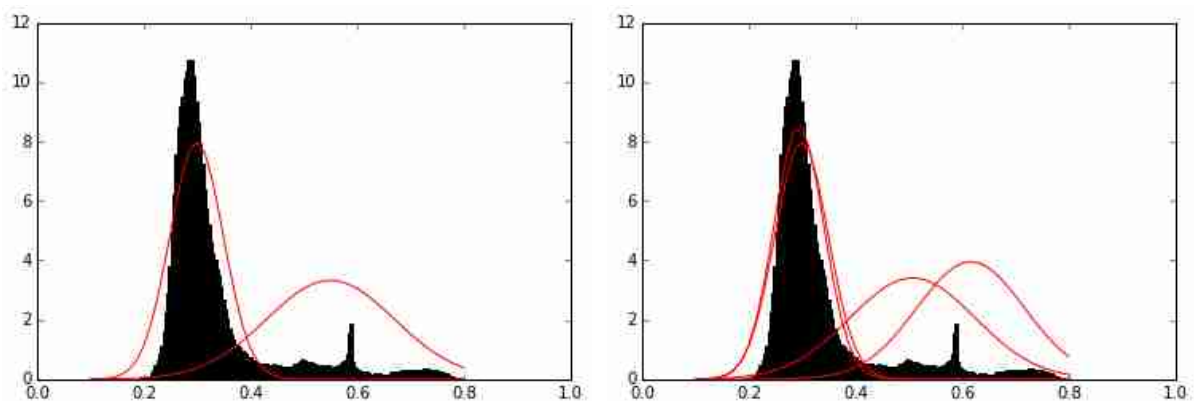


Figura 13: Exemplos de como o histograma é decomposto no modelo de duas gaussianas (esq.), pelo método de Otsu bimodal, e quatro gaussianas (dir.), pelo modelo de mistura gaussiana. O limiar foi estimado em 0.40 e 0.57 respectivamente.

Pelo modelo de mistura gaussiana, tomou-se como referência as curvas gaussianas que representam os dois grupos de pixels mais intensos. Devido às baleias se destacarem no canal **a** de forma mais clara do que todos os outros elementos da imagem, o limiar é calculado como o ponto de intersecção mais alto desses dois grupos (que pode ser calculado aproximadamente como a média dos centroides de cada um dos grupos). A diferença nos dois modos de limiarização pode ser observado na figura 14.

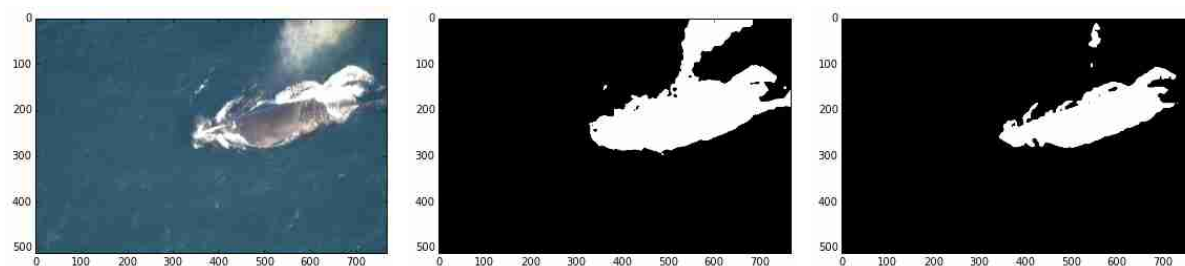


Figura 14: Imagem original, limiarizada pelo modelo bimodal e pelo modelo multimodal respectivamente.

O modelo multimodal permite o cálculo de um limiar mais preciso para a separação da baleia de todo o restante da imagem. Nota-se, inclusive, que o limiar foi tal que o próprio vapor de água expelido pela baleia foi deixado de fora.

3.2.3 Método por agrupamento de superpixels

A ideia deste método de segmentação foi primeiramente proposto por Ren e Malik (2003) e popularizado por Achanta et al. (2010), através do algoritmo chamado SLIC (*Simple Linear Iterative Clustering*). Este algoritmo realiza uma super-segmentação da imagem através do modelo de k-médias e utilizando como referência todos os valores dos pixels de todos canais do espaço de cores Lab.

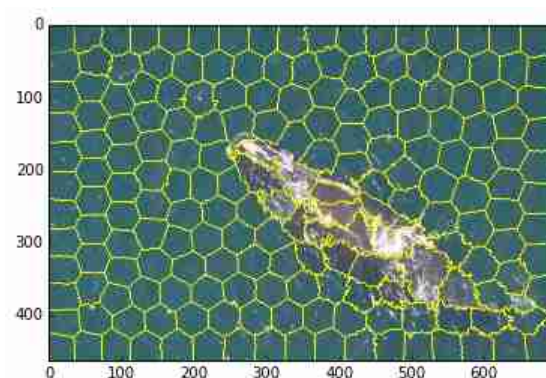


Figura 15: Exemplo de criação de superpixels pelo algoritmo SLIC.

Os pixels mais semelhantes entre si são agrupados em um “superpixel”, podendo ou não estarem conectados uns aos outros, de acordo com os parâmetros impostos ao algoritmo. Por utilizar k-médias, este algoritmo de criação de superpixels é um dos mais eficientes

computacionalmente (ACHANTA et al., 2012). Contudo, mesmo assim o tempo de execução foi demasiadamente alto para a aplicação deste método de segmentação em 100 imagens de teste. A redução da imagem para 20% do tamanho original diminuiu o tempo de execução sem prejudicar a eficiência consideravelmente.

Após a criação dos superpixels, estes são agrupados em três grupos maiores através de um algoritmo de clusterização aglomerativa, no qual realiza uma clusterização hierárquica dos superpixels baseando-se na distância euclidiana. Esta etapa é sugerida por Pantofaru e Herbert (2005), cujo trabalho utiliza um algoritmo de superpixels chamado “mean-shift”, antecessor do SLIC.

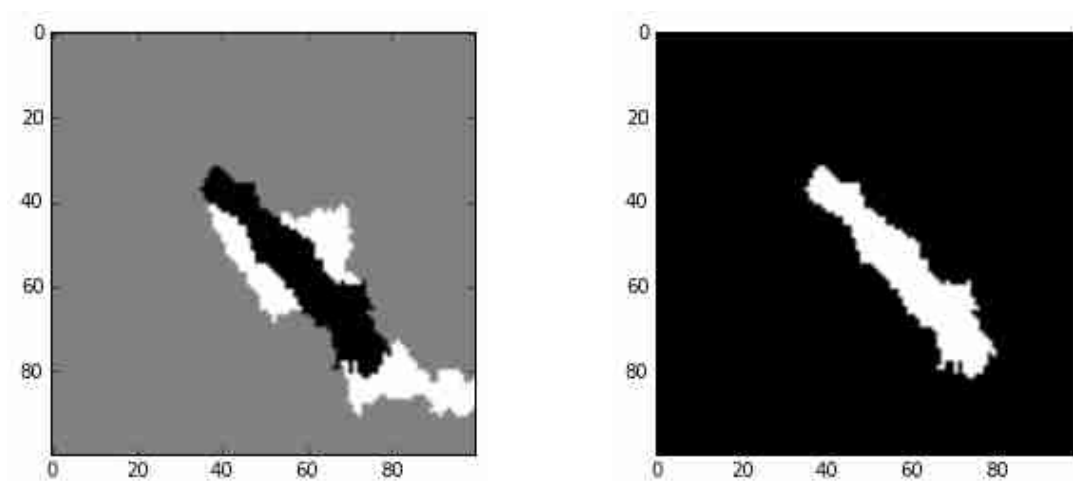


Figura 16: Número de superpixels reduzidos por clusterização aglomerativa (esq.) e seleção do cluster mais compatível à iluminação da imagem original (dir.)

Após o processo de redução do número de clusters, o cluster mais relevante entre os 3 estimados é selecionado. Na grande maioria das fotos investigadas, o corpo da baleia possui pixels de maior valor no canal R do que nos outros canais. Convertendo o espaço RGB para RGB normalizado, para garantir que o método de seleção seja padronizado em diferentes iluminações, seleciona-se o segmento cuja razão dos pixels do canal R normalizado em relação ao restante da imagem é o maior (Figura 16).

3.2.4 Método por similaridade de histogramas

Este método é mais simples que os métodos anteriores, porém mais robusto em relação a características variantes entre uma imagem e a outra. Este algoritmo se baseia em dois fatos provenientes da natureza das fotos em questão: a) a baleia ocupa uma área pequena da imagem e b) a maior parte da imagem, que é ocupada pela água, é praticamente uniforme.

Partindo destes dois princípios, o algoritmo divide a imagem em quatro janelas e calcula o histograma de cada uma delas no espaço HSV, uma vez que este é parcialmente invariante a diferentes condições de iluminação, mas retorna histogramas diferenciáveis entre regiões. Se uma delas possuir um histograma estatisticamente diferente do histograma que gerou essas 4 janelas (determinado pelo coeficiente de correlação entre os valores de cada banda), esta janela é marcada em uma máscara e então ele subdivide a janela em mais 4 janelas e assim por diante.



Figura 17: Exemplo de busca recursiva por regiões com histograma diferente em relação ao histograma da imagem completa.

Graças à recursividade do algoritmo, a varredura se torna mais eficiente, pois a computação dos histogramas se foca nas regiões (e sub-regiões) que possuam somente pixels de interesse, em vez de pixels de fundo.

3.3 Pós-processamento

Para todos os métodos de segmentação utilizados, a etapa seguinte consiste em eliminar segmentos pequenos através da operação morfológica de abertura (na qual é efetuada uma erosão seguida da dilatação da imagem).

Mais de um segmento pode permanecer na imagem resultante nesta etapa. Supondo que a baleia seja o maior elemento de frente contido em todas as imagens, basta filtrá-la de modo que apenas o maior segmento permaneça. A detecção e cálculo de cada mancha presente na imagem é feita através do processo de crescimento de regiões. Neste processo, a imagem binarizada é varrida até que seja encontrado um pixel do valor desejado (no caso, 1). A partir deste ponto, a imagem é varrida bilateralmente para que seja estimado até onde se propaga este segmento de valor 1. Após ter varrido toda a extensão contínua de valores 1 e haver localizado a extensão da mancha, esta tem a área contabilizada e seus pixels removidos da imagem. O algoritmo então continua sua busca a partir do ponto anterior.

4 RESULTADOS E DISCUSSÃO

4.1 Desempenho

Não existe um consenso na comunidade da visão computacional sobre qual a melhor maneira de avaliar segmentações (ÖZDEMİR et al, 2010). A lista de métricas disponíveis é tão grande quanto grupos de pesquisa na área, bem como seus prós e contras. Portanto, o bom senso deve ditar qual é a métrica mais relevante para o tipo de tarefa a qual o trabalho deseja avaliar. Não teria significado prático para este trabalho, por exemplo, utilizar a “taxa de falsos negativos”. Como estimar o número de falsos negativos, se todas as imagens no conjunto de testes são positivas (ou seja, possuem pelo menos uma baleia)?

A métrica escolhida para avaliar o desempenho dos algoritmos neste trabalho foi a “sobreposição” (overlap), como definida pelo *Pascal Visual Object Classification Challenge*, que é uma competição mundial de algoritmos de visão computacional tida como referência (UIJILINGS et al., 2013). A fórmula da sobreposição é dada por

$$O = \frac{\text{área}_{real} \cap \text{área}_{detectada}}{\text{área}_{real} \cup \text{área}_{detectada}}$$

A sobreposição pode variar entre 0 e 1, sendo um valor maior que 0.5 considerado “detectado”. A fórmula acima garante que os algoritmos sejam recompensados quanto mais o retângulo estimado da baleia se aproximar do retângulo verdadeiro. Em contrapartida, a pontuação é prejudicada se a baleia ainda estiver dentro do retângulo estimado, mas o retângulo estimado for excessivamente maior do que o retângulo real. Da mesma forma, retângulos estimados muito menores do que o real também serão penalizados.

Apesar da métrica de superposição proposta no *Pascal VOC* parecer sensata no escopo deste trabalho, a comparação com algoritmos de outros trabalhos ainda é ambígua, mesmo que utilizem a mesma métrica. Deve-se levar em conta a quantidade e natureza das imagens do conjunto de testes, assim como quão “permissível” foi a pessoa que rotulou em cada foto a posição real do objeto a ser detectado. No caso das baleias deste trabalho, as localizações reais foram rotuladas levando-se em consideração toda a parte humanamente visível dos cetáceos, mesmo a porção submersa quase indistinta a olho nu. Outros autores, no entanto, poderiam preferir a delimitação das baleias nas fotos somente na parte fora d’água e de fácil observação. Outros ainda poderiam preferir delimitar a região da baleia deixando uma certa margem de folga no interior do retângulo. Portanto, a estatística da tabela 2 tem como propósito a comparação intra-estudo somente.

Algoritmo	Baleias detectadas (%) ou precisão
Detecção de Contornos (Sobel)	59.6
Limiarização (Modelo de Mistura Gaussiana)	69.6
Superpixels (SLIC)	31.3
Similaridade de Histogramas	77.7

Tabela 2: Porcentagem de fotos em que as baleias foram corretamente detectadas (sobreposição > 0.5).

O desempenho da Similaridade de Histogramas foi notável, levando-se em consideração a simplicidade do conceito. O desempenho da segmentação por superpixels não foi boa como se esperava, já que muitos outros trabalhos a utilizam com sucesso. Acredita-se ser uma questão de ajustes no algoritmo.

Os métodos por limiarização e detecção de contorno também mostraram excelentes resultados, dada a variedade de condições luminosas em cada imagem. A escolha do espaço de cor mostrou ter uma relação direta com o desempenho final do algoritmo, de acordo com a figura 18.

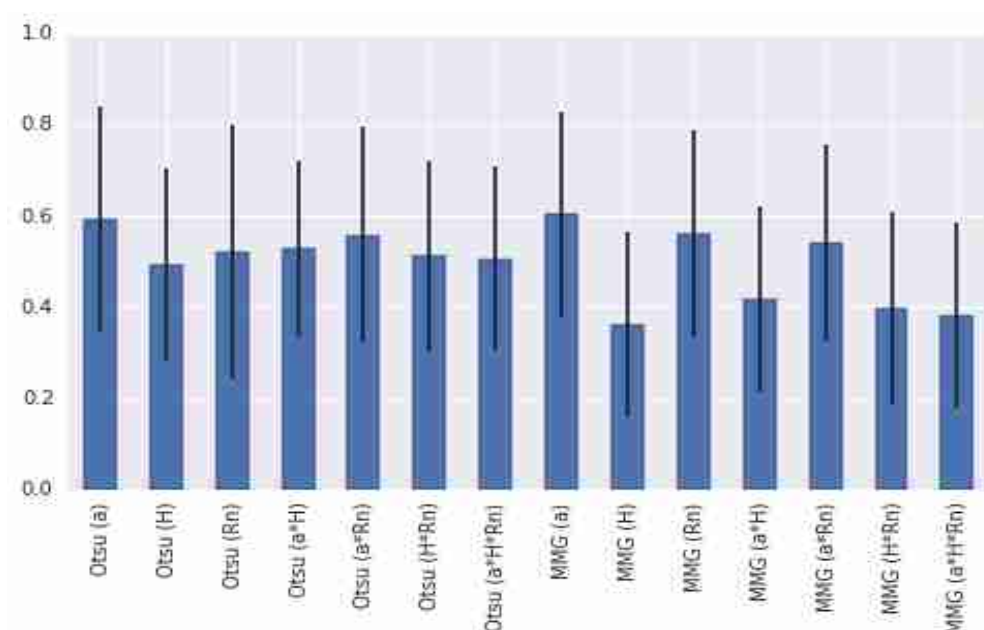


Figura 18: Precisão das limiarizações feitas com o método de Otsu e modelo de mistura gaussiana nos canais a, H, Rn e combinações destes.

De acordo com os resultados de diferentes configurações de limiarização, percebe-se que o canal utilizado como referência é mais importante até mesmo que o método de limiarização do histograma (Otsu ou MMG).

No Anexo 7.1 se encontram alguns exemplos de identificação de cada algoritmo. Os retângulos azuis indicam a área onde a baleia está presente, os retângulos vermelhos indicam as duas piores classificações pelos algoritmos, e os retângulos verdes indicam as duas melhores classificações. Nota-se que nuvens de vapor foi o elemento que mais prejudicou todos os desempenhos, seguido de condições luminosas onde as baleias não se destacam do fundo.

4.2 Comparação com outros trabalhos

O trabalho mais destacado em relação à contagem remota de baleias-francas foi feito por Fretwell et al. (2014), a partir das bandas “costal” e “infravermelho próximo” do satélite WorldView2, cujo sensor possui resolução máxima de 0,5 m por pixel. Os autores experimentaram quatro métodos diferentes de segmentação em suas imagens obtidas a partir do satélite de alta resolução. Foram estimados o número de acertos, número de falhas e número de falsos positivos e número de falsos negativos. Os testes foram realizados em várias imagens concatenadas de satélite, no qual os autores identificaram manualmente 91 baleias-francas. Não foi especificado como os acertos e falhas de identificação foram contabilizados nas estatísticas. As estatísticas do trabalho estão sumarizadas sob a métrica de precisão, dada por

$$Precisão = \frac{VP}{VP + FP}$$

Onde VP é a quantidade de verdadeiros positivos e FP é a quantidade de falsos positivos. Como o presente trabalho não possui uma relação de verdadeiros negativos, devido

às fotos de teste conterem sempre pelo menos um indivíduo, as estatísticas não puderam ser sumarizadas pela métrica F1, como é de praxe em trabalhos com classificadores.

Tabela 3: Valores de precisão obtidos pelos algoritmos deste estudo e o estudo de Fretwell et al. (2014). As linhas em negrito representam os melhores desempenhos de ambos.

Algoritmo	Precisão (%)
Detecção de Contornos (Sobel)	59,6
Limiarização (Modelo de Mistura Gaussiana)	69,9
Superpixels (SLIC)	31,3
Similaridade de Histogramas	77,7
*Iso-médias	38,6
*K-médias	51,9
*Limiarização (pancromática)	89,0
*Limiarização (banda 5)	76,2

* Fretwell et al., 2014

Os experimentos de Fretwell et al. (2014) mostram que o melhor desempenho obtido naquele estudo foi através da limiarização das imagens pancromáticas Worldview2. Entretanto, os autores afirmam que este método com desempenho excepcional foi realizado “com auxílio do usuário” -- isto é, ajustando iterativamente um valor de limiar que maximizasse sua precisão, ao contrário deste trabalho, cujos algoritmos foram executados de forma totalmente não-supervisionada.

Os autores também mencionam que a região escolhida para monitoramento proporcionou imagens com águas calmas e livres de nuvens. Estas duas características facilitam o processo de busca por elementos marítimos (ABILEAH, 2002). É comentado ainda que em trabalhos anteriores, os reflexos e sombras produzidos pela incidência de luz solar sobre as ondulações transformam a segmentação de baleias em uma tarefa impraticável.

Como o presente trabalho continha tais elementos em várias imagens e, levando em consideração que os algoritmos atuaram de forma não-supervisionada, o desempenho obtido pode ser considerado razoável.

5 CONCLUSÕES

A área da visão computacional pode trazer bons frutos para os mais diversos campos do conhecimento. Este trabalho exemplifica de maneira sucinta esta afirmação. O trabalho exaustivo de localização e rotulação de baleias em imagens aéreas pode ser automatizada com precisão de até 78%, utilizando-se apenas métodos estatísticos básicos para a segmentação, mas também um minucioso trabalho de pré-processamento das imagens. Devemos ter em mente, ainda, que parte dos resultados insatisfatórios se deve à enorme variância de propriedades das fotos utilizadas, as quais mostraram diferentes formas e intensidades de iluminação, textura da água causada pelos ventos, distâncias entre o sensor fotográfico e as baleias, focos, equipamentos fotográficos, ângulos e posições das baleias, porções das baleias que permaneceram submersas etc.

A precisão obtida neste estudo talvez possa ser melhorada através de métodos mais sofisticados de segmentação, aliado a um retificador capaz de compreender o formato do segmento e estimar a probabilidade dele ser uma baleia ou apenas impureza na imagem. O método criado por Uijilings et al. (2013) parece bastante promissor para o problema, devido a sua alta eficiência computacional na segmentação, deixando margem para que classificadores robustos como o SVM ou redes neurais de convolução possam classificar corretamente os segmentos.

6 REFERÊNCIAS

- ABILEAH, R. Marine mammal census using space satellite imagery. **U. S. Navy Journal of Underwater Acoustics**, v. 52, n. 3, jul. 2002.
- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. **Slic superpixels**. [s.l.: s.n.].
- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 34, n. 11, p. 2274–2282, nov. 2012.
- BRILLANT, S. W.; VANDERLAAN, A. S.; RANGELEY, R. W.; TAGGART, C. T. Quantitative estimates of the movement and distribution of North Atlantic right whales along the northeast coast of North America. **Endangered Species Research**, v. 27, p. 141–154, 2015.
- CHEN, S. Y.; DAI, S. G.; MU, P. A. Correction of CCD Camera Lens Vignetting. **Applied Mechanics and Materials**, v. 373–375, p. 603–607, ago. 2013.
- EBERHARDT, L. L.; CHAPMAN, D. G.; GILBERT, J. R. A review of marine mammal census methods. **Wildlife Monographs**, v. 63, p. 3–46, 1979.
- FINN, C.; DUYCK, J.; HUTCHEON, A.; VERA, P.; SALAS, J.; RAVELA, S. Relevance feedback in biometric retrieval of animal photographs. In: **Pattern Recognition**. [s.l.] Springer, 2014. p. 281–290.
- FRETWELL, P. T.; STANILAND, I. J.; FORCADA, J. Whales from Space: Counting Southern Right Whales by Satellite. **PLoS ONE**, v. 9, n. 2, p. e88655, 12 fev. 2014.
- FUJIWARA, M.; CASWELL, H. Demography of the endangered North Atlantic right whale. **Nature**, v. 414, n. 6863, p. 537–541, 2001.
- GILLESPIE, D. Detection and classification of right whale calls using an “edge” detector operating on a smoothed spectrogram. **Canadian Acoustics**, v. 32, n. 2, p. 39–47, 1 jun. 2004.
- KAGGLE. **Description - Right Whale Recognition | Kaggle**. Disponível em: <<https://www.kaggle.com/c/noaa-right-whale-recognition>>. Acesso em: 17 nov. 2015.
- KRAUS, S. D.; BROWN, M. W.; CASWELL, H.; CLARK, C. W.; FUJIWARA, M.; HAMILTON, P. K.; KENNEY, R. D.; KNOWLTON, A. R.; LANDRY, S.; MAYO, C. A. North Atlantic right whales in crisis. **SCIENCE-NEW YORK THEN WASHINGTON-**, v. 5734, p. 561, 2005.
- KRAUS, S. D.; MOORE, K. E.; PRICE, C. A.; CRONE, M. J.; WATKINS, W. A.; WINN, H. E.; PRESCOTT, J. H. The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). **Rep. int. Whal. Commn**, p. 145–51, 1986.

LAIST, D. W.; KNOWLTON, A. R.; PENDLETON, D. Effectiveness of mandatory vessel speed limits for protecting North Atlantic right whales. **Endangered Species Research**, v. 23, p. 133–147, 2014.

LIU, D.; YU, J. Otsu Method and K-means. In: Ninth International Conference on Hybrid Intelligent Systems, 2009. HIS '09, **Anais...** In: NINTH INTERNATIONAL CONFERENCE ON HYBRID INTELLIGENT SYSTEMS, 2009. HIS '09. ago. 2009.

MACDONALD, D. (ed.). **The encyclopedia of mammals**. 2 edition ed. Oxford ; New York: OUP Oxford, 2009.

MATE, B. R.; NIEUKIRK, S. L.; KRAUS, S. D. Satellite-Monitored Movements of the Northern Right Whale. **The Journal of Wildlife Management**, v. 61, n. 4, p. 1393–1405, 1997.

NEW ENGLAND AQUARIUM. **Right Whale Callosity Pattern Identification**. Disponível em:

<http://www.neaq.org/conservation_and_research/projects/endangered_species_habitats/right_whale_research/right_whale_projects/monitoring_individuals_and_family_trees/identifying_with_photographs/how_it_works/callosity_patterns.php>. Acesso em: 17 nov. 2015.

NOAA. **Automatic Whale Detector, Version 1.0 :: NOAA Fisheries**. Disponível em: <http://www.fisheries.noaa.gov/stories/2015/02/gray_whale_survey_thermal_imaging.html>. Acesso em: 17 nov. 2015.

OLIPHANT, T. E. Python for scientific computing. **Computing in Science & Engineering**, v. 9, n. 3, p. 10–20, 2007.

OTSU, N. A Threshold Selection Method from Gray-Level Histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62–66, 1979.

ÖZDEMİR, B.; AKSOY, S.; ECKERT, S.; PESARESI, M.; EHRLICH, D. Performance measures for object detection evaluation. **Pattern Recognition Letters**, v. 31, n. 10, p. 1128–1137, 2010.

REN, X.; MALIK, J. Learning a classification model for segmentation. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, **Anais...**IEEE, 2003.

UIJLINGS, J. R.; VAN DE SANDE, K. E.; GEVERS, T.; SMEULDERS, A. W. Selective search for object recognition. **International journal of computer vision**, v. 104, n. 2, p. 154–171, 2013.

VAN DER WALT, S.; SCHÖNBERGER, J. L.; NUNEZ-IGLESIAS, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T. scikit-image: image processing in Python. **PeerJ**, v. 2, p. e453, 2014.

YACK, T. M.; BARLOW, J.; ROCH, M. A.; KLINCK, H.; MARTIN, S.; MELLINGER, D. K.; GILLESPIE, D. Comparison of beaked whale detection algorithms. **Applied Acoustics**, Proceedings of the 4th International Workshop on Detection, Classification and Localization of Marine Mammals Using Passive Acoustics and 1st International Workshop on Density Estimation of Marine Mammals Using Passive Acoustics. v. 71, n. 11, p. 1043–1049, nov. 2010.

ZERBINI, A. N.; BAUMGARTNER, M. F.; KENNEDY, A. S.; RONE, B. K.; WADE, P. R.; CLAPHAM, P. J. Space use patterns of the endangered North Pacific right whale *Eubalaena japonica* in the Bering Sea. 2015.

7 ANEXO

7.1 Exemplos de detecção de cada algoritmo

Os retângulos em verde representam os dois melhores resultados, e os retângulos vermelhos representam os dois piores resultados da rodada de testes.



