

ALGORITMOS PARA SEGMENTAÇÃO DE IMAGENS INFRAVERMELHAS

Sergio Rodrigues Neves

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Gelson Vieira Mendonça, Ph.D.

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Prof. Jacques Facon, Dr. Ing.

Prof. Carla Liberal Pagliari, Ph.D.

Prof. Marco Antônio Grivet Mattoso Maia, Ph.D.

Prof. Murilo Bresciani de Carvalho, D. Sc.

RIO DE JANEIRO, RJ - BRASIL
JUNHO DE 2003

NEVES, SERGIO RODRIGUES

Algoritmos para Segmentação de
Imagens Infravermelhas [Rio de Janeiro]
2003

X, 121 p, 29,7cm (COPPE/UFRJ, D.Sc.,
Engenharia Elétrica, 2003).

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. Segmentação de imagens
2. Morfologia matemática
3. Transformada wavelet
4. Imagem infravermelha

I. COPPE/UFRJ II. Título (série)

Para Isabela.

AGRADECIMENTOS

Levei muito tempo para escrever os agradecimentos dessa tese. Trabalho difícil este de descrever sentimentos em palavras. Não tem uma equação, tabela ou gráfico para ajudar...

Assim começo a encarar essa tarefa com o agradecimento ao Gelson e o Eduardo, que, mais do que orientadores, foram amigos que ofereceram sua grande inteligência e capacidade à serviço dessa tese e de meu auxílio.

Minha família, mãe (Luiza), pai (Delton), esposa (Carmen), filha (Isabela), irmão (Jorge) e sobrinhas (Carla e Aline), e meus amigos (todos, todos) merecem o meu agradecimento por sua compreensão, paciência e carinho mesmo nos momentos de “stress” maior.

Por fim fica o agradecimento a quem utilizar essa tese de qualquer forma, pois é isso que faz todo esse trabalho valer a pena.

Resumo da Tese apresentada a COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.).

ALGORITMOS PARA SEGMENTAÇÃO DE IMAGENS INFRAVERMELHAS

Sergio Rodrigues Neves

Junho/2003

Orientadores: Gelson Vieira Mendonça
Eduardo Antônio Barros da Silva

Programa: Engenharia Elétrica

Num sistema de classificação de alvos, é essencial que se possa extrair características precisas destes, com a intenção de potencializar os acertos do sistema. A segmentação de imagens sem supervisão é um dos mais desafiadores problemas para a visão artificial. Uma imagem infravermelha é um mapeamento em duas dimensões da radiação térmica de uma cena e, apresenta como características importantes para extração, as bordas dos elementos. A extração de objetos de imagens em infravermelho é de grande interesse tanto para aplicações no setor militar como no comercial. Este trabalho descreve um método automático para a segmentação de imagens infravermelhas utilizando a transformação watershed a partir de um marcador que combina a transformada wavelet módulo máximo com técnicas morfológicas, e que apresenta diversas vantagens sobre os métodos convencionais tais como a extração de objetos abaixo ou acima da temperatura ambiente, robustez frente ao ruído inerente a imagens infravermelhas e solução para o problema da linha do horizonte com resultados precisos e com baixo número de detecções falsas.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.).

INFRARED IMAGE SEGMENTATION ALGORITHMS

Sergio Rodrigues Neves
June/2003

Advisors: Gelson Vieira Mendonça
Eduardo Antônio Barros da Silva

Department: Electrical Engineering

On a target classification system it is essential that precise characteristics can be extracted, with the intention of improving the matches. Automatic image segmentation is one of the most challenging problems for artificial vision. An infrared image is a thermal radiation two-dimensional mapping. Its edges are an important characteristic for segmentation purposes. The infrared object image extraction is of great interest for military purpose and also for commercial applications. This work describes an infrared image segmentation method that employs a watershed transformation using a marker that combines a wavelet transform modulus maxima with morphological techniques. This approach has many advantages compared with the traditional methods, as its ability of extracting objects beyond or below the environment temperature, robustness to noise inherent to infrared images, as well as a solution of the horizon problem with precise results and with low false detection rate.

ÍNDICE

AGRADECIMENTOS	iv
ÍNDICE.....	vii
ÍNDICE DE FIGURAS	ix
1 – INTRODUÇÃO.....	1
2-IMAGENS INFRAVERMELHAS	7
3 - MORFOLOGIA MATEMÁTICA	12
3.1 - Introdução	12
3.2 – Operações Básicas.....	13
3.2.1 – Elemento Estruturante.....	14
3.2.2 - Erosão	15
3.2.3 - Dilatação	17
3.2.4 – Propriedades da erosão e da dilatação	19
3.2.5 – Operadores de Minkowski	21
3.3 – Gradiente Morfológico	22
3.4 - Abertura	25
3.5 - Fechamento	27
3.6 – Propriedades da abertura e fechamento.....	28
3.7 – Transformação “hit or miss”	29
3.8 - Afinamento	30
3.9 – Espessamento	31
3.10 - Esqueleto.....	32
3.11 – Esqueleto por zona de influência	35
3.12 – Transformações geodésicas.....	36
3.12.1 – Dilatação geodésica	36
3.12.2 – Erosão geodésica.....	37
3.12.3 – Reconstrução Morfológica.....	38
3.13 – Filtragem Morfológica	40
3.13.1– Abertura por área	40
3.14 – Segmentação Morfológica e Transformação Watershed	41
3.14.1 – Transformação Watershed	41
3.15 – Segmentação controlada por marcador	49
3.15.1 – Escolha do marcador.....	50
4 –SEGMENTAÇÃO UTILIZANDO A TRANSFORMAÇÃO WATERSHED	52
4.1 – Introdução	52
4.2 – Uso da transformação watershed em imagens infravermelhas	53

4.3 – Considerações práticas com o paradigma de Beucher-Meyer como fio condutor...	54
4.4 – A escolha do método para extração do marcador	63
5 - TRANSFORMADA WAVELET MÓDULO MÁXIMO.....	64
5.1 - Introdução	64
5.2 - Definição da Transformada Wavelet	65
5.2.1 - Análise por multiresolução	67
5.2.2 - Filtros em quadratura	69
5.2 – Transformada Wavelet Binária	72
5.3 – Transformada Wavelet Módulo Máximo	74
5.4 – Vantagens e aspectos práticos da transformada wavelet módulo máximo para a técnica desenvolvida	76
5.4.1 – Introdução	76
5.4.2 – Extraíndo os mapas de bordas.....	77
5.4.3 – Escolhendo os mapas de borda	82
6 – DESCRIÇÃO DO MÉTODO PROPOSTO	87
6.1 – Introdução	87
6.2 – O primeiro passo: detectando os objetos semânticos e construindo o conjunto de marcadores	88
6.2.1- Detectando as bordas dos objetos semânticos	88
6.2.2 – Escolha das linhas de borda.....	89
6.2.3 – O problema do horizonte	92
6.2.4 – Finalizando o marcador.....	96
6.3 – Segundo passo: a simplificação da imagem.....	99
6.4 – Último passo: a transformação watershed.....	100
7 - RESULTADOS.....	102
7.1 - Introdução	102
7.2 – Generalidade do resultado.....	103
7.3 – Solução para o problema da linha do horizonte	104
7.4 – Pequenas estruturas e reflexos	107
7.5 – Imagens com vídeo sintético	109
7.6 - Conclusão.....	111
8 – CONCLUSÕES E SUGESTÕES.....	112
APÊNDICE A	114
Algoritmo para realização da transformada wavelet binária e sua inversa	114
APÊNDICE B.....	117
Funções utilizadas para a implementação dos operadores morfológicos.....	117
BIBLIOGRAFIA	119

ÍNDICE DE FIGURAS

Figura 1: Diagrama em blocos de um sistema de classificação de alvos.....	5
Figura 2: Transmitância atmosférica (UV – ultravioleta, V- visível).....	8
Figura 3: Exemplos de imagens em infravermelho.....	9
Figura 4: Imagem infravermelho e histograma multimodal.....	11
Figura 5: Operações de máximo ou supremo e mínimo ou ínfimo.....	13
Figura 6: Elementos estruturantes planos isotrópicos para grade retangular e hexagonal.....	15
Figura 7: Erosão do objeto A pelo elemento estruturante B.....	16
Figura 8: Dilatação do objeto A pelo elemento estruturante B.....	18
Figura 9: Subtração de Minkowski.....	21
Figura 10: Adição de Minkowski.....	22
Figura 11: Gradiente básico ou de Beucher.....	23
Figure 12: Gradiente interno e gradiente externo.....	24
Figure 13 : Gradiente espesso.....	24
Figure 14: Gradiente direcional.....	25
Figura 15: Abertura em termos de dilatação e erosão.....	26
Figura 16: Fechamento em termos de dilatação e erosão.....	28
Figura 17: Afinamento.....	31
Figura 18: Espessamento.....	32
Figura 19: Esqueleto através da noção de disco máximo.....	33
Figura 20 : Esqueleto por abertura.....	34
Figura 21 : Esqueleto por afinamento homotópico.....	35
Figura 22: Esqueleto por zona de influência.....	36
Figura 23: Dilatação geodésica.....	37
Figura 24 : Erosão geodésica.....	38
Figura 25 : Reconstrução morfológica por dilatação.....	39
Figura 26: Reconstrução morfológica por erosão.....	40
Figura 27: Mínimos, bacias de segmentação e Watersheds.....	42
Figura 28: Relações de inundação de objetos em uma imagem.....	44
Figura 29: Imagem original e sobre-segmentação ocasionada pela aplicação da transformada watershed diretamente sobre a imagem original.....	49
Figura 30: Imposição de mínimo.....	51
Figura 31: Imagem em infravermelho com destaque para , pelo menos, três regiões uniformes de cinza internas ao objeto semântico.....	53
Figura 32: Diagrama em blocos da primeira tentativa para a segmentação de imagens infravermelhas.....	55
Figura 33: Segmentação morfológica com o gradiente como marcador.....	55
Figura 34: Contorno da imagem determinado manualmente.....	56
Figura 35: Teste dos elementos estruturantes que dilatam o conjunto de bordas escolhidas.....	57
Figura 36: Teste dos elementos estruturantes que dilatam o conjunto de bordas escolhidas.....	58
Figura 37: Elementos estruturante escolhido para dilatar o conjunto de bordas escolhidas.....	58
Figura 38: Teste da limiarização em oito níveis como simplificação para a imagem que será submetida à transformação watershed.....	59
Figura 39: Teste da filtragem passa-baixas como simplificação para a imagem que será submetida à transformação watershed.....	60
Figura 40: Teste do gradiente morfológico como simplificação para a imagem que será submetida à transformação watershed.....	61
Figura 41: Teste do gradiente morfológico como simplificação para a imagem que será submetida à transformação watershed.....	62
Figura 42: Algoritmo para decomposição da imagem em componente da transformada wavelet.....	71
Figura 43: Representação wavelet tradicional para uma e duas camadas ou níveis.....	71
Figura 44 : Diagrama em blocos esquemático da transformada wavelet binária.....	73
Figura 45 : Módulos máximos ao longo da escala.....	77

Figura 46: Mapas de módulos máximos com busca bidimensional e tradicional para a imagem original da figura 44.	78
Figura 47 : Mapas de módulos máximos realizados com a comparação com limiar e sem a comparação para a transformada wavelet módulo máximo realizada para os níveis 1, 2 e 3 respectivamente a partir da imagem original da figura 45.	79
Figura 48 : Imagens de módulo para os níveis 1, 2, 3, 4, 5 e 6 a partir da imagem original da figura 44. .	80
Figura 49 : Teste para determinar o limiar para validar os módulos máximos.	81
Figura 50 : Teste para determinar o limiar para validar os módulos máximos. Limiar baseado na média para os níveis 1,2 e 3.	82
Figura 51 : Mapas de borda para os níveis 1,2 e 3 da figura original.	83
Figura 52 : Mapas de borda para os níveis 1,2 e 3 da figura original.	84
Figura 53 : Mapas de borda para os níveis 1,2 e 3 da figura original.	84
Figura 54 : Mapas de borda para os níveis 1,2 e 3 da figura original.	85
Figura 55 : Mapas de borda combinando os níveis 2 e 3 das imagens 52 e 53.	86
Figura 56 : Diagrama em blocos do método proposto.	87
Figura 57 : Diagrama em blocos da detecção das bordas do objeto semântico.	89
Figura 58 : Exemplos de teste para determinar o limiar da abertura por área que simplifica o mapa de bordas obtido pela combinação dos níveis 2 e 3 da TWMM.	92
Figura 59 : Diagrama em blocos do processo de retirada da linha do horizonte.	93
Figura 60 : Retirada de objetos que tocam simultaneamente as duas margens da imagem.	94
Figura 61 : Problema da linha do horizonte.	95
Figura 62 : Gradiente horizontal como parte do processo para a solução do problema da linha do horizonte.	96
Figura 63 : Mapa de bordas antes e depois do tratamento do problema do horizonte.	96
Figura 64 : Diagrama em blocos do processo de finalização do marcador para a transformação watershed.	97
Figura 65 : Esqueletonização do marcador.	98
Figura 66 : Marcadores propostos.	99
Figura 67 : Imagem original e imagem simplificada.	100
Figura 68 : Watersheds antes e depois do filtro final de abertura por área.	101
Figura 69 : Resultados da segmentação.	104
Figura 70 : Resultados da segmentação de imagens com a linha do horizonte.	105
Figura 71 : Resultados da segmentação de imagens.	106
Figura 72 : Resultados da segmentação de imagens com estruturas pequenas ou finas.	108
Figura 73 : Resultados da segmentação de imagens.	109
Figura 74 : Resultados da segmentação de imagens com vídeo sintético.	110

1 – INTRODUÇÃO

Com apenas alguns segundos de entrada no mundo, aqueles que dentre nós foram abençoados com o presente da visão, começam a perceber imagens do que está ao seu redor. Pouco tempo depois o processamento, a análise e o entendimento de imagens por seres humanos se torna quase uma rotina.

A partir do surgimento dos computadores digitais, o processamento e aquisição de imagens a partir destes se tornou uma tarefa possível, porém complexa e mobilizadora de vários grupos de trabalho nas áreas de processamento de sinais e imagens, que emergiram em várias partes do mundo para a solução de tais desafios. Os constantes avanços nas tecnologias de processamento de sinais e de computadores trazem consigo novos desafios e a necessidade de investimento nessas áreas, para a satisfação de suas expectativas. Atualmente a velocidade de processamento de um microprocessador utilizado em um microcomputador dobra aproximadamente a cada dezesseis meses. Isso aumenta significativamente a capacidade de processamento aumentando assim a flexibilidade dos “softwares”, sua complexidade e a simplicidade de interface junto ao usuário. O avanço das tecnologias digitais também expandiu a capacidade e a velocidade das redes de comunicação. Consequentemente, muitas novas aplicações e serviços relacionados ao processamento de imagens, tais como TV digital, videoconferência, multimídia interativa, processamento de imagens médicas, bibliotecas de arquivos de vídeo, entre outras, vem assumindo cada vez mais importância, e tem papel de destaque na tentativa de criação de uma “estrada internacional de informações”, ou seja, um fluxo contínuo internacional de informações, dados, áudio e vídeo. É claro que tais temas não estão restritos a questões técnicas apenas, mas estão relacionados também a questões econômicas, políticas e sociais, e estão intimamente ligadas à transformação das sociedades. Entre as várias áreas de pesquisa que se relacionam com esta infra-estrutura de informação global, um tópico é de interesse para vários tipos de indústria: a extração de informação de imagens.

“I only have eyes for you”¹. Como no “standard” de jazz americano, a todo momento escolhemos um objetivo em uma cena ou situação para seguir ou prestar atenção. Pode ser o foco de nosso desejo, como a pessoa amada da música, ou qualquer outra coisa que chame a nossa atenção, como a linha que estamos lendo num livro. A todo o momento realizamos segmentações com o interesse de extrair objetos semânticos de acordo com o que a situação impõe. Entende-se por objeto semântico aquele objeto que traduz uma especial relevância em uma imagem ou seqüência de imagens, de acordo com os objetivos relacionados. A operação de extração de objetos semânticos, para o cérebro, é tão automática e veloz que poucas vezes é possível se dar conta de sua complexidade. Essa complexidade é bem percebida quando tenta-se reproduzir o processo de segmentação via computador. Vários problemas se relacionam com a extração de um objeto:

- Deve-se ter um conhecimento a priori do que se vai extrair. Ao se pensar no “mecanismo de segmentação” humano, pode ser observado que ele funciona de forma específica: de acordo com cada motivação, o cérebro coordena a segmentação de um determinado elemento na situação e cena do dia a dia. Por exemplo, quando assistimos a um jogo de futebol, nossa atenção está, principalmente, voltada para o jogador que está com a bola; quando lemos quadrinhos, “separamos” da página o quadrinho na seqüência que se está lendo. Em outras palavras, o cérebro precisa de algum conhecimento a priori para realizar a tarefa de segmentação.
- Deve-se ter as condições de contorno da imagem ou conjunto de imagens de onde se vão extrair elementos. É importante saber se temos uma imagem ou uma seqüência, qual é seu mapa de cores e a partir de que tipo de sensor ela foi originada.

Assim, para realizar um método artificial de segmentação de imagens, temos que partir de características intrínsecas ao objeto que se deseja extrair e sua imagem, ou seqüência de imagens, sejam elas cores, tonalidades, descontinuidades ou movimento, entre outras.

Esse trabalho apresenta um método automático para a segmentação de imagens infravermelhas. A segmentação de uma imagem infravermelha, tem o mesmo objetivo relacionado a qualquer outro tipo de detecção realizada por outro tipo de sensor que origine um mapeamento em duas dimensões: extrair elementos que serão os fundamentos para detectar ou mesmo reconhecer alvos, materiais ou padrões.

¹ Em “I only have eyes for you” – Warren –Dubin.

Uma imagem digital monocromática pode ser considerada como uma matriz $m \times n$ onde os índices de linha e coluna identificam um ponto na imagem e o valor numérico do elemento correspondente na matriz está associado a alguma grandeza física medida nesse ponto. Os elementos da matriz são chamados pixels representando a abreviação de “picture elements” [1]. Na tela do computador esta matriz é apresentada associando a um ou mais pontos da tela um tom de cinza proporcional ao valor numérico de cada elemento da matriz. Uma imagem infravermelha [2] é produzida através da radiação térmica de uma cena, compondo um mapa em duas dimensões da irradiação observada representando as variações de temperatura, emissividade e reflexividade da cena. A radiação térmica é produzida por todas as superfícies acima de 0°K, e pode ser produzida sem qualquer fonte de luz visível. Essa característica de mapeamento térmico da cena justifica uma grande variabilidade para um mesmo cenário, dependendo de variações climáticas e temporais tais como, hora do dia, umidade relativa e temperatura. Essas variações climáticas introduzem uma série de ruídos, que podem ser considerados como alvos falsos, e que se juntam a outros introduzidos artificialmente, especialmente por cameras infravermelhas de uso militar, como vídeo sintético. Compreende-se por vídeo sintético, uma série de informações a respeito da cena que são introduzidas na imagem por meio de caracteres.

O conhecimento a priori para a segmentação de imagens infravermelhas, é que o objeto de interesse está sempre a uma temperatura acima ou abaixo da temperatura ambiente. As regiões de interesse são, então, caracterizadas por transientes como picos e bordas, que se destacam do fundo, localizado nas regiões de temperatura ambiente.

O processo de extrair um objeto de uma imagem pode ser dividido em duas tarefas: reconhecer e delinear o objeto. Reconhecer o objeto consiste em determinar mais ou menos sua localização e distinguir este objeto das outras entidades similares na imagem. Delinear consiste em definir precisamente a extensão do objeto na imagem. Operadores humanos (especialistas no domínio da aplicação) normalmente são capazes de reconhecer o objeto com maior facilidade que os algoritmos de computador. A maior razão para esta limitação dos algoritmos é a dificuldade de transformar o conhecimento global relevante sobre o objeto em operações locais computáveis. Isso tem sido um dos maiores obstáculos das técnicas de segmentação de imagens. Por outro lado, os algoritmos de computador normalmente executam a tarefa de delinear um objeto na

imagem bem melhor (mais repetitivo, mais rápido e mais exato) do que os operadores humanos. Vários métodos de segmentação de imagem têm sido criados e estudados para os mais diversos objetivos. Para o caso da extração de objetos em imagens infravermelhas, um método popular é a técnica de limiarização (thresholding) [3]. Apesar desta técnica ser de fácil implementação, ela é bastante dependente da seleção do limiar (threshold) para a validação de um pixel como pertencente a um alvo. Por outro lado, técnicas que calculam o limiar baseado em um histograma [4] freqüentemente recaem em problemas quando o valor do pixel muda através do objeto ou do fundo. Outra técnica popular para a segmentação de imagens infravermelhas é a abordagem baseada em um modelo [5]. Sua desvantagem é a de requerer conhecimento a priori do formato do objeto.

A morfologia matemática [6] [7] [8] [9] é uma poderosa ferramenta para o processamento de imagens, especialmente pela sua habilidade em lidar diretamente com o objeto da imagem. Uma das ferramentas mais eficientes e utilizadas para a segmentação provém da morfologia matemática : é a transformação watershed (divisor das águas)[10]. Porém, a não ser por poucos casos em que o objeto alvo é bem diferenciado do fundo (ou vice e versa) e a textura é uniforme, a watershed não pode ser aplicada diretamente na imagem sob pena de sobre-segmentação. Para eliminar, ou, pelo menos, minimizar este problema, foi elaborado um método baseado no uso da transformação watershed, o paradigma de Beucher-Meyer [11], que consiste basicamente em duas etapas: a primeira que seleciona marcadores para os objetos de interesse (considerada como etapa “inteligente”) e um passo que realiza efetivamente a transformação watershed para a extração destes.

O método proposto nesse trabalho pode ser considerado como baseado na filosofia do modelo de Beucher-Meyer. Uma das contribuições desse trabalho é o uso de métodos multiescala na seleção dos marcadores. Especificamente, é proposta a utilização da transformada wavelet módulo máximo [12] [13], combinada com operadores morfológicos, o que funciona como uma pré-segmentação. Esse marcador oferece à transformação watershed uma região aproximada da localização do contorno do objeto de interesse, que, combinada com outras operações morfológicas, esta transformação vai extrair com precisão.

A extração de um alvo com a maior precisão possível é crucial para um sistema de detecção e classificação automática de alvos. Esse sistema poderia ser concebido baseado no diagrama em blocos a seguir.

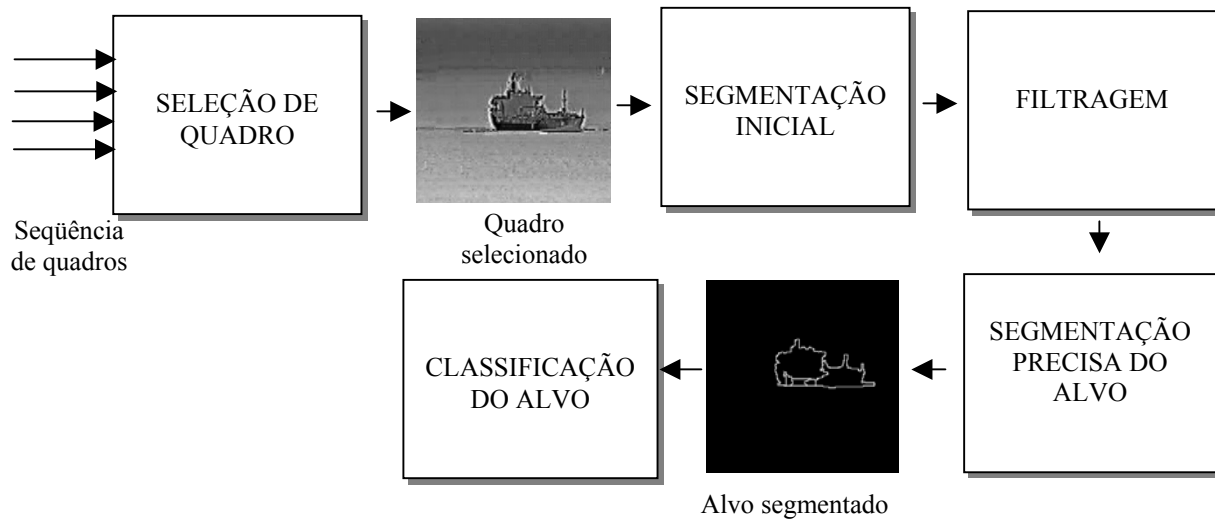


Figura 1: Diagrama em blocos de um sistema de classificação de alvos

A partir da seleção de um quadro numa seqüência (que pode ser executado, por exemplo, a partir da detecção do movimento do alvo, ou da detecção de certa temperatura no cenário, ou de maneira não automática), é realizada uma segmentação inicial. Uma fase de filtragem, então, é necessária com o objetivo de se evitarem ruídos que aumentem a detecção indesejada de alvos falsos. Note que essa filtragem pode ser do tipo não linear com o objetivo de não modificar os objetos semânticos e pode utilizar operadores morfológicos. Uma segmentação precisa do alvo o torna apto para servir de entrada a um método para a sua classificação.

A segmentação precisa para um objeto semântico, no caso um alvo, tem especial relevância visto que quanto mais precisa a retirada de seu contorno maior acerto ter-se-á na classificação do alvo. Inclusive, a partir dessa segmentação precisa, pode ser desenvolvido um modelo matemático para a parametrização do objeto extraído com vistas a utilização deste como dado de entrada para uma rede neural de classificação.

O trabalho aqui proposto concentra seus esforços na segmentação de uma imagem infravermelha já selecionada de determinada seqüência. Ou seja, na figura 1 seria representado pelos blocos de segmentação inicial, filtragem e segmentação final.

No capítulo 2 é realizada uma breve introdução ao imageamento infravermelho e aos métodos convencionais de segmentação de imagens em infravermelho. No capítulo 3 uma introdução às técnicas morfológicas utilizadas no método tem seu lugar com destaque para a segmentação morfológica e os operadores utilizados na técnica desenvolvida. No capítulo 4 são relacionados aspectos práticos desse trabalho com a segmentação morfológica. No capítulo 5 é apresentada a transformada wavelet módulo máximo que é utilizada como o principal componente do processo de extração do marcador, e onde alguns aspectos práticos também são descritos. No capítulo 6 é detalhado o método proposto. No capítulo 7 são apresentados e comentados os resultados, e, por fim, no capítulo 8 temos as conclusões e sugestões de pesquisa.

2-IMAGENS INFRAVERMELHAS

A descoberta da radiação infravermelha tem dois séculos e é creditada ao alemão Friedrich Wilhelm Herschel (1738-1822). Quando estudava as cores do espectro solar transmitidas através de filtros que permitiam a passagem de apenas uma cor, Herschel observou que cada um dos filtros deixava passar uma quantidade específica de calor. A partir de um experimento que media a temperatura das diferentes cores ele verificou que a temperatura diminuía do vermelho para o violeta. Ao medir essa temperatura da região do espectro antes da cor vermelha, Herschel notou que era superior àquela emitida pelo vermelho.

Herschel deduziu que este fenômeno estaria associado a uma forma de radiação invisível que transmitia calor e lhe deu o nome de raios calóricos. Para investigar a natureza desta suposta nova forma de radiação, ele realizou experiências envolvendo fenômenos da luz visível e verificou que os raios calóricos tinham natureza ondulatória como a luz. Assim, Herschel demonstrou a existência de uma nova forma de luz, invisível, e a chamou de infravermelho. Cinquenta anos mais tarde James Clerk Maxwell (1831-1879) constataria que essa radiação infravermelha e a luz visível eram apenas uma pequena parte do espectro da radiação eletromagnética.

Ao incidir sobre um prisma, o raio de luz solar forma o espectro, devido à propriedade do vidro de desviar menos, através do prisma, os componentes de menor frequência da luz solar (como o vermelho) do que os de maior frequência (como a cor violeta). Esse espectro corresponde ao espectro da luz visível que representa uma janela ínfima do espectro da radiação eletromagnética, composta também por ondas de rádio, microondas, infravermelho, ultravioleta, raios X e raios gama.

A faixa de luz visível possui comprimentos de onda variando de 4×10^{-7} m a 7×10^{-7} m. No infravermelho, os comprimentos de onda variam de valores próximos ao do vermelho até 1mm.

A região do infravermelho é, por conveniência, subdividida em quatro partes com denominações que relacionam sua proximidade com a região visível (0,4 a 0,7 μm) [2] (figura 2).

- Infravermelho próximo (NIR – “near infrared”) – região de 0,75 μm (400THz) a 3 μm (100THz).
- Infravermelho médio (MIR – “middle infrared”) – região de 3 μm (100THz) a 6 μm (50THz).
- Infravermelho distante (FIR – “far infrared”) – região de 6 μm (50THz) a 15 μm (20THz).
- Infravermelho extremo (XIR – “extreme infrared”) – região de 15 μm (20THz) a 1mm (300GHz).

As regiões de particular interesse para as tarefas de detecção e reconhecimento são as faixas do infravermelho médio e do infravermelho distante, por estas possuírem condições de propagação mais favoráveis.

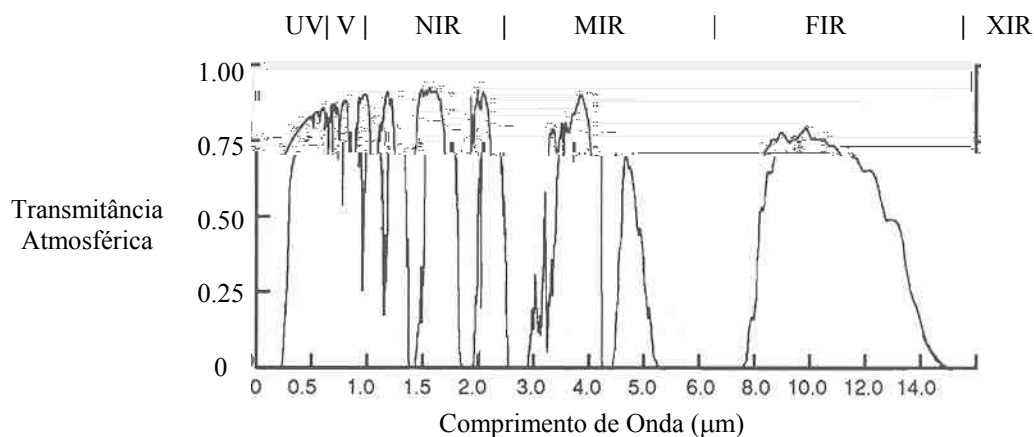


Figura 2: Transmitância atmosférica (UV – ultravioleta, V- visível)

Uma imagem infravermelha, como pode ser visto na figura 3, é um mapeamento bidimensional em tons de cinza que representa a captação da radiação infravermelha por um detector específico e representa a caracterização e quantificação da energia térmica emitida por um objeto.



Figura 3: Exemplos de imagens em infravermelho.

A utilização de imagens em infravermelho se aplica em diversas áreas de atuação tais como no monitoramento do meio ambiente (através de imagens obtidas via satélite é possível monitorar queimadas em matas, o aumento da poluição urbana ou a temperatura de superfície do mar), na astronomia, onde a região de radiação infravermelha é uma região chave para diversos fenômenos, no monitoramento de segurança na indústria e em sistemas de defesa, entre outras.

Nos sistemas de defesa, o infravermelho tem uma grande importância nas operações de detecção e reconhecimento. Por ser um instrumento de detecção passiva (não há emissão de onda eletromagnética, como no caso do radar, por exemplo), sua eficácia em patrulhamento de fronteiras e do litoral, onde é importante detectar sem ser detectado, é comprovada.

Entre o emissor da radiação infravermelha percebida e o detetor existe atenuação de radiação, fazendo com que o detetor capte apenas parte da energia emitida. Essa atenuação é governada por uma lei exponencial, ou seja, a transmissão através de um percurso da atmosfera pode ser expressa com:

$$\varepsilon = e^{-kx}$$

onde k é o coeficiente de extinção e x a distância considerada na propagação.

Essa atenuação é devida, de um modo geral, a dois mecanismos básicos:

1. Absorção molecular pelos gases da atmosfera (vapor d'água e gás carbônico).
2. Difusão pelas gotículas d'água e partículas existentes na atmosfera tais como poeira e neblina.

A atenuação pela atmosfera é altamente dependente do comprimento de onda da radiação, fazendo com que existam duas janelas atmosféricas clássicas, a primeira na região de 3 a 5 μm em que se situam objetos nas temperaturas de 300 a 600°C e a segunda na região de 8 a 14 μm cujas emissões de energia são produzidas por objetos com temperatura próxima ao ambiente.

De uma maneira geral, para a segmentação de um objeto semântico de uma imagem infravermelha, o contraste da radiação emitida por ele e o fundo é o principal parâmetro de entrada.

Uma das técnicas mais populares para a extração de regiões de interesse em uma imagem infravermelho é a que se utiliza da técnica de limiarização [3]. Se for considerado que μ_a é um pixel do alvo, que é proporcional à sua temperatura, e μ_f é um pixel do fundo que é proporcional à temperatura do ambiente, uma métrica simples de análise é a que executa a sua diferença:

$$m = \mu_a - \mu_f$$

Como a temperatura do alvo pode estar a uma temperatura maior ou menor que o ambiente, é mais adequado considerar:

$$m = |\mu_a - \mu_f|$$

O problema maior desse tipo de abordagem se encontra na determinação do valor de limiar m que valida o pixel como alvo. Na técnica baseada no histograma da imagem [5], o limiar é escolhido de forma a ser um valor de cinza no vale de um histograma bimodal, no qual uma moda representa o objeto e outra o fundo. Um problema freqüentemente associado a este tipo de técnica é que tanto o objeto como o fundo podem ser multimodais. Então, o vale do histograma é gradual e seu fundo não é bem definido como pode ser visto na figura 4.

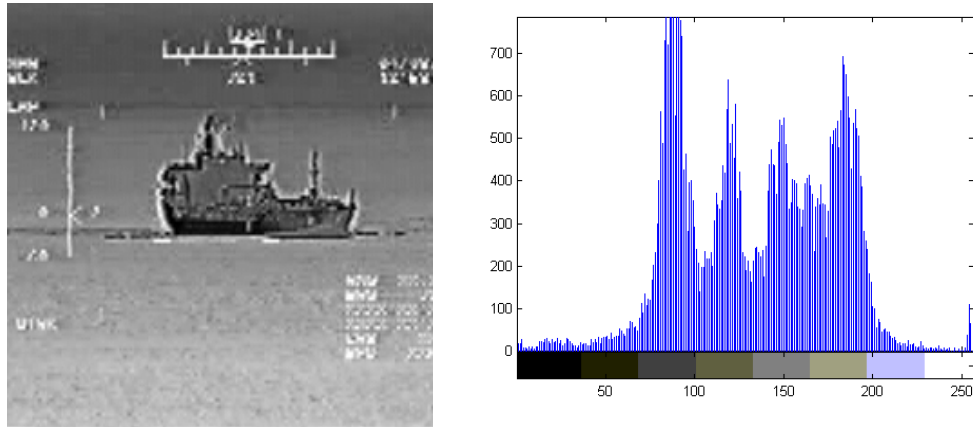


Figura 4: Imagem infravermelho e histograma multimodal.

Em [3] esse problema é contornado, mas a solução é especializada para segmentação de alvos mais quentes que o ambiente, especialmente aviões.

Uma técnica para segmentação de navios em imagens em infravermelho por detecção de bordas, que é uma alternativa um método de segmentação desenvolvido pela marinha australiana, é apresentada em [14]. Vários problemas relacionados ao estudo da marinha australiana são solucionados tais como, a necessidade de se conservar o navio em uma posição central e a eliminação de pequenas estruturas que podiam gerar alvos falsos. Propõe-se a realização de histogramas de pequenos blocos da imagem para determinar as posições de alvo através de um processo que calcula e valida o limiar, mesmo sendo este histograma multimodal. A partir da determinação da posição dos alvos é computado o crescimento de região até a obtenção do alvo. A parada do processo de crescimento de região é realizado por detecção de bordas Apesar de, positivamente, combinar as técnicas de crescimento de região e detecção de bordas, a técnica implica em pelo menos dois problemas: no caso de múltiplos alvos o processo deve ser reiniciado a cada captura de alvo, e, como o fim do crescimento de região é determinado pela detecção de bordas, a linha do horizonte pode determinar algum problema, no caso do navio estar sobre ela.

Em nosso trabalho é apresentada uma nova abordagem que combina wavelet binária, técnicas morfológicas e watersheds com resultados superiores às técnicas convencionais.

3 - MORFOLOGIA MATEMÁTICA

3.1 - Introdução

A morfologia matemática ou simplesmente morfologia, pode ser definida como uma teoria para a análise da forma e da estrutura [6] [7] [8] [9]. A morfologia é matemática no sentido em que é baseada em teoria dos conjuntos, geometria integral e álgebra de reticulados. Mas a morfologia matemática não é só uma teoria: é também uma poderosa ferramenta para a análise de imagens.

A origem da morfologia matemática se remonta a meados dos anos 60 e ao estudo de materiais porosos de G. Matheron. Em seu trabalho pioneiro (*Eléments pour une théorie des milieux poreux* – 1967) Matheron propôs as primeiras transformações morfológicas. Quase que simultaneamente, em 1972, o desenvolvimento e a comercialização de um “hardware” especialista por J. C. Klein e J. Serra permitiu o uso dessas transformações para a solução de problemas práticos de análise de imagem. A partir dessas primeiras definições e formalizações de operações morfológicas para a investigação da geometria de objetos em uma imagem binária foi formalizada toda uma teoria por G. Matheron e J. Serra [9].

De uma maneira geral, podem ser considerados dois tipos básicos de imagens às quais se aplicam as operações morfológicas: imagens binárias e em tons de cinza.

Na morfologia aplicada à imagens binárias, nas quais os únicos valores de pixel possíveis são 0 ou 1 (branco e preto), na vizinhança de cada pixel da imagem original é testada uma configuração de pontos pretos e brancos (o elemento estruturante), de acordo com uma proposição específica (por exemplo : o elemento cabe no objeto ou toca o objeto).

Na morfologia aplicada à imagens em tons de cinza, além do procedimento descrito para as operações binárias é importante o conhecimento do valor dos pixels envolvidos tanto na imagem original, quanto no elemento estruturante, para a execução das operações de mínimo ou ínfimo (\wedge) e máximo ou supremo (\vee). Essas operações

podem ser definidas para cada ponto x de duas imagens f e g de domínios idênticos como:

$$(f \vee g)(x) = \max[f(x), g(x)]$$

e

$$(f \wedge g)(x) = \min[f(x), g(x)]$$

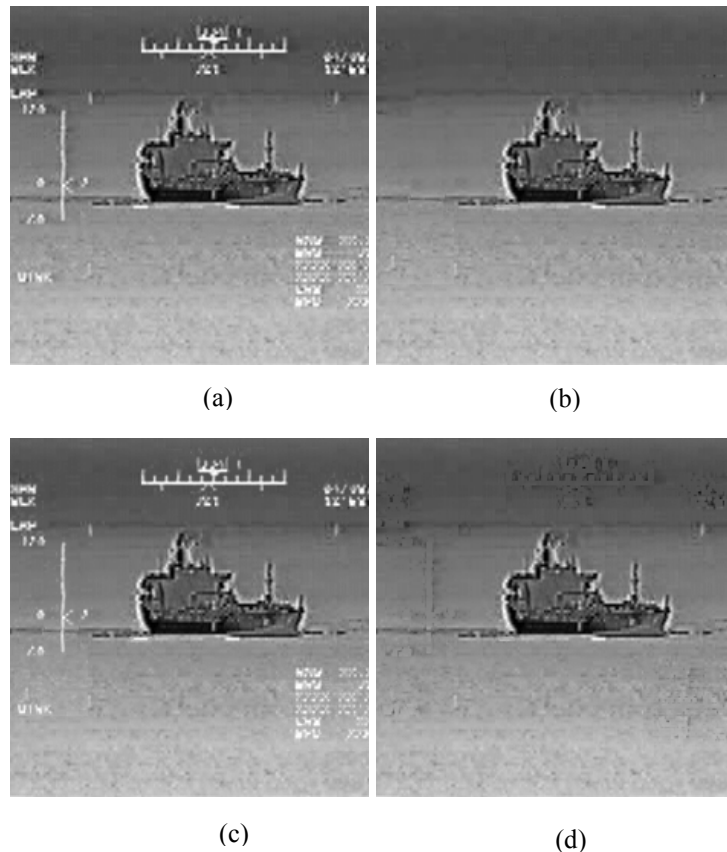


Figura 5: Operações de máximo ou supremo e mínimo ou ínfimo. Da esquerda para a direita e de cima para baixo: (a) imagem de entrada f , (b) imagem de entrada g , (c) operação de máximo ou supremo e (d) mínimo ou ínfimo. Note que na operação de máximo, o vídeo sintético que se encontra perto do branco (valor mais alto na escala de cinza), é conservado enquanto na operação de mínimo este se perde.

3.2 – Operações Básicas

Podemos considerar as operações morfológicas como sendo, basicamente, comparações com um elemento estruturante escolhido de acordo com um conhecimento a priori das estruturas geométricas a serem analisadas e cuja finalidade é retirar ou

alterar a forma dos objetos que estão relacionados (maiores, menores ou iguais) de alguma maneira a estes elementos estruturantes.

A grosso modo podemos classificar as operações morfológicas em três grandes grupos: filtragem, segmentação ou medidas de imagens. Todas essas operações são construídas através de dois operadores fundamentais da morfologia: a erosão e a dilatação. Esses dois operadores duais podem ser considerados como as duas letras fundamentais do “alfabeto morfológico”, já que todos os outros operadores podem ser expressos em termos de dilatações e erosões.

3.2.1 – Elemento Estruturante

O elemento estruturante é um conjunto plenamente conhecido e definido (forma e tamanho) usado para ser comparado ou medido na imagem a ser estudada a partir de uma transformação. O formato e o tamanho do elemento estruturante permitem o teste e a quantização da maneira em que este está ou não contido na imagem.

O elemento estruturante é freqüentemente referenciado como elemento estrutural plano (flat) quando ele é bidimensional e é normalmente utilizado para operações com imagens bidimensionais. De maneira similar, ele é referenciado como não plano, volumétrico ou de escala de cinza, quando se refere a elementos estruturantes compostos por pequenas imagens em tons de cinza, aplicáveis a operações em imagens em tons de cinza.

Os operadores morfológicos fundamentais necessitam da especificação de uma origem para cada elemento estruturante. De uma forma geral, pode ser dito que a mudança na posição da origem num mesmo elemento estruturante resulta em uma translação do resultado final.

O formato e o tamanho dos elementos estruturantes devem ser adaptados às propriedades geométricas dos objetos das imagens a serem processadas. Por exemplo, quando o interesse é a extração de objetos lineares em uma cena, elementos estruturantes lineares são os mais adequados. Na maioria dos casos os elementos estruturantes devem ser os mais simples possíveis e são escolhidos em função das propriedades de convexidade, não convexidade, isotropia e anisotropia [7].

Na figura 6 temos três elementos estruturantes planos elementares para as grades retangular e hexagonal.

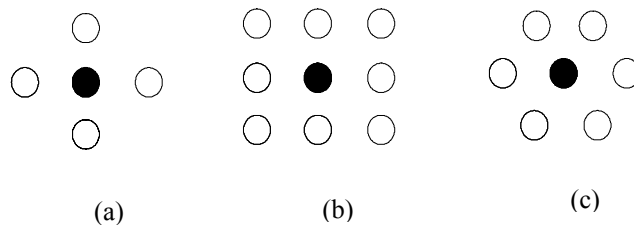


Figura 6: Elementos estruturantes planos isotrópicos para grade retangular e hexagonal. (a) cruz elementar, (b) quadrado elementar e (c) hexágono. A origem destes elementos estruturantes é o seu centro.

Apesar de ser possível estabelecer tipos muito diversos de elementos estruturantes, somente um número limitado destes é usado em aplicações práticas. Alguns destes são:

- *Segmentos de linhas* – usados, normalmente, para extrair estruturas alongadas de imagens. Dois parâmetros importantes para este tipo de elemento estruturante são o tamanho e a orientação.
- *Aproximação digital de discos* – devido a sua isotropia, discos são elementos estruturantes muito utilizados. Várias aproximações de discos são possíveis como a cruz elementar e o quadrado elementar vistos na figura 6.
- *Par de pontos* – a erosão pelo par de pontos permite caracterizar o estado de dispersão de uma estrutura [7].
- *Elementos estruturantes compostos* – compreendem dois elementos estruturantes disjuntos que possuem a mesma origem e são principalmente utilizados na transformada “hit or miss” (seção 3.7).

Em nosso trabalho foi utilizado o elemento estruturante quadrado da figura 6(b), como elemento estruturante elementar, quando não há outra indicação.

3.2.2 - Erosão

Uma das primeiras perguntas que podem ser feitas quando se compara um elemento estruturante e um conjunto ou objeto é: “O elemento estruturante está totalmente dentro do objeto?”. O objeto erodido é o lugar geométrico dos centros dos elementos estruturantes onde esta resposta é afirmativa (vide figura 7). Isso qualifica a erosão binária com os seguintes efeitos [7]:

- . diminuir os objetos.

- . eliminar objetos menores que o elemento estruturante.
- . aumentar os buracos.
- . permitir a separação de objetos próximos.

Considerando X como o conjunto a ser erodido e B como o objeto estruturante escolhido, a erosão (ε_B) pode ser definida como o lugar geométrico dos pontos x nos quais B é incluído em X , quando sua origem está posicionada em x [6]:

$$\varepsilon_B(X) = \{x | B_x \subseteq X\}$$

Essa equação pode ser reescrita como uma interseção de translações do objeto, com essas translações sendo determinadas pelo elemento estruturante [6]:

$$\varepsilon_B(X) = \bigcap_{b \in B} X_{-b}$$

A translação de uma imagem f por um vetor b (f_b) é definido por:

$$f_b(x) = f(x - b)$$

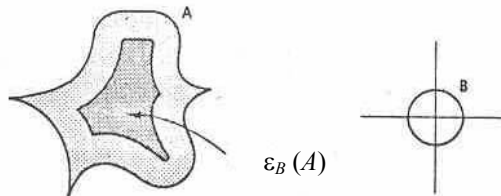


Figura 7: Erosão do objeto A pelo elemento estruturante B

A definição de erosão binária pode ser estendida diretamente em tons de cinza: a erosão de uma imagem f por um elemento estruturante B é definida como o mínimo ou ínfimo das translações de f pelos vetores $-b$ de B .

Considerando ponto a ponto [6]:

$$\varepsilon_B(f) = \bigwedge_{b \in B} f_{-b}$$

Assim, o valor da erosão em um pixel x é o valor mínimo da imagem na janela definida pelo elemento estruturante quando sua origem está em x [6]:

$$[\varepsilon_B(f)](x) = \min_{b \in B} f(x + b)$$

Os efeitos da erosão em tons de cinza são [7]:

- . escurecer a imagem.
- . alargar e engordar os vales (padrões escuros).
- . conectar vales próximos.
- . reduzir e as vezes eliminar picos (padrões claros).
- . separar picos próximos.

Para um elemento estruturante volumétrico (B_v), este tem um valor de cinza associado a cada ponto de seu domínio de definição, e pode-se definir a erosão para esse caso como [6]:

$$[\varepsilon_{B_v}(f)](x) = \min_{b \in B_v} \{f(x + b) - B_v(b)\}$$

Os valores de cinza de um elemento estruturante volumétrico devem ser escolhidos de acordo com os valores de intensidade da imagem. Devido a dificuldade de se estabelecer essa condição de antemão, e ao aumento das operações computacionais por conta da subtração, o elemento estruturante volumétrico é raramente utilizado na prática.

3.2.3 - Dilatação

A dilatação é o operador dual da erosão e está baseada na seguinte pergunta: “O elemento estruturante toca o objeto em análise?”. Onde a resposta é afirmativa o objeto é acrescido do elemento estruturante, compondo assim o objeto dilatado (vide figura 8). Os efeitos da dilatação binária são então [7]:

- . engordar o objeto.
- . preencher pequenos buracos.
- . conectar objetos próximos.

Considerando X como o conjunto a ser dilatado, B o elemento estruturante escolhido, a dilatação (∂_B) pode ser definida como o local dos pontos x onde B toca X quando sua origem coincide com x [6]:

$$\partial_B(X) = \{x | B_x \cap X \neq \emptyset\}$$

Essa equação pode ser rescrita em termos de uma união de translações do objeto, com essas translações sendo definidas pelo elemento estruturante [6]:

$$\partial_B(X) = \bigcup_{b \in B} X_{-b}$$

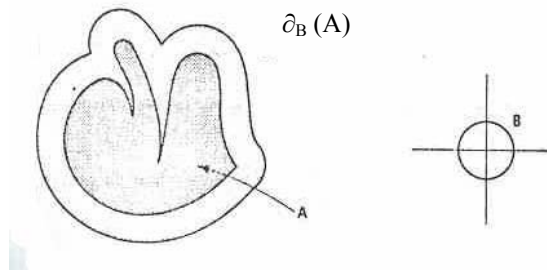


Figura 8: Dilatação do objeto A pelo elemento estruturante B.

Essa última definição pode ser estendida para imagens em tons de cinza: a dilatação de uma imagem f por um elemento estruturante B é definida como o máximo ou supremo das translações de f pelos vetores $-b$ de B :

Considerando ponto a ponto:

$$\partial_B(f) = \bigvee_{b \in B} f_{-b}$$

Em outras palavras, o valor da dilatação em um ponto x é o valor máximo da imagem na janela definida pelo elemento estruturante quando sua origem está em x .

$$[\partial_B(f)](x) = \max_{b \in B} f(x + b)$$

Os efeitos da dilatação em níveis de cinza são [7]:

- . clarear a imagem.
- . alargar e engordar os picos (padrões claros).

- . conectar picos próximos.
- . reduzir e as vezes eliminar vales (padrões escuros).
- . separar vales próximos.

A dilatação por um elemento estruturante não plano é definida por [6]:

$$[\partial_{B_v}(f)](x) = \max_{b \in B_v} \{f(x + b) + B_v(b)\}$$

Diferente de imagens erodidas ou dilatadas por elementos estruturais planos, as imagens dilatadas ou erodidas por elementos estruturantes volumétricos não estão limitadas pelos valores mínimos ou máximos dos tons de cinza da imagem de entrada. Em particular, a erosão com elemento estruturante em tons de cinza pode levar a imagens com resultados negativos, devendo ser usado com bastante parcimônia.

3.2.4 – Propriedades da erosão e da dilatação

1) A dilatação e a erosão são transformações duais, com respeito à complementação. A complementação de uma imagem f é denotada por f^c e é definida para qualquer pixel como o maior valor que um pixel pode assumir (t_{max}) conforme o tipo de imagem (por exemplo 1 para imagens binárias) menos o valor da imagem f na posição x :

$$f^c(x) = t_{max} - f(x)$$

O operador de complementação é representado por C :

$$C(f) = f^c$$

Então para a erosão e dilatação [6]:

$$\varepsilon_B = C \partial_B C$$

Isso significa que qualquer erosão de uma imagem é equivalente a complementação da dilatação de uma imagem complementada com o mesmo elemento estruturante (e vice e versa). Em outras palavras, a erosão encolhe o objeto e expande o fundo, enquanto ocorre o oposto com a dilatação. Convém, entretanto, observar que a

erosão e a dilatação não possuem transformação inversa, por não preservar a homotopia da imagem de entrada.

2) A erosão e a dilatação são invariantes à translação. Essas operações também preservam a ordenação de objetos dentro das imagens, sendo assim, transformadas crescentes [6].

$$f \leq g \Rightarrow \begin{cases} \varepsilon(f) \leq \varepsilon(g) \\ \partial(f) \leq \partial(g) \end{cases}$$

3) A dilatação é distributiva com relação ao operador máximo (\vee) e a erosão com relação ao operador mínimo (\wedge) [6]:

$$\begin{aligned} \partial(\vee_i f_i) &= \vee_i \partial(f_i) \\ \varepsilon(\wedge_i f_i) &= \wedge_i \varepsilon(f_i) \end{aligned}$$

4) As duas equações seguintes são relacionadas com composições referentes a dilatação e erosão [6]:

$$\begin{aligned} \partial_{B_2} \partial_{B_1} &= \partial_{(\partial_{B_2} B_1)} \\ \varepsilon_{B_2} \varepsilon_{B_1} &= \varepsilon_{(\partial_{B_2} B_1)} \end{aligned}$$

Essas duas propriedades, na prática, são bastante úteis porque elas permitem decompor uma operação morfológica com um elemento estruturante grande numa seqüência com elementos estruturantes menores.

5) A erosão de um conjunto com respeito a um elemento estruturante é sempre menor ou igual a dilatação realizada com o mesmo objeto estruturante. Então, se o elemento estruturante B contém sua origem, é possível estabelecer a seguinte relação de ordem (I é o operador identidade) [6]:

$$\varepsilon_B \leq I \leq \partial_B$$

3.2.5 – Operadores de Minkowski

A subtração de Minkowski [6] de um conjunto B por um conjunto X é a interseção das translações de X pelos vetores de B (figura 9), portanto:

$$X \ominus B = \bigcap_{b \in B} X_b$$

A erosão pode ser definida em termos da subtração de Minkowski:

$$\varepsilon_B(x) = X \ominus \check{B}$$

Onde \check{B} é a transposição de um conjunto B que pode ser definido por sua simetria em relação à origem:

$$\check{B} = \{-b / b \in B\}$$

A erosão é, então, equivalente a subtração de Minkowski quando o elemento estruturante é simétrico em relação a origem. Nesse caso:

$$B = \check{B}$$

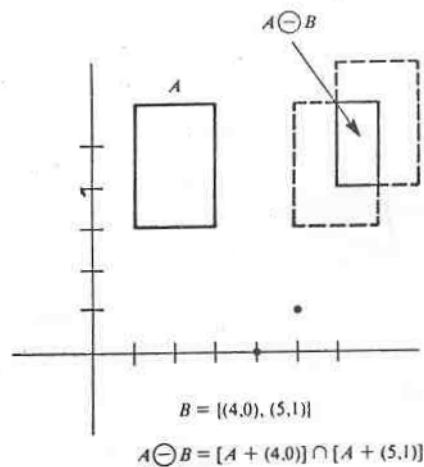


Figura 9: Subtração de Minkowski

A adição de Minkowski [6] (vide figura 10) é definida por:

$$X \oplus B = \bigcup_{b \in B} X_b$$

Da mesma maneira que é possível se rescrever a erosão em termos de operadores de Minkowski temos para a dilatação:

$$\hat{\partial}_{\mathbf{B}}(x) = X \oplus \tilde{B}$$

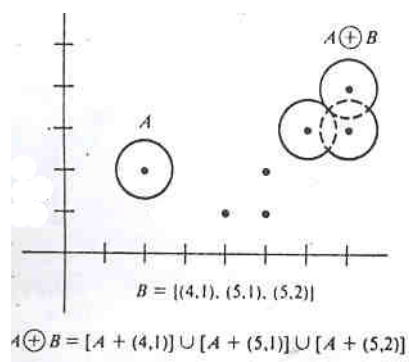


Figura 10: Adição de Minkowski

3.3 – Gradiente Morfológico

É comum assumir que, numa análise de imagem, os objetos estejam contidos em regiões de níveis de cinza mais ou menos homogêneos. Assim, os limites ou bordas desses objetos estão localizados em regiões de variações grandes de níveis de cinza. Os operadores gradiente são normalmente usados para potencializar essas variações. No entanto, alguns cuidados devem ser tomados, como uma filtragem preliminar no caso de imagens sujeitas a ruído, sob pena de se potencializar o componente de ruído.

Os gradientes morfológicos são operadores que aumentam as variações da intensidade de um pixel numa vizinhança determinada por um elemento estruturante. A erosão / dilatação extrai para cada pixel o mínimo / máximo valor da imagem da vizinhança definida pelo elemento estruturante. Variações de combinações entre esses

operadores elementares conduzem a diferentes resultados. As combinações mais usadas são [6]:

- Diferença aritmética entre a dilatação e a erosão.
- Diferença aritmética entre a dilatação e a imagem original.
- Diferença aritmética entre a imagem original e a erosão.

A partir dessas combinações possíveis, os gradientes podem ser denominados como gradiente morfológico básico ou de Beucher, gradiente metade, gradiente espesso ou gradiente direcional.

Gradiente morfológico básico ou gradiente de Beucher – é definido pela diferença entre a dilatação e a erosão pelo elemento estruturante elementar B para a imagem considerada (figura 11):

$$\rho_B = \hat{\sigma}_B - \varepsilon_B$$

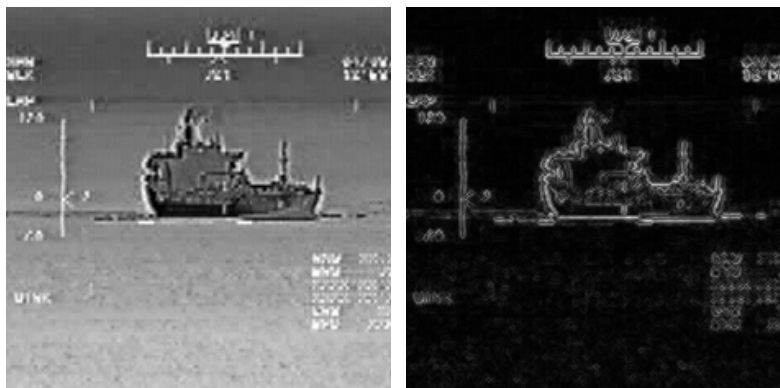


Figura 11: Gradiente básico ou de Beucher. Da esquerda para a direita: imagem original e gradiente básico ou de Beucher

Gradientes metade – podem ser de dois tipos gradiente metade por erosão ou gradiente interno e gradiente metade por dilatação ou gradiente externo (figura 12).

- *Gradiente metade por erosão ou gradiente interno* – é definido pela diferença entre a imagem original e a imagem erodida:

$$\rho_B^- = I - \varepsilon_B$$

- *Gradiente metade por dilatação ou gradiente externo* – é definido pela diferença entre a imagem dilatada e a imagem original:

$$\rho_B^+ = \partial_B - I$$

Os gradientes metade são de bastante valia quando são usados para a detecção da fronteira interna ou externa de uma borda.

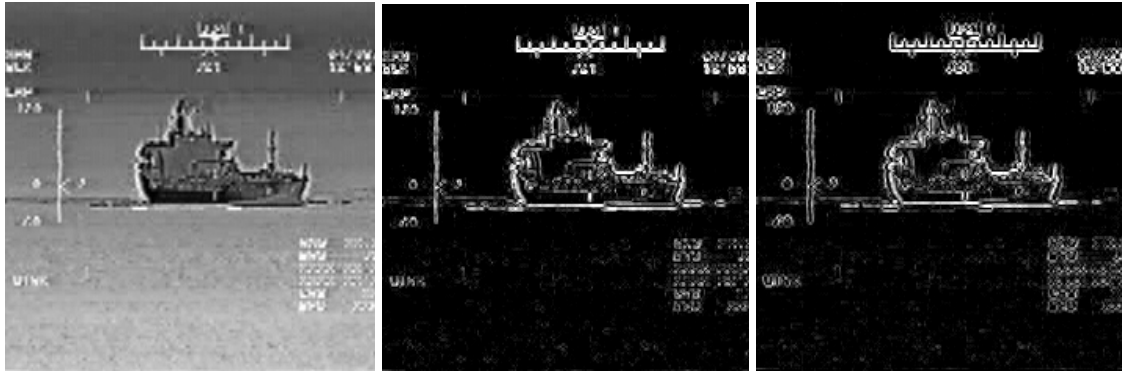


Figure 12: Gradiente interno e gradiente externo. Da esquerda para a direita: imagem original, gradiente interno e gradiente externo.

Gradiente espesso – se o elemento estruturante aplicado no processo do gradiente básico é maior que elemento estruturante elementar (em nosso caso o quadrado elementar), o gradiente morfológico é conhecido como gradiente espesso, predominando bordas dilatadas (figura 13). Sua representação (nB) significa que o elemento estruturante elementar terá seu tamanho aumentado n vezes. Por exemplo o quadrado elementar de 3×3 para um elemento estruturante em que o n de nB valesse 2, corresponderia a um quadrado de 6×6 . Temos então que:

$$\rho_{nB} = \partial_{nB} - \varepsilon_{nB}$$

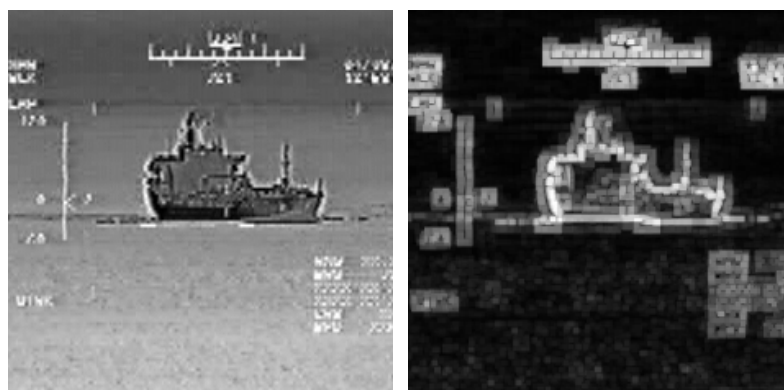


Figure 13 : Gradiente espesso. Da esquerda para direita: imagem original e gradiente espesso.

Gradientes direcionais – são definidos pela substituição de um elemento estruturante isotrópico por um segmento de linha L numa direção α :

$$\rho_{L\alpha} = \partial_{L\alpha} - \varepsilon_{L\alpha}$$

Com o gradiente direcional é possível a filtragem de elementos em determinada direção como pode ser visto na figura 14. Essa característica é de especial importância para a solução do problema da linha do horizonte, como será visto na seção 6.2.3.



Figure 14: Gradiente direcional. Da esquerda para a direita: imagem original, gradiente horizontal (com um elemento estruturante horizontal) e gradiente vertical (com um elemento estruturante vertical).

3.4 - Abertura

A erosão de uma imagem não remove somente todas as estruturas que não contém o elemento estruturante, mas também “encolhe” todas as outras. A procura por um operador capaz de recuperar a maioria das estruturas perdidas pela erosão levou a definição do operador morfológico abertura. A idéia por trás do operador abertura, então, é dilatar a imagem erodida de maneira a recuperar o máximo possível da imagem original.

A abertura de uma imagem f por um elemento estruturante B é definida em termos de erosão e dilatação [6]:

$$\gamma_B(f) = \partial_{\bar{B}}[\varepsilon_B(f)]$$

Uma erosão corresponde a uma interseção de translações. Uma união de translações na direção contrária (isto é, uma dilatação pelo elemento estruturante transposto) deve ser considerado quando se pretende recuperar a imagem original. Conseqüentemente, a abertura de uma imagem é independente da origem do elemento estruturante.

Outra formulação para a abertura é possível em termos geométricos, conservando a mesma pergunta utilizada para a erosão: “O elemento estruturante cabe no objeto?” é mantido, se a resposta for afirmativa, o elemento estruturante completo (no caso da erosão só a origem é mantida). Assim o objeto submetido ao operador abertura é aquele que representa a união de todos os elementos estruturantes que cabem no objeto [6]:

$$\gamma_B(x) = \bigcup \{B | B \subseteq X\}$$

Em linhas gerais pode ser observado que a abertura tem a propriedade de retirar pequenas estruturas (menores que o elemento estruturante), mas modificando também todos os objetos da imagem em cujo o operador abertura foi aplicado.

Na figura 15 pode ser vista a abertura em termos de dilatação e erosão.

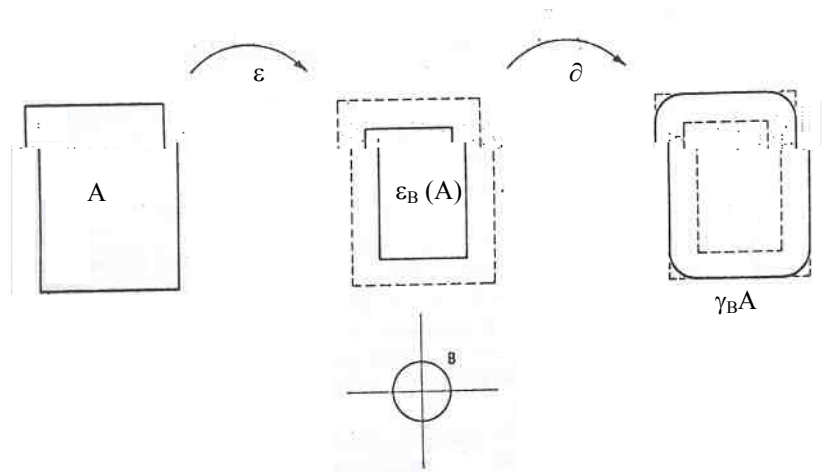


Figura 15: Abertura em termos de dilatação e erosão

3.5 - Fechamento

A idéia por trás do fechamento morfológico é a recuperação do formato inicial de estruturas que tenham sido dilatadas. Essa operação é realizada através da erosão do objeto dilatado.

O fechamento de uma imagem f por um elemento estruturante B pode ser escrito em termos de erosão e dilatação como [6]:

$$\varphi_B(f) = \varepsilon_{\bar{B}}[\delta_B(f)]$$

A questão “O elemento estruturante cabe no fundo da imagem?” pode ser aplicada para definição do fechamento. Se a resposta for afirmativa, todos os pontos do elemento estruturante pertencentes ao complemento do fechamento do objeto serão considerados [6]:

$$\varphi_B(x) = [\cup\{B \mid B \subseteq X^c\}]^c$$

Uma formulação equivalente a esta é aquela que considera a interseção de todas as translações do complemento do elemento estruturante B de tal modo que ele contenha x [6]:

$$\varphi_B(x) = \cap\{B^c \mid x \subseteq B^c\}$$

Todas as estruturas do fundo da imagem que não contém o elemento estruturante são preenchidas pelo fechamento.

Na figura 16 pode ser visto o fechamento em termos de dilatação e erosão.

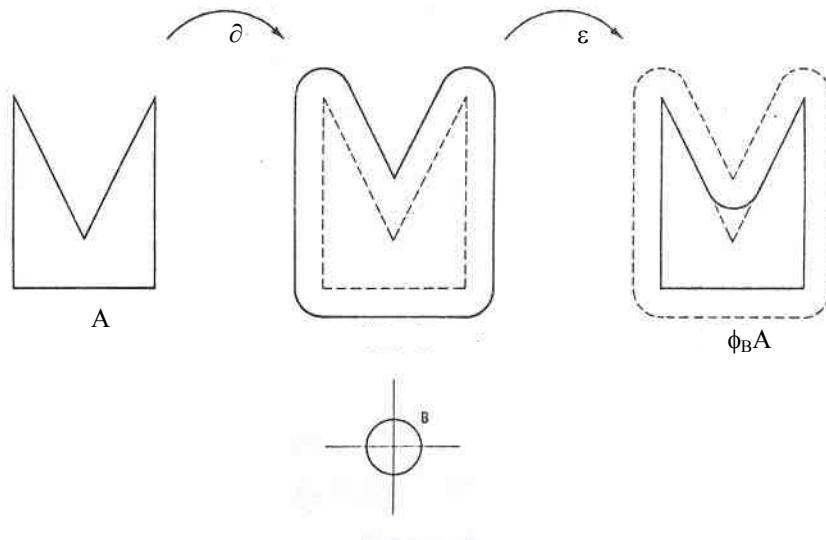


Figura 16: Fechamento em termos de dilatação e erosão.

3.6 – Propriedades da abertura e fechamento

1) Uma abertura remove os pixels do objeto que não podem ser cobertos pelas translações dos elementos estruturantes que tocam os objetos da imagem. O fechamento tem um comportamento oposto porque ele adiciona pixels no fundo que não podem ser cobertos pelas translações que tocam os objetos da imagem. Assim, abrir uma imagem corresponde a fechar o complemento da imagem e depois, complementar o resultado. Portanto a abertura e o fechamento são operações duais [6]:

$$\gamma_B = C\phi_B C$$

2) Aberturas são transformações anti-extensivas (alguns pixels são removidos) e fechamentos são extensivas (alguns pixels são acrescentados). Portanto estes operadores sempre satisfazem a seguinte relação de ordem (I é o operador identidade) [6]:

$$\gamma \leq I \leq \phi$$

3) A abertura e o fechamento são transformações crescentes:

$$f \leq g \Rightarrow \begin{cases} \gamma(f) \leq \gamma(g) \\ \phi(f) \leq \phi(g) \end{cases}$$

4) A abertura e o fechamento também são idempotentes, ou seja, depois de executadas uma vez, sua posterior execução não modifica a imagem.

$$\begin{aligned} \gamma\gamma &= \gamma \\ \phi\phi &= \phi \end{aligned}$$

A idempotência é considerada uma importante propriedade para um filtro porque garante que a imagem não sofrerá modificações futuras pela iteração da transformação.

3.7 – Transformação “hit or miss”

A idéia básica por trás da transformação “hit or miss” consiste em se extrair pixels de uma imagem binária que possuam uma dada configuração de uma vizinhança, como um pixel da frente isolado por pixels traseiros. Em outras palavras, pretende-se, com a transformada “hit or miss” testar a parte interna e externa de um conjunto. Para isso são necessários dois elementos estruturantes, que formarão um elemento estruturante composto que, obrigatoriamente, possuam uma única origem.

A partir do deslocamento do elemento estruturante composto, para toda a posição possível da imagem pode ser feita uma pergunta, para a realização da transformada “hit or miss”: “O primeiro elemento estruturante toca o objeto, enquanto, simultaneamente, o segundo elemento estruturante o perde (ou toca o fundo)?” Se a resposta for positiva, o ponto definido como a origem do elemento estruturante composto é um ponto da imagem transformada pelo operador “hit or miss”.

Pode ser definida a transformação “hit or miss” de uma imagem X por um elemento estruturante $B(B_1, B_2)$ como [6]:

$$HMT_B(X) = \{x | (B_1)_x \subseteq X \text{ e } (B_2)_x \subseteq X^c\}$$

A transformação “hit or miss” também pode ser definida em termos de uma interseção de duas erosões [6]:

$$HMT_B(X) = \varepsilon_{B_1}(X) \cap \varepsilon_{B_2}(X^c)$$

Por definição B_1 e B_2 devem possuir a mesma origem. Estes precisam também ser disjuntos [6]:

$$B_1 \cap B_2 = 0$$

De outra maneira, a saída da transformação “hit or miss” será um conjunto vazio para qualquer que seja X .

A transformação “hit or miss” é usada, geralmente, para encontrar configurações específicas de pixels em imagens.

3.8 - Afinamento

O afinamento consiste em remover os pixels que pertencem a objetos que tem uma dada configuração. Em outras palavras, a transformada “hit or miss” da imagem é subtraída da imagem original.

O afinamento de uma imagem binária ($X \circ B$) por um elemento estruturante composto B é definida pela diferença de conjuntos entre X e a transformada “hit or miss” de X por B [6] (figura 17):

$$X \circ B = X - HMT_B(X)$$

A origem do elemento estruturante tem que pertencer ao componente B_1 do elemento estruturante composto, de outra maneira a operação terá o mesmo resultado que a transformada identidade. Por definição a operação de afinamento é anti-extensiva e não crescente.

Para imagens em tons de cinza, o valor de cinza na posição do espessamento de uma imagem X por um elemento estruturante composto B num pixel x é igualado a valor da dilatação da imagem pelo elemento estruturante B_2 se e somente se este for menor que o valor da imagem e esse valor da imagem for igual a erosão da imagem por B_1 na posição x . De outra forma o valor de x não é modificado. Assim [6]:

$$(f \circ B)(x) = \begin{cases} [\partial_{B_2}(f)](x) & \text{se } [\partial_{B_2}(f)](x) < f(x) \text{ e } f(x) = [\varepsilon_{B_1}(f)](x) \\ f(x) & \text{qualquer outro} \end{cases}$$



Figura 17: Afinamento. Da esquerda para a direita: imagem original e imagem afinada

3.9 – Espessamento

O espessamento consiste em adicionar pixels de fundo com uma configuração específica para o conjunto de pixels do objeto.

Para o caso binário o espessamento ($X \odot B$) de uma imagem X por um elemento estruturante composto B é definido como a união da transformação “hit or miss” da imagem com a imagem original [6] (figura 18):

$$X \odot B = X \cup HMT_B(X)$$

Para o espessamento, a origem do elemento estruturante deve pertencer a B_2 . O espessamento é extensivo e decrescente. O afinamento e o espessamento são funções duais [6]:

$$X \odot B = (X^c \circ B^c)$$

Onde $B = (B_1, B_2)$ e $B^c = (B_2, B_1)$.

Para o caso da imagem em tons de cinza, o espessamento de uma imagem X por um elemento estruturante composto B num pixel x é definido como o valor erodido da imagem por B_1 , se esse valor é maior que o valor original da imagem em x e se o

valor dilatado por B_2 é igual ao valor da imagem em x . De outra maneira o valor do espessamento é o próprio valor de x . Assim:

$$(f \odot B)(x) = \begin{cases} [\varepsilon_{B_2}(f)](x) & \text{se } [\partial_{B_2}(f)](x) = f(x) \text{ e } f(x) < [\varepsilon_{B_1}(f)](x) \\ f(x) & \text{qualquer outro} \end{cases}$$



Figura 18: Espessamento. Da esquerda para a direita: imagem original e imagem após o espessamento.

3.10 - Esqueleto

Em reconhecimento de padrões e análise de formato são necessários métodos para extração das características de um objeto. Uma abordagem possível consiste em afinar o objeto com a intenção de formar um conjunto de linhas finas que condensem a informação do objeto original preservando sua homotopia. Esse conjunto de linhas finas é o chamado esqueleto do objeto. A transformação associada a produção do esqueleto chama-se esqueletização e a detecção de pontos finais, pontos múltiplos e malhas fechadas a partir do esqueleto são algumas das ações importantes para as tarefas de reconhecimento de formato.

Várias definições para o esqueleto de um conjunto euclidiano estão disponíveis, mas conduzem sempre a um mesmo esqueleto. Como exemplo, podem ser citados os processos conhecidos como “fogo na grama” ou “propagação de frente de onda” e “discos máximos”.

No caso do “fogo na grama” ou “propagação de frente de onda” se for assumido que fogo, ou uma frente de onda, é iniciada dos pontos de contorno de um objeto X e estes se propagam em velocidade uniforme dentro do objeto, o esqueleto de X é, então, o conjunto de pontos onde o fogo ou as ondas se encontram. Por exemplo o

esqueleto de um disco é o seu centro, porque é o lugar onde o fogo ou as frentes de onda se encontram simultaneamente.

Um disco B é máximo em X se não existe nenhum outro disco maior que B contido em X com o mesmo centro. O esqueleto de um objeto X é o conjunto dos centros desses discos máximos (vide figura 19).

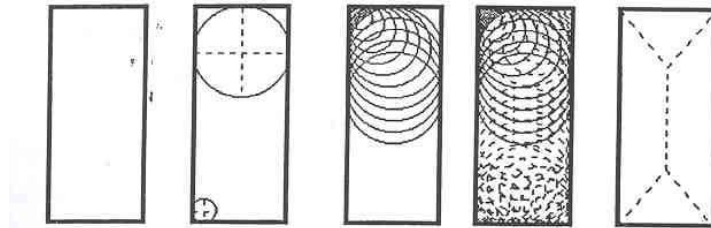


Figura 19: Esqueleto através da noção de disco máximo.

A extensão da definição de esqueleto para imagens discretas (que é o que nos interessa) conduz a uma variedade de esqueletos. Noções como as descritas anteriormente como “propagação de frente de onda” ou “disco máximo” não tem equivalente discreto único e direto. Então a adaptação das definições existentes para o caso contínuo conduzem a uma grande variedade de esqueletos que não compartilham das mesmas propriedades, sendo dependentes do elemento estruturante escolhido. Além disso, muitas adaptações dos processos de esqueletização conduzem a esqueletos não conectados, que são úteis em processos de compressão, mas para a descrição de formato são de pouca utilidade. Alguns procedimentos para a extração do esqueleto discreto serão brevemente descritos.

a) por abertura – Esse processo utiliza quadrados como aproximação dos discos como elemento estruturante e pode ser definido como [6]:

$$SK(X) = \bigcup_{\lambda \geq 0} \{\mathcal{E}_{\lambda B}(X) - \gamma_B[\mathcal{E}_{\lambda B}(X)]\}$$

Onde X é a imagem, B é o quadrado elementar de tamanho três e λ_B é um quadrado de comprimento $2\lambda+1$. Na figura 20 pode ser observado que o esqueleto resultante não é conectado nem possui espessura de um pixel.



Figura 20 : Esqueleto por abertura. Da esquerda para a direita: imagem original e esqueleto por abertura.

b) *afinamento homotópico* – Um esqueleto não conectado como o da figura 20 é de pouca utilidade para aplicações de descrição de formatos de objetos porque a homotopia não é preservada e pontos característicos, como pontos de extremo, não são preservados como no caso contínuo. Uma esqueletização de uma imagem discreta pode preservar estas propriedades importantes se esta for executada com elementos estruturantes que preservem sua homotopia, ou seja, elementos estruturantes homotópicos (figura 21). Um elemento estruturante é definido como homotópico se o afinamento de um objeto por este elemento estruturante não modifica a homotopia deste objeto. A homotopia é uma propriedade topológica de um conjunto e pode ser definida como [7]:

Dois conjuntos X e Y são homotópicos se, por meio de deformação contínua, o primeiro conjunto possa sempre ser sobreposto ao segundo. Uma transformação é homotópica se ela transforma qualquer conjunto X em outro conjunto homotópico.

Como exemplo, uma transformação homotópica aplicada em um conjunto conexo simples geraria outro conjunto conexo simples, um conjunto apresentando um furo geraria um outro conjunto apresentando um furo. Um elemento estruturante é homotópico, então, se o número de componentes conectados do objeto e do fundo não são modificados.

O esqueleto é obtido através do afinamento da imagem de entrada por um ou uma série de elementos estruturantes homotópicos e sua rotação até a estabilidade. Esse processo é chamado de afinamento seqüencial [6]. A fórmula para o afinamento seqüencial ($X \underline{O} B$) de X por n rotações de um elemento estruturante B ($\theta_1 B, \theta_2 B, \dots \theta_n B$) é:

$$X \underline{O} B = (\dots ((X \underline{O} \theta_1 B) \underline{O} \theta_2 B) \dots) \underline{O} \theta_n B$$

O afinamento seqüencial é repetido até a estabilidade, isto é até que a entrada tenha sido reduzida a um conjunto de linhas afinadas conectadas. Cada elemento estruturante homotópico define um tipo diferente de esqueleto.



Figura 21 : Esqueleto por afinamento homotópico. Da esquerda para a direita: imagem original e esqueleto por afinamento homotópico.

3.11 – Esqueleto por zona de influência

O conjunto de pixels de uma imagem binária que estão mais perto de um dado objeto conectado que qualquer outro objeto conectado define a zona de influência do objeto conectado considerado. O esqueleto por zona de influência (SKIZ) é definido como os pontos que não pertencem a nenhuma zona de influência. Então, as fronteiras das zonas de influência definem o esqueleto por zona de influência (figura 22).

Em termos matemáticos uma imagem X que possui os objetos conectados K pode ser representada por[6]:

$$X = \bigcup_{i=1}^n K_i, \quad i \neq j \Rightarrow K_i \neq K_j$$

A zona de influência de uma partícula K_i é o conjunto de pixels de X que estão mais perto de K_i que quaisquer outros [6]:

$$IZ(K_i) = \{x | \forall j \in \{1, \dots, n\}, i \neq j \Rightarrow d(x, K_i) < d(x, K_j)\}$$

O esqueleto por zona de influência é definido, então, como os pontos que não pertencem a nenhuma zona de influência [6]:

$$SKIZ(X) = \left[\bigcup_i IZ(K_i) \right]^c$$

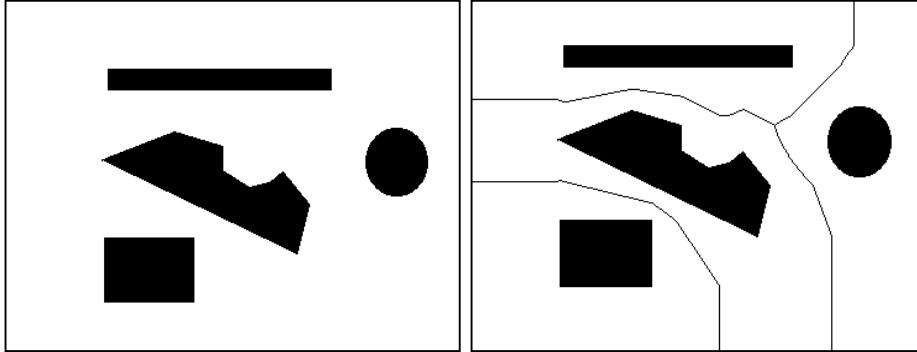


Figura 22: Esqueleto por zona de influência: a esquerda imagem original, a direita imagem original mais esqueleto por zona de influência.

Uma aplicação do esqueleto por zona de influência pode ser encontrada na tarefa de estabelecer as watersheds da imagem (vide seção 3.14.1.2).

3.12 – Transformações geodésicas

A abordagem relacionada com as transformações geodésicas conduz à utilização de duas imagens de entrada. Esse tipo de operação se torna bastante útil quando é necessário alterar apenas algumas estruturas da imagem deixando outras intactas. Um operador morfológico é aplicado à primeira imagem e o resultado é forçado a permanecer maior ou menor do que a segunda imagem, dependendo da operação realizada. Serão brevemente descritas aqui algumas transformações geodésicas: a dilatação geodésica, a erosão geodésica e a reconstrução morfológica.

3.12.1 – Dilatação geodésica

A dilatação geodésica envolve duas imagens: uma imagem marcadora e uma imagem máscara. Ambas as imagens devem possuir o mesmo domínio, e a imagem máscara deve ser maior ou igual à imagem marcadora (vide figura 23(a) e 23(b)). Uma imagem f é maior ou igual a uma imagem g de mesmo domínio se todos os valores dos

elementos x de f são maiores ou iguais aos seus correspondentes (localizados na mesma posição) de g . A imagem marcadora é dilatada, primeiramente, por um elemento estruturante elementar. A imagem dilatada resultante é forçada a permanecer menor que a imagem máscara. A imagem máscara age como se fosse um limite para a dilatação da imagem marcadora (figura 23).

Considerando f como a imagem marcadora e g a imagem máscara a dilatação geodésica ($\partial_g^{(1)}$) pode ser definida como [6]:

$$\partial_g^{(1)} = \partial^{(1)}(f) \wedge g$$

Onde $\partial^{(1)}$ é a dilatação elementar (realizada a partir do elemento estruturante elementar definido em 3.2.1) e $f \leq g$.

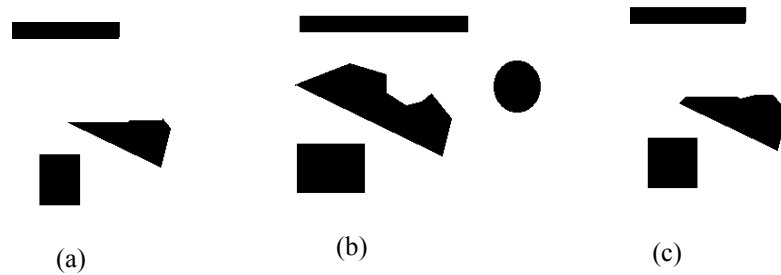


Figura 23: Dilatação geodésica: (a) Imagem marcadora (b) Imagem máscara (c) Dilatação geodésica realizada 10 vezes com quadrado elementar como elemento estruturante. Note que, como não existe uma semente para o círculo na imagem de marcadora, este não aparece no resultado da dilatação. Observe ainda que o limite imposto pela imagem máscara é respeitado nos outros objetos.

3.12.2 – Erosão geodésica

A erosão geodésica ($\varepsilon_g^{(1)}$) é a transformação dual da dilatação geodésica com respeito ao complemento da imagem (figura 24). Pode ser definida como [6]:

$$\varepsilon_g^{(1)}(f) = \varepsilon^{(1)}(f) \vee g$$

Onde $\varepsilon^{(1)}$ é a erosão elementar (realizada a partir do elemento estruturante elementar definido em 3.2.1) e $f \geq g$.

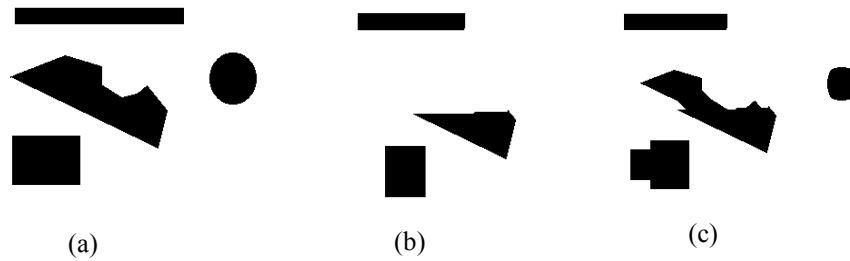


Figura 24 : Erosão geodésica: (a) Imagem marcadora (b) Imagem máscara (c) Erosão geodésica realizada 10 vezes com o quadrado elementar como elemento estruturante. Observe que existe um limite para os retângulos: a erosão destes é interrompida ao se atingir o tamanho destes objetos na imagem máscara.

3.12.3 – Reconstrução Morfológica

As dilatações e erosões geodésicas de um tamanho determinado são raramente usadas na prática [6]. Sua iteração até a estabilidade, no entanto, possibilita a definição de poderosos algoritmos de reconstrução morfológica.

Dilatações e erosões geodésicas sempre convergem depois de um número finito de iterações em que o espessamento ou afinamento da imagem marcadora é controlado pela imagem máscara.

3.12.3.1 – Reconstrução Morfológica por Dilatação

A reconstrução morfológica por dilatação [6] pode ser compreendida como uma dilatação condicional. A partir de uma “semente” definida entre os objetos de uma cena, na imagem marcadora, pode ser realizada uma dilatação a partir de um elemento estruturante elementar. O resultado dessa operação será comparado com uma máscara, e o resultado a ser considerado será a interseção dessas duas imagens. O processo é repetido até o equilíbrio, ou seja, até que o resultado seguinte seja igual ao anterior. Com essa operação é possível a extração ou a recuperação de elementos, desde de que se saiba sua posição (figura 25).

A reconstrução morfológica por dilatação ($R_g(f)$) de uma imagem f com imagem máscara g pode ser definida como:

$$R_g(f) = \partial_g^i(f)$$

Onde i é tal que:

$$\partial_g^i(f) = \partial_g^{i+1}(f)$$

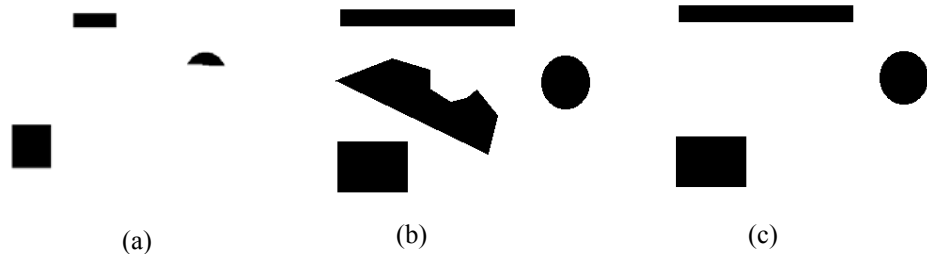


Figura 25 : Reconstrução morfológica por dilatação: (a) Imagem marcadora (b) Imagem máscara (c) Reconstrução morfológica por dilatação. Repare que só os elementos que possuem “sementes” na imagem marcadora são recuperados.

3.12.3.2 – Reconstrução morfológica por erosão

Na reconstrução morfológica por erosão, a erosão geodésica é que realiza a iteração até a estabilidade. Como, nesse caso, a imagem f é menor que a imagem máscara g , os objetos contidos na imagem são encolhidos até o limite imposto por g (figura 26).

A reconstrução morfológica por erosão ($R_g^*(f)$) de uma imagem f com uma imagem máscara g :

$$R_g^*(f) = \varepsilon_g^i(f)$$

Onde i é tal que:

$$\varepsilon_g^i(f) = \varepsilon_g^{i+1}(f)$$

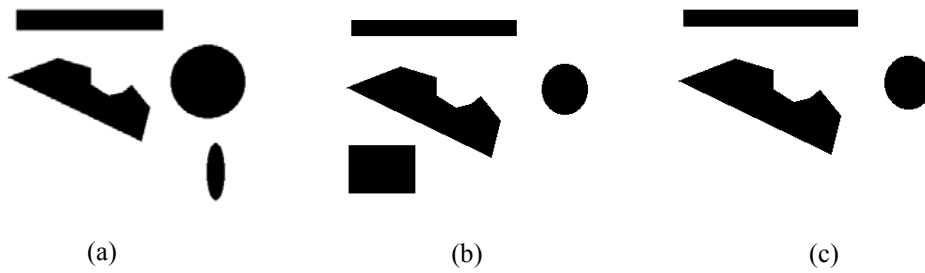


Figura 26: Reconstrução morfológica por erosão. (a) Imagem marcadora (b) Imagem máscara (c) Reconstrução morfológica por erosão . Repare que só os elementos que possuem limites na imagem máscara e “sementes” na imagem marcadora permanecem.

3.13 – Filtragem Morfológica

Em processamento de sinais a filtragem linear geralmente é usada, por exemplo, para as tarefas de remover ou extrair frequências altas ou baixas de imagens ou, genericamente falando, resolver problemas ligados a fenômenos lineares.

Filtros morfológicos são filtros não lineares que são adequados, principalmente, a dois tipos de tarefas: a restauração de imagens corrompidas por algum tipo de ruído e a remoção seletiva de objetos ou estruturas de imagens conservando todas as outras intactas. Para essa última tarefa pode ser citado o filtro de abertura por área.

3.13.1– Abertura por área

Numa imagem discreta, pode acontecer que pequenos elementos formados por pixels conectados sejam ruídos ou objetos irrelevantes para essa imagem. Uma filtragem de muita utilidade para esses casos vem a ser a que remove todos os componentes conectados cuja área da superfície seja menor que um valor de limiar fornecido (λ). Esse é a chamada abertura por área e pode ser definido como γ_λ :

$$\gamma_\lambda = \bigvee_i \{ \gamma_{B_i} / B_i \text{ é conectado e a área } B_i > \lambda \}$$

A abertura por área é uma espécie de filtro adaptativo em que não é importante o formato do elemento estruturante e sim seu tamanho. Pode ser dito que este adapta-se ao formato do elemento a ser retirado.

3.14 – Segmentação Morfológica e Transformação Watershed

Assumindo que objetos numa imagem são regiões formadas por pixels conectados de mesmo valor, no caso de imagens binárias, ou com pequenas variações de tons de cinza, no caso de imagens em tons de cinza, é possível extrair esses objetos ou regiões usando as propriedades de sua vizinhança. De fato, uma diferença muito grande em níveis de cinza entre dois pixels vizinhos pode indicar que estes pertençam a dois objetos distintos.

Basicamente, existem duas abordagens de segmentação: detecção de bordas e crescimento de regiões. No caso do crescimento de regiões, primeiramente são localizadas regiões homogêneas. O crescimento destas regiões é baseado em medidas de similaridade combinando atributos espaciais ou espectrais e é realizado até que todos os pixels da imagem estejam associados a uma região. As fronteiras das regiões, que delimitam o objeto segmentado, são criadas quando o processo de crescimento das regiões faz com que duas ou mais regiões se encontrem.

As técnicas de detecção de borda atuam de maneira oposta. Um objeto contido numa imagem normalmente possui, dentro dele, pouca variação em tons de cinza. Então, uma variação brusca de nível de cinza na vizinhança de um pixel caracteriza essas bordas. Assim a tarefa de um detetor de bordas é encontrar essas variações.

A principal ferramenta da morfologia matemática para as técnicas de segmentação é a transformação watershed, que consiste numa combinação das técnicas de crescimento de região e detecção de bordas.

3.14.1 – Transformação Watershed

O conjunto de todos os pontos $\{x, f(x)\}$ pertencentes a uma imagem pode ser considerado como uma superfície topográfica S . Quanto mais claro é o tom de cinza f no

ponto x , maior é a altitude do ponto correspondente. Partindo disso, o conceito de watersheds é relativamente simples: a imagem vista topograficamente possuirá “vales” e “picos” de diversas alturas. Se for considerado que água escorre por esses “vales” até atingirem seus pontos mais baixos, e essa água continua a escorrer até que esses “vales” sejam inundados um a um com diques em suas linhas divisórias, para impedir que a inundação de um vale invada outro, essas linhas divisórias resultantes dessas múltiplas inundações são as watersheds (vide figura 27). Historicamente, a segmentação baseada na transformação watershed foi introduzida por Beucher e Lantuéjoul [15], e foi definida matematicamente por Meyer [16] e Najman e Schmitt [17].

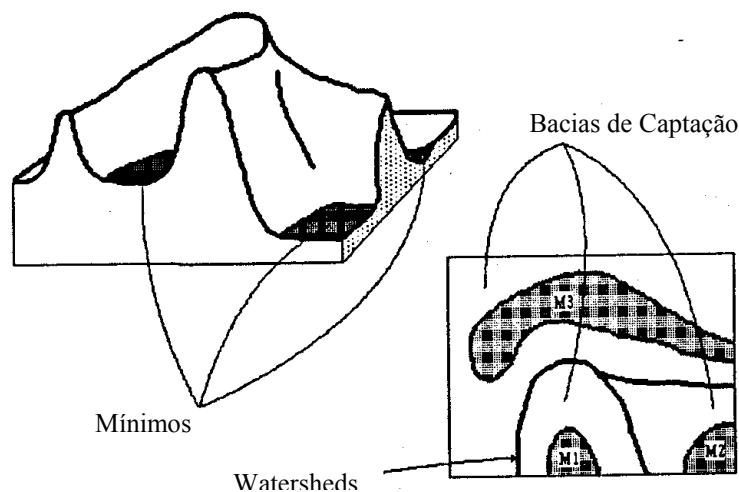


Figura 27: Mínimos, bacias de segmentação e Watersheds.

3.14.1.1 – Definição em termos de inundação

O conceito de água escorrendo para encher os vales da imagem considerada como superfície topográfica não é adequado para uma definição formal, pois existem muitas situações onde a direção da água escorrendo em um ponto não é determinada. Entretanto, uma definição em termos de inundação, como será vista a seguir, é bastante útil a esse propósito.

Considera-se o modelo da imagem como uma superfície semelhante a um relevo composto por vales e picos diversas altitudes. Só que em cada um desses vales, agora, existe um pequeno furo. Imaginado que essa superfície é lentamente submersa

em uma piscina com água entrando pelos furos, os vales serão inundados vagarosamente. Inicialmente as bacias de menor altitude começam a encher, mas cada vez que a água atinge a altitude de um mínimo regional, uma nova bacia começa a encher. Cada um desses vales representa uma bacia de captação ou retenção, e o processo de enchimento de cada uma dessas bacias prossegue até que toda a superfície esteja submersa.

No instante em que duas bacias de captação se juntam, deve ser construído um dique para evitar que as águas se misturem e, a medida que o nível da água sobe, este dique deve ser aumentado para impedir que as águas se misturem. Esses diques formam as watersheds.

Agora esse processo de inundação será formalizado. O menor valor assumido para uma imagem em tons de cinza f no domínio D_f é denotado por h_{min} e o maior valor h_{max} . A bacia de captação associada a um mínimo M é denominada de $BC(M)$. Os pontos, na bacia de captação, que possuem altitude menor ou igual a h são chamados de $BC_h(M)$:

$$BC_h(M) = \{p \in BC(M) / f(p) \leq h\}$$

X_h é o subconjunto de todas as bacias de captação que possuem um valor menor ou igual a h :

$$X_h = \bigcup_i BC_h(M_i)$$

Finalmente, o conjunto de pontos que pertencem ao mínimo regional de elevação h é denominado por $RMIN_h(f)$.

As bacias de captação são construídas progressivamente, pela simulação do processo de inundação. Os primeiros pontos a serem atingidos pela água são os pontos de menor valor. Esses pontos pertencem a um mínimo regional da imagem no nível h_{min} . Estes também são equivalentes a $X_{h_{min}}$:

$$X_{h_{min}} = T_{h_{min}}(f) = RMIN_{h_{min}}(f)$$

Onde T é um operador limiar que obedece a seguinte relação:

$$T_h(f)(x) = \begin{cases} x, & \text{se } h \leq f(x) \\ 0, & \text{qualquer outro} \end{cases}$$

A definição de $X_{h_{\min}+1}$ é baseada na análise do processo de inundação até a elevação $h_{\min}+1$. A água tanto expande as regiões que já foram atingidas como inicia a captação em bacias de captação cujo mínimo tem uma altitude igual a $h_{\min}+1$. Isso é ilustrado pela figura 28 onde são visualizadas as três possíveis relações entre três componentes conexos Y , pertencentes a imagem, e $X_{h_{\min}}$. Esses componentes conexos representados por Y são formados pela inundação de um objeto da imagem.

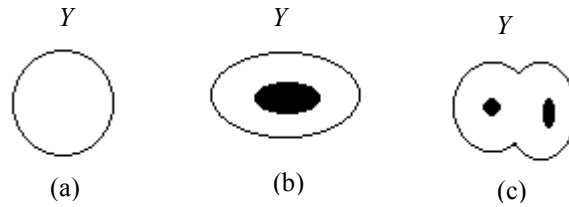


Figura 28: Relações de inundação de objetos em uma imagem. Em (a) um novo mínimo vai ser encontrado num nível maior ou igual a $h_{\min}+1$. Em (b) A região inundada da bacia de captação cujo mínimo está em h_{\min} está expandindo. Em (c) as regiões da bacia de captação de dois mínimos distintos no nível h_{\min} estão expandindo e imergirão juntos.

As relações são:

a) $Y \cap X_{h_{\min}} = \emptyset$ (figura 28 (a)) o processo de inundação ainda não formou Y , então a altura mínima de Y estará num novo mínimo regional de f num nível maior ou igual a $h_{\min}+1$:

$$\forall p \in Y \begin{cases} p \notin X_{h_{\min}} & \Rightarrow f(p) \geq h_{\min} + 1 \\ p \in Y & \Rightarrow f(p) \leq h_{\max} \end{cases}$$

b) $Y \cap X_{h_{\min}} \neq \emptyset$ e é conectado (figura 28 (b)). Nesse caso o processo de inundação já atingiu o mínimo de Y e esse caso corresponde, então, à expansão de um nível já atingido em h_{\min} e que pode ser descrito por:

$$Y = BC_{h_{\min}+1}(Y \cap X_{h_{\min}}) = IZ_Y(Y \cap X_{h_{\min}})$$

Onde $IZ_Y(Y \cap X_{h_{\min}})$ é a zona de influência geodésica de $(Y \cap X_{h_{\min}})$ dentro de Y . A zona de influência geodésica $(IZ_A(K_i))$ de um componente conectado K_i de um conjunto X que está contido num conjunto maior A é o lugar geométrico dos pontos de A onde a distância geodésica para K_i é menor que sua distância geodésica para qualquer outro componente de X [6]:

$$IZ_A(K_i) = \{p \in A, \forall j \in [1, N] - \{i\}, d_A(p, K_i) < d_A(p, K_j)\}$$

A distância geodésica $d_A(p, q)$ entre dois pixels p e q em A é o menor comprimento L do conjunto de caminhos $P = (p_1, p_2, \dots, p_l)$ que liga p a q que é incluído em A [6]:

$$d_A(p, q) = \min\{L(P) \mid P \text{ é o conjunto de caminhos incluído em } A\}$$

Do mesmo modo a distância geodésica de um ponto p e um subconjunto Y contido em A é a menor distância geodésica entre p e qualquer pixel y de Y [6]:

$$d_A(p, Y) = \bigwedge_{y \in Y} d_A(p, y)$$

c) $Y \cap X_{h_{\min}} \neq \emptyset$ e não é conectado. Nesse caso, Y contém mais do que um mínimo de f no nível h_{\min} . Então dois ou mais mínimos estão expandindo juntos. Esses podem ser chamados de Z_1, Z_2, \dots, Z_k e considerando Z_i como um destes, pode-se estabelecer que a melhor aproximação para $BC_{h_{\min}+1}(Z_i)$ corresponde a zona de influência geodésica de Z_k dentro de Y (IZ_y):

$$BC_{h_{\min}+1}(Z_i) = IZ_y(Z_i)$$

As relações b) e c) correspondem a expansão de regiões de bacias de captação já alcançadas pela água. Essas regiões expandidas podem ser definidas em termos de uma única zona de influência geodésica de $X_{h_{\min}}$. Assim, $X_{h_{\min}+1}$ é definida como a união dessas zonas de influência geodésicas com os mínimos regionais descobertos mais recentemente:

$$X_{h_{\min}+1} = RMIN_{h_{\min}+1}(f) \cup IZ_{T_{\leq h_{\min}+1}(f)}(X_{h_{\min}})$$

Essa recursão é válida para todos os níveis h . O conjunto de bacias de captação de uma imagem f é igual a $X_{h_{\max}}$, desde que todos os níveis tenham sido inundados:

$$\begin{aligned} i) X_{h_{\min}} &= \text{todos os } h_{\min} \text{ de } f \\ ii) \forall h \in [h_{\min}, h_{\max}-1], X_{h+1} &= RMIN_{h+1}(f) \cup IZ_{T_{\leq h+1}(f)}(X_h) \end{aligned}$$

O conjunto das bacias de captação de uma imagem representa uma imagem com regiões rotuladas correspondentes às bacias de captação relacionadas a mínimos regionais. As watersheds de f correspondem às fronteiras das bacias de captação de f .

Um algoritmo rápido para a implementação do processo de inundação é apresentado em [6] e [18].

3.14.1.2 – Definição em termos de geodésia

3.14.1.2.1 - Introdução

As bacias de captação de uma imagem podem também ser consideradas como zonas de influência dos mínimos regionais da imagem [6]. Nesse sentido, as watersheds são simplesmente os esqueletos por zona de influência dos mínimos da imagem. Essa abordagem pode ser formalizada a partir do uso de algumas noções de geodésia genérica, que serão brevemente descritos aqui.

3.14.1.2.2 – Geodésia Genérica

A geodésia genérica representa as operações geodésicas com marcadores e máscaras como imagens em tons de cinza. Assim a expansão ou encolhimento da marcadora é controlado pela intensidade dos valores de pixel da máscara em tons de cinza.

Tempo geodésico: Considerando o caminho P definido no domínio da imagem, ou função integrável f , o tempo requerido para cobrir P é definido como a integral de f ao longo de P , e é chamado de tempo geodésico [6]:

$$\tau_f(P) = \int_P f(s) ds$$

Para imagens discretas o tempo geodésico ($\tau_f(P)$) necessário para cobrir o caminho discreto P de tamanho l definido numa imagem discreta f é igual a média dos valores de f tomados dois a dois ao longo de P [6]:

$$\tau_f(P) = \sum_{i=1}^l \frac{f(p_{i-1}) + f(p_i)}{2} = \frac{f(p_0)}{2} + \frac{f(p_1)}{2} + \sum_{i=1}^{l-1} f(p_i)$$

$\tau_f(P)$ é independente da direção usada para cobrir o caminho [6].

O tempo geodésico $\tau_f(p, q)$ separando dois pontos p e q numa imagem em tons de cinza f é a menor quantidade de tempo necessária para unir p a q em f [6]:

$$\tau_f(p, q) = \min\{\tau_f(P) \mid P \text{ une } p \text{ a } q\}$$

O tempo geodésico separando o ponto p e o conjunto, ou objeto, Y numa imagem em tons de cinza f é a menor quantidade de tempo para unir o ponto p a qualquer ponto y de Y [6]:

$$\tau_f(p, Y) = \min_{y \in Y} \tau_f(p, y)$$

Se p pertence a Y então o tempo geodésico de p para Y é zero. Associando cada ponto de uma imagem f com seu tempo geodésico para um objeto Y contido em D_f , pode ser definida a função tempo geodésico $T_{f,Y}$ [6].

Zona de influência baseada em geodésia genérica: considerando X como um conjunto definido no domínio D_f , de uma imagem em tons de cinza f composto de N componentes desconectados X_i , a zona de influência baseada na geodésia genérica de X_i contido em X é o conjunto dos pontos pertencentes ao domínio de f cujo tempo geodésico para X_i é menor que para qualquer outro componente conectado de X [6]:

$$IZ_f(X_i) = \{p \in D_f, \forall j \in [1, N] - \{i\}, \tau_f(p, X_i) < \tau_f(p, X_j)\}$$

As zonas de influência de X em f ($IZ_f(X)$) são definidas como a união das zonas de influência de todos os componentes conectados de X [6]. As fronteiras dessas zonas de influência correspondem ao esqueleto por zonas de influência de X em f .

3.14.1.2.3 – Watersheds em termos de geodésia

Na introdução dessa seção foi dito que as bacias de captação de uma imagem em tons de cinza podem ser consideradas como zonas de influência dos mínimos regionais da imagem e que as watersheds podem,

3.15 – Segmentação controlada por marcador

A idéia básica de uma segmentação controlada por marcador é transformar a imagem de tal forma que as watersheds dessa imagem correspondam a limites de um objeto semântico. Na prática uma realização direta da transformação watershed sobre a imagem produz uma sobre segmentação devido à presença de mínimos espúrios (figura 29). Assim, é necessária a filtragem da imagem antes da computação da transformação watershed para a retirada de todos os mínimos irrelevantes. Qualquer técnica de filtragem pode ser utilizada, mas o uso de uma imagem marcadora que assinale os objetos relevantes potencializará o resultado da transformação watershed.

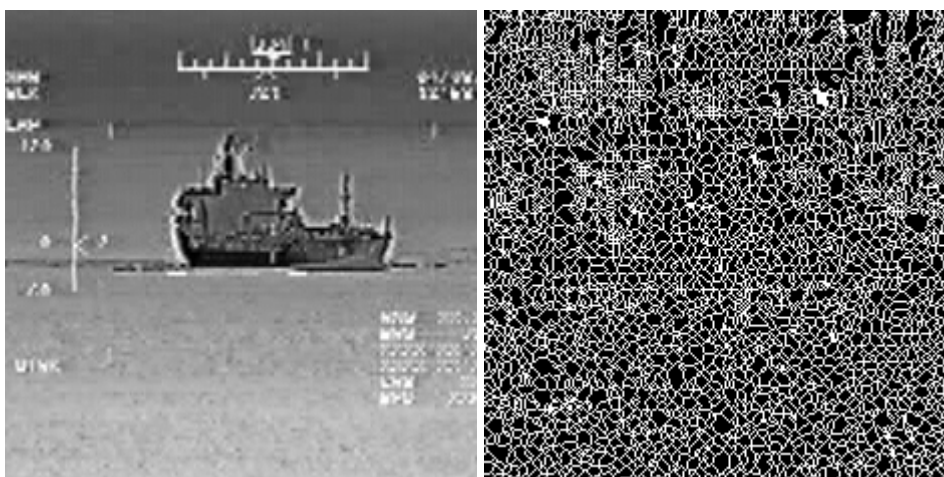


Figura 29: Imagem original e sobre-segmentação ocasionada pela aplicação da transformada watershed diretamente sobre a imagem original.

Esse tipo de abordagem também é conhecido como o paradigma de Beucher-Meyer. Este é o resultado da tentativa de se estabelecer uma metodologia para a segmentação morfológica. Essa metodologia consiste em dois passos: um passo “inteligente” cujos propósitos são a determinação do conjunto de marcadores que irão apontar os objetos a serem extraídos, e um passo que se pode chamar de “mecânico” que executa a transformação watershed a partir de marcadores, sobre a imagem simplificada.

O tamanho e a espessura do marcador também pode variar bastante. Por exemplo, marcadores maiores tem melhor desempenho em imagens ruidosas.

O processo para a escolha do marcador, ou seja, o passo “inteligente”, ganha especial relevância e existem várias propostas para sua execução[19] [20].

No trabalho que é aqui apresentado utilizamos a transformada wavelet módulo máximo (ver capítulos 4 e 5).

3.15.1 – Escolha do marcador

A escolha do marcador depende do que se está procurando segmentar na imagem. É possível, para uma mesma imagem, ter-se um conjunto de marcadores muito diverso.

Em linhas gerais, a imagem de marcação é uma imagem em dois estados, contendo marcadores dos objetos e é definida como:

$$f_m(x) = \begin{cases} 0 & \text{se } x \text{ pertence a um marcador} \\ t_{\max} & \text{caso contrário} \end{cases}$$

Em nosso caso foi convencionado que para as imagens marcadoras para a transformação watershed o valor zero corresponde a cor branca e o valor um a cor preta.

O marcador é aplicado à imagem que foi preparada (f_1) a partir da imagem original (f_0) com o objetivo de se executar a transformação watershed. Essa preparação compreende algum tipo de filtragem realizada na imagem antes da computação da transformação watershed. Esta pode ser composta também de um conjunto de filtros com o objetivo de retirar espúrios ou aumentar a definição das bordas. A imagem f_m é, então, modificada pela técnica de imposição de mínimo resultando em uma nova imagem f_2 :

$$f_2 = R_{(f_1 \wedge f_m)}^*(f_m)$$

A imposição de mínimo é executada em dois passos. Primeiro uma operação de ínfimo é realizado pixel a pixel entre a imagem de entrada e a imagem marcadora ($f_1 \wedge f_m$). Assim, mínimos são criados, onde não existiam, nas posições

correspondentes aos marcadores e com isso tem-se a certeza que a imagem resultante (f_2) é menor ou igual à imagem marcadora. O segundo passo consiste em uma reconstrução morfológica por erosão (R^* - vide seção 3.12.3.2) de $f_1 \wedge f_m$ a partir da imagem f_m . Na figura 30 é possível notar que os únicos mínimos locais da imagem resultante são os impostos pelo marcador.



Figura 30: Imposição de mínimo. Da esquerda para a direita imagem de entrada, imagem marcadora e imposição de mínimo.

Só então é aplicada a transformação watershed:

$$Seg(f) = W_s(f_2)$$

Com isso é possível evitar a sobre-segmentação, pois a transformação watershed é executada a partir do limite imposto pelo marcador.

4 –SEGMENTAÇÃO UTILIZANDO A TRANSFORMAÇÃO WATERSHED

4.1 – Introdução

A segmentação de imagens pode ser considerada como um processo de baixo-nível no qual a análise de imagens está fundamentada. Dois processos relacionados estão envolvidos:

- 1) A segmentação de imagens em regiões que são homogêneas com respeito a uma dada característica da imagem; por exemplo, segmentação em regiões de cinza uniforme.
- 2) A geração de uma imagem de bordas e a extração das bordas mais importantes.

Os métodos de segmentação se concentram tipicamente em uma ou outra dessas duas abordagens.

Métodos baseados em regiões são dependentes de elementos da estatística da imagem como média e variância. O método de crescimento de regiões, por exemplo, é dependente não só da escolha dos pontos iniciais de segmentação, como também do algoritmo de crescimento utilizado. Regiões distintas podem ser juntadas porque a mudança de característica dos pixels fronteiros pode ser subjugada pelas características gerais das duas regiões.

Métodos baseados em detecção de bordas representam um compromisso entre simplicidade e estabilidade por conta da possibilidade de ruído e interferência. Além disso técnicas de segmentação baseadas em bordas frequentemente têm dificuldade em estabelecer a conectividade dos segmentos de borda. Podem, então, ocorrer discontinuidades devido ao ruído ou trocas de direção das bordas.

Essas duas abordagens da segmentação de imagens, por região e por bordas, podem ser consideradas como complementares: a primeira depende das variações locais de intensidade e a segunda de parâmetros das regiões para que seus resultados ganhem

em precisão. Vários métodos já unificam essas duas abordagens como em [21] [22] e a transformação watershed é um deles.

4.2 – Uso da transformação watershed em imagens infravermelhas

Uma imagem em infravermelho, como já foi visto no capítulo 2, apresenta características importantes para detecção localizada nas bordas dos objetos semânticos. Entretanto, se considerarmos as regiões pertencentes a um objeto semântico em uma imagem infravermelha, pode ser visto que essas estão relacionadas à regiões de pouca variação interna e possibilidade de muita variação entre si. Por exemplo, ao considerar um carro de combate de perfil como o da figura 31 que tenha andado algum tempo temos pelo menos três regiões de destaque: as rodas mais quentes devido ao atrito com o chão, a região do motor e a carcaça propriamente dita, que inclusive pode estar a uma temperatura inferior a do ambiente. Além disso, devido a características inerentes ao processo de aquisição das imagens infravermelhas estas estão sujeitas a ruído, principalmente de alta frequência.

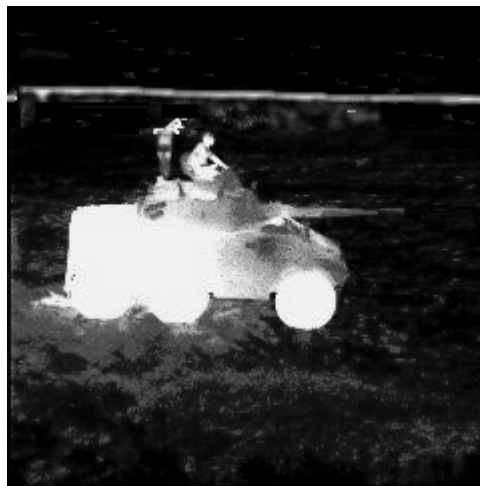


Figura 31: Imagem em infravermelho com destaque para , pelo menos, três regiões uniformes de cinza internas ao objeto semântico.

É comum se associar a segmentação de imagens em infravermelho com técnicas de limiarização [3] [4]. No entanto, como já foi observado no capítulo 2, essa técnica possui o inconveniente de contar com uma estatística da imagem bem

comportada, onde o cálculo ou a extração de um único ou de múltiplos limiares é relativamente simples. Porém isso não é verdade quando se possui uma imagem impregnada de ruído ou com o objeto semântico com múltiplas regiões de temperatura, como o da figura 31, onde seria difícil separar as imagens do carro de combate da região da temperatura ambiente. Além disso, as técnicas de detecção associadas a limiarização só permitem a detecção de um objeto por vez.

Assim, o estudo de um método baseado na transformação watershed para a segmentação de imagens em infravermelho é justificável não só pelas características descritas, que indicam a necessidade da utilização de técnicas baseadas nas bordas combinadas com as baseadas em região, como também pela escassez de trabalhos com esse tipo de abordagem.

4.3 – Considerações práticas com o paradigma de Beucher-Meyer como fio condutor

Na seção 3.15 vimos o efeito da transformação watershed quando aplicado diretamente na imagem. A sobre segmentação inerente à transformada watershed é um problema que encontra no paradigma de Beucher-Meyer [10] [11] uma solução. Como Beucher relaciona [10], o processo de determinação dos marcadores utiliza técnicas que podem diferir de acordo com o tipo de imagem a ser analisado. É aí que reside o grande desafio dessa técnica morfológica: estabelecer um método eficiente para a extração dos marcadores.

Em nosso caso, a primeira tentativa de se estabelecer uma segmentação eficiente para as imagens infravermelhas, foi o de utilizar, simplesmente, marcadores baseados no gradiente morfológico e aplicar a transformada watershed diretamente sobre a imagem, como pode ser visto no diagrama da figura 32.

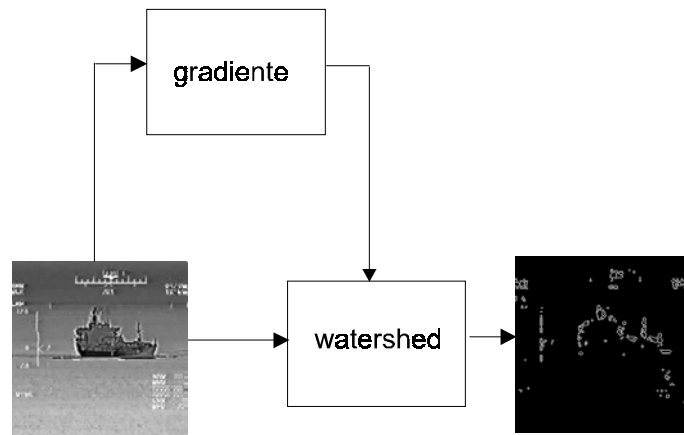


Figura 32: Diagrama em blocos da primeira tentativa para a segmentação de imagens infravermelhas.

Esse tipo de abordagem é muito simples e, infelizmente, não produziu um bom resultado. Na figura 33 pode ser observado que sua aplicação não extraiu um conjunto de bordas fechadas para o objeto semântico, e foram detectados muitos alvos falsos, por conta do ruído da imagem. De fato, uma das características desse tipo de abordagem é a de que o gradient4 0 s8ça todosos component4s de alta freqüência da imagem e, assim, component4s de ruído de alta freqüência são selecionados pelo marcador e resultam em alvos falsos.

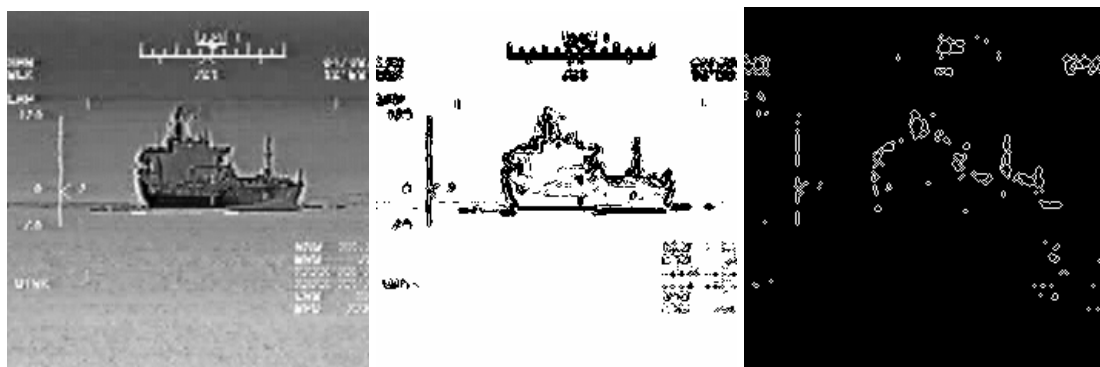


Figura 33: Segmentação morfo8ógica com o gradient4 como marcador. Imagem original, marcador baseado no gradiente e transformada watershed sobre imagem original.

Observando métodos semi-automáticos baseados na transformação watershed [19] [23], partiu-se para uma abordagem semi-automática, na tentativa de

isolar os dois passos propostos pela técnica desenvolvida por Beucher e Meyer. Isto é, trata-se um passo de cada vez, encontrando-se a técnica mais adequada para cada um. Como em [19] os contornos dos objetos semânticos foram determinados manualmente (veja a figura 34), através de um editor de imagens.

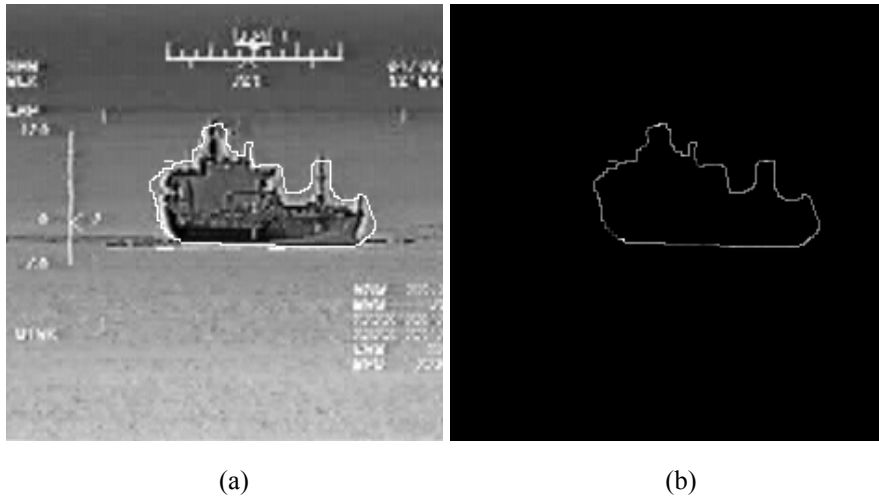


Figura 34: Contorno da imagem determinado manualmente. Utilizando um editor de imagens é contornado o objeto semântico da imagem (a). Depois essa imagem é subtraída da original para obtenção do mapa de bordas (b).

Com o mapa de bordas obtido é realizada uma dilatação. Essa dilatação tem o objetivo de aumentar a área em que a watershed será realizada de fato (ver seção 3.15), garantindo que as bordas do objeto semântico estarão contidas no conjunto de marcadores para a transformação watershed. O elemento estruturante dessa dilatação foi objeto de estudos e testes. Foram utilizados seis tipos de elementos estruturantes: cinco foram quadrados digitais de dimensões 10X10, 8X8, 6X6, 4X4 e 2X2; o outro foi a cruz elementar. Esses tamanhos estabelecem a espessura final do marcador para a transformação watershed.

Um marcador válido deve ser constituído de uma região de pixels pretos com, pelo menos, um pixel branco em seu interior. Se o elemento estruturante da dilatação for muito espesso, objetos semânticos muito finos ou pequenos, ou parte deles, não são segmentados pela transformação watershed como pode ser visto na figura 35, devido ao fato de, por conta de sua espessura, a dilatação não produzir marcadores

válidos. Note que no primeiro navio (figura 35 (a)), por ser uma estrutura fina, a utilização do elemento estruturante composto por um quadrado de dimensões 10X10 pixels conduziu a não detecção pela transformação watershed. No segundo navio (figura 35 (d)) um dos mastros do navio não é detectado.

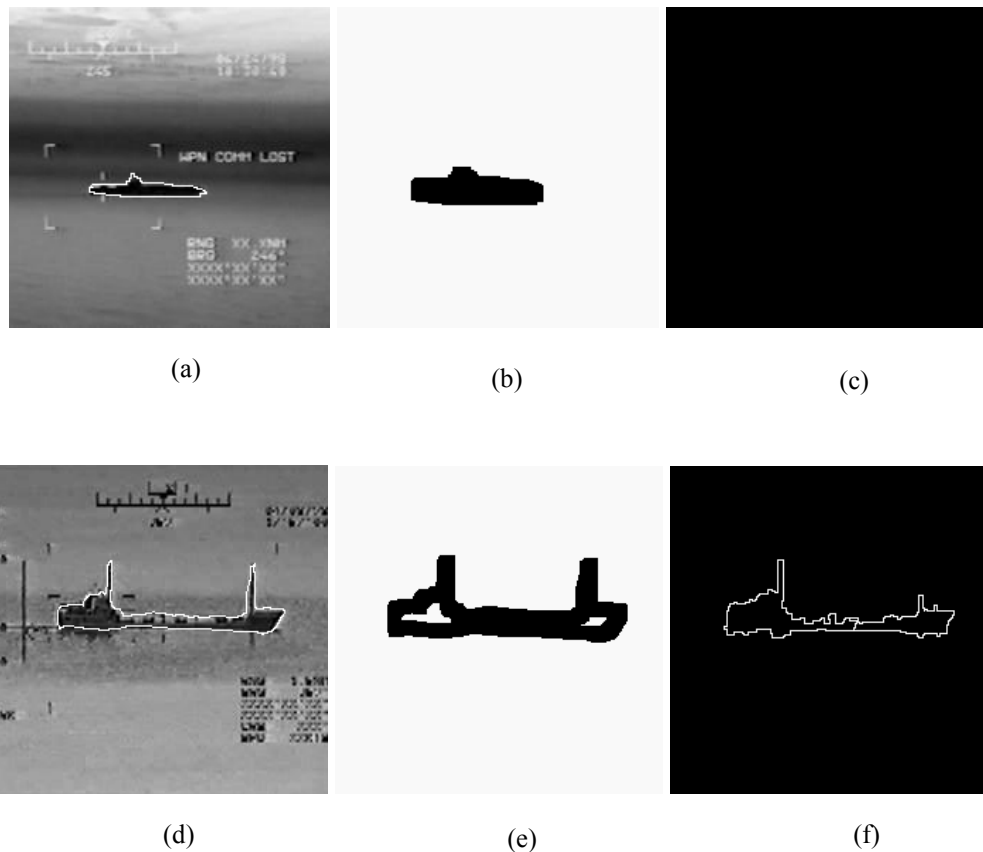


Figura 35: Teste dos elementos estruturantes que dilatam o conjunto de bordas escolhidas. Imagem original com bordas escolhidas manualmente (a e d), marcadores realizados a partir da dilatação das bordas escolhidas manualmente, a partir da dilatação por um elemento estruturante quadrado de 10X10 pixels (b e e) e watersheds extraídas (c e f)

Elementos estruturantes muito finos podem conduzir a dois tipos de problemas: se a imagem estiver muito contaminada de ruído, uma extração automática dos marcadores pode conduzir a muitos alvos falsos por conta do provável número elevado de marcadores válidos; por outro lado, por conta de pequena espessura dos marcadores, se o conjunto de bordas, extraídos automaticamente ou não, não for

preciso, existe a tendência da transformação watershed acompanhar a imagem oferecida como marcador que pode conter erro (vide figura 36).

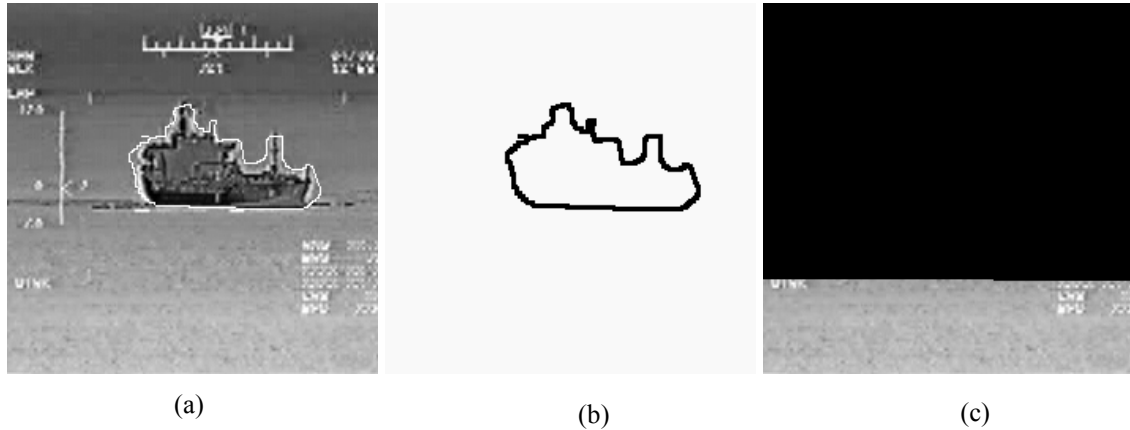


Figura 36: Teste dos elementos estruturantes que dilatam o conjunto de bordas escolhidas. Imagem original com bordas escolhidas manualmente (a) marcadores realizados a partir da dilatação das bordas escolhidas manualmente, a partir da dilatação por um elemento estruturante quadrado de 2X2 pixels (b) e watersheds extraídas sobreposta a imagem original (c)

A melhor relação custo benefício, em termos de tamanho do elemento estruturante para dilatação do conjunto de bordas extraída para ser marcador na computação da transformação watershed foi obtida com o elemento estruturante quadrado, binário de tamanho 6X6 pixels (figura 37). É importante observar que, ao longo do processo de determinação dos passos do método, essa escolha foi testada e confirmada algumas vezes.



Figura 37: Elemento estruturante escolhido para dilatar o conjunto de bordas escolhidas. Marcador realizado a partir da dilatação das bordas escolhidas manualmente, a partir da dilatação por um elemento estruturante quadrado de 6X6 pixels (esquerda) e watersheds extraídas sobreposta a imagem original (direita)

Com a determinação desse tamanho de elemento estruturante foi possível isolar o segundo passo para testes. Algumas abordagens de simplificação para a execução da transformada watershed foram testadas no conjunto de imagens de testes: limiarização em oito níveis, filtragem passa-baixas e gradiente morfológico.

A limiarização em oito níveis consiste na análise pixel a pixel da imagem e sua aproximação para o menor valor dos seguintes intervalos: $[0,32[$, $[32, 64[$, $[64,96[$, $[96,128[$, $[128,160[$, $[160,192[$, $[192,224[$, $[224,256[$. Essa limiarização em oito níveis resultou em perda de bordas importantes visto que patamares próximos, mas que deveriam ser diferenciados foram fundidos num só. Além disso dois patamares que estavam próximos, porém em intervalos diferentes, podem ser afastados demais, se transformando em uma borda falsa que poderia ser detectada. Note que na figura 38 o resultado da limiarização conduz à formação de uma região entre o céu e o mar menos suave que na imagem original. Além disso, o resultado da segmentação, que utilizou o marcador obtido manualmente da figura 35(d), indica que dois limiares próximos que deveriam ser diferenciados, não foram, visto que os mastros do navio não foram segmentados.



Figura 38: Teste da limiarização em oito níveis como simplificação para a imagem que será submetida à transformação watershed. Da esquerda para direita: imagem original, imagem limiarizada em oito níveis e resultado da transformação watershed.

A filtragem passa-baixas de uma imagem produz um efeito de suavização dessa imagem. Com isso diminui o ruído de alta frequência que esteja contido nessa

imagem. No entanto, o efeito de suavização pode produzir o mesmo efeito de fusão de limiares próximos que deviam ser diferenciados. Isso pode ser observado no exemplo da figura 39 em que um conjunto de marcadores não muito preciso levou a segmentação final a identificar o reflexo do navio como parte deste, e a seguir o marcador (utilizar o marcador como borda real do objeto) não identificando o mastro da proa. Além disso a filtragem passa-baixas pode conduzir a um deslocamento das bordas e com isso a segmentação final também apresentar pequenas deslocações.

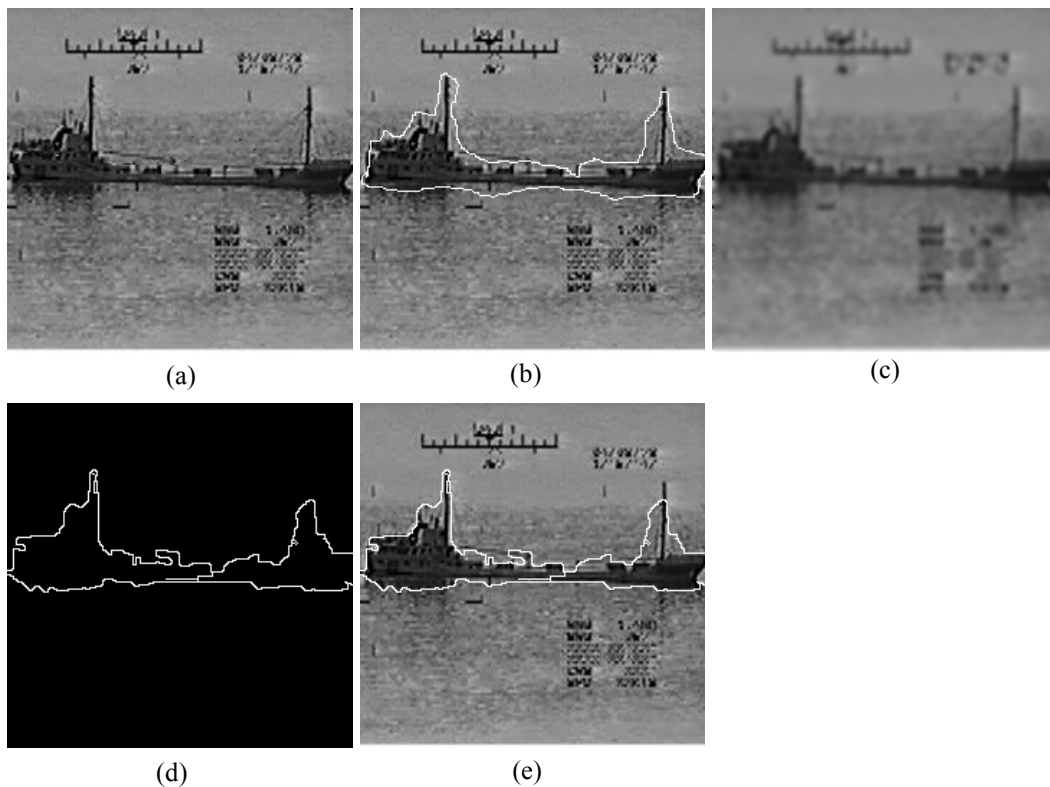


Figura 39: Teste da filtragem passa-baixas como simplificação para a imagem que será submetida à transformação watershed. Da esquerda para direita e de cima para baixo: imagem original (a), marcador estabelecido manualmente (b), imagem filtrada (c), resultado da transformação watershed (d) e resultado da transformação watershed sobreposto a imagem original (e).

Apesar de sua característica de ressaltar as bordas, o que também ressalta os ruídos de alta frequência que estejam contaminando a imagem, de todos os métodos de simplificação da imagem testados, o que apresentou os melhores resultados, para a execução da transformação watershed com um marcador preciso, foi o gradiente

morfológico. Foram testados dois tipos de gradiente morfológico, o gradiente normal e o gradiente espesso (seção 3.3), e, como pode ser visto na figura 40, o melhor resultado para a simplificação da imagem foi o associado ao gradiente espesso. Na comparação entre os dois resultados para a mesma imagem (figuras 40 (c) e (e)) aquela em que a transformação watershed foi realizada a partir do gradiente espesso teve um melhor resultado. Note que em algumas regiões da borda a segmentação final acompanhou o marcador extraído manualmente. No exemplo da figura 41 a segmentação final realizada a partir do gradiente normal não obteve êxito em extrair o objeto semântico.

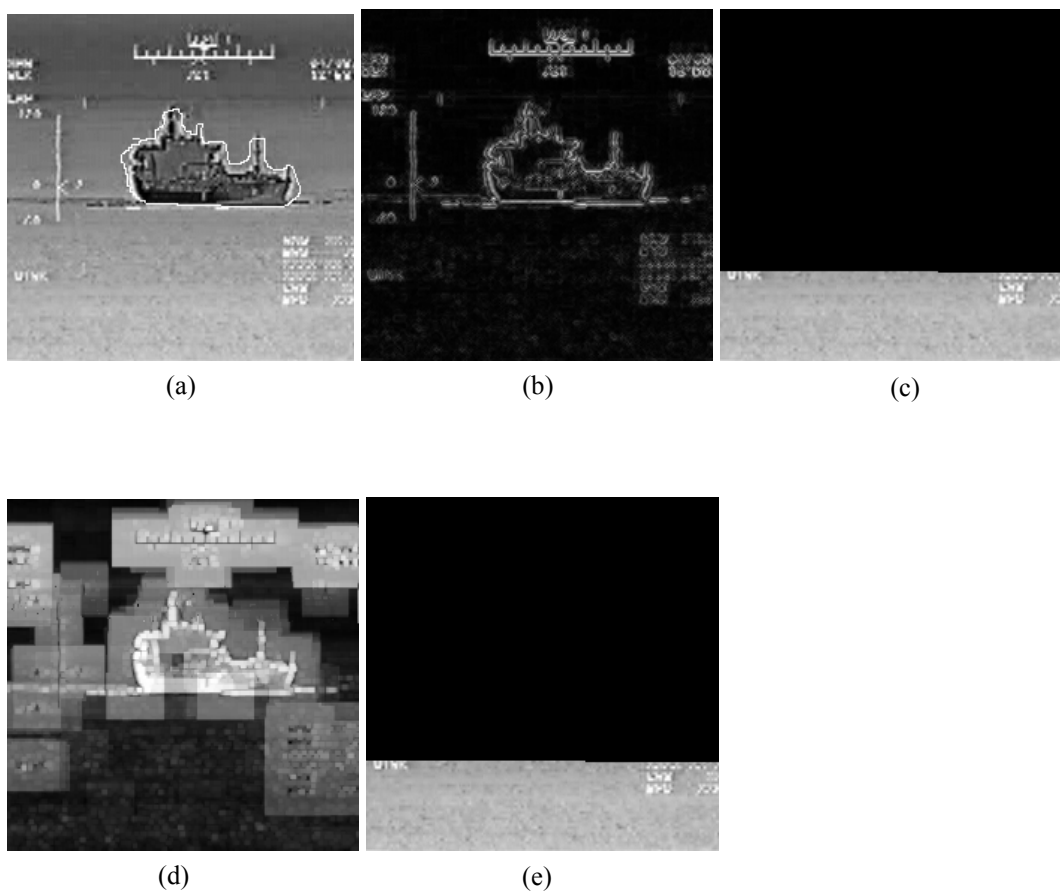


Figura 40: Teste do gradiente morfológico como simplificação para a imagem que será submetida à transformação watershed. Da esquerda para direita e de cima para baixo: marcador obtido manualmente sobreposto à imagem original (a), gradiente morfológico normal (b), resultado da transformação watershed utilizando o gradiente normal sobreposto a imagem original (c), gradiente morfológico espesso (d) e resultado da transformação watershed utilizando o gradiente espesso sobreposto a imagem original (e).

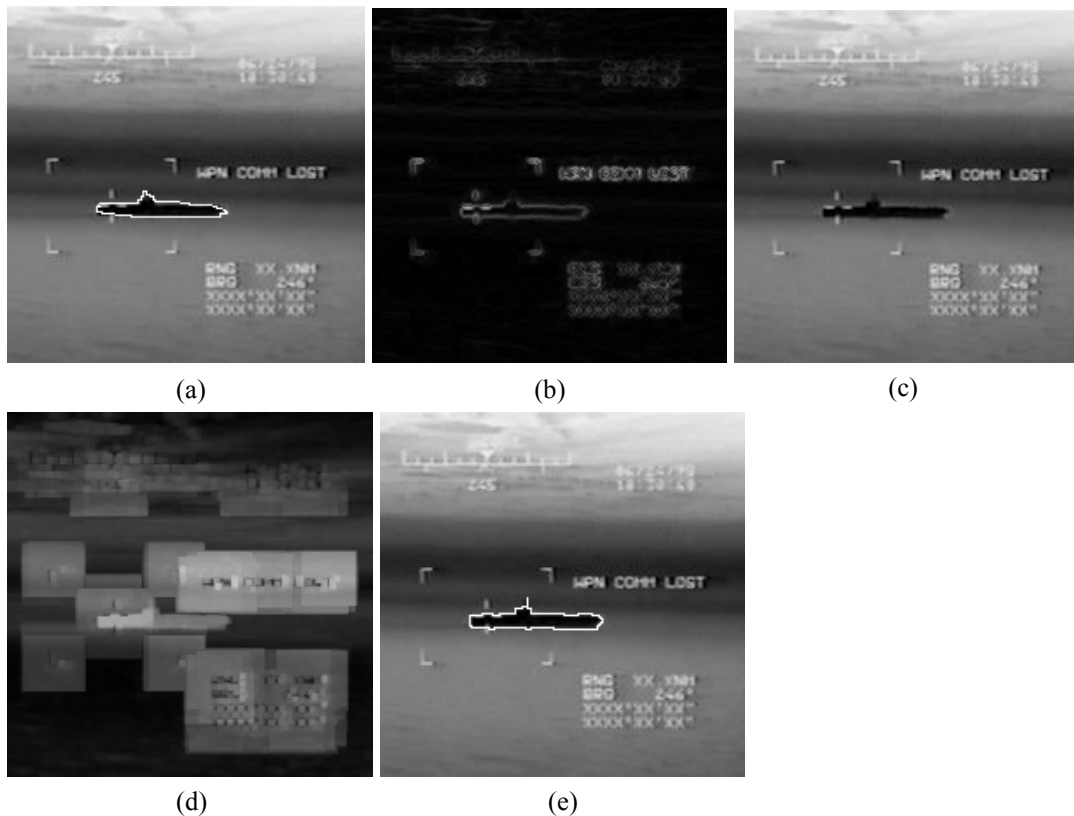


Figura 41: Teste do gradiente morfológico como simplificação para a imagem que será submetida à transformação watershed. Da esquerda para direita: marcador obtido manualmente sobreposto à imagem original (a), gradiente morfológico normal (b), resultado da transformação watershed utilizando o gradiente normal sobreposto a imagem original (c), gradiente morfológico espesso (d) e resultado da transformação watershed utilizando o gradiente espesso sobreposto a imagem original (e).

Convém observar que os testes para essas opções de simplificação foram realizados para todo o conjunto de imagens de testes e foram analisados de acordo com a melhor relação possível de bons resultados. Existiram resultados iguais para os vários métodos ou até alguns em que foi superado o método de simplificação (gradiente espesso) escolhido.

Outra conclusão que pode ser tirada dos testes realizados até aqui, é o da tremenda importância de se ter um método de extração do conjunto de marcadores para os objetos semânticos da imagem o mais preciso possível.

4.4 – A escolha do método para extração do marcador

A escolha da técnica para extração do conjunto de marcadores é, portanto, o problema central a ser resolvido para o estabelecimento de um método de segmentação para imagens infravermelhas. Este deve ter as seguintes características:

- Deve ser automático – considerando a utilização em classificação e detecção de alvos é essencial que a extração de alvos seja automática.
- Deve ter imunidade ao ruído ou permitir um tratamento diferenciado deste – devido ao ruído inerente ao processo de aquisição da imagem em infravermelho é necessário o acesso a uma técnica que permita a eliminação de ruído por técnicas lineares ou não-lineares.
- Deve fornecer marcadores conectados – a transformação watershed precisa de marcadores conectados para executar uma correta extração do objeto semântico.

Conforme foi visto na seção 4.3 a exatidão desse conjunto de marcadores é essencial. Em muitos casos (vide figuras 36 e 38) a transformada watershed tem a tendência, quando a imposição de mínimo não inclui a região de borda de interesse, de seguir o resultado estabelecido para o marcador.

Por conta desses requisitos a técnica escolhida para execução do passo de detecção e extração dos marcadores da imagem foi a transformada wavelet módulo máximo que será vista a seguir.

5 - TRANSFORMADA WAVELET

MÓDULO MÁXIMO

5.1 - Introdução

A teoria que envolve a transformada wavelet fornece uma abordagem unificada para várias técnicas que foram desenvolvidas independentemente para aplicações em processamento de sinais [24]. Por exemplo, o processamento de sinais em multiresolução usado em visão artificial; codificação em subbandas, utilizada em compressão de imagens e voz e expansão de séries wavelet desenvolvida em matemática aplicada, podem ser reconhecidas como diferentes visões de uma mesma teoria.

A transformada wavelet pode ser vista como a decomposição de um sinal por um conjunto de funções baseadas na dilatação e compressão de uma única “wavelet mãe” ($\psi(t)$) [25] [26]: $\left\{ \psi\left(\frac{t-b}{a}\right), (a,b) \in R^+ \times R \right\}$ normalmente $a = 2^{-j}$ e $b = k2^j$ onde k e j inteiros. Na prática sua realização é bem simples e se utiliza de filtros em quadratura, que vão decompor o sinal ou imagem [25] [26].

A transformada wavelet representa uma alternativa à transformada de Fourier e à transformada de Fourier de tempo curto para a análise de sinais. A transformada wavelet estabelece a noção de escala como alternativa à frequência conduzindo a uma representação tempo-escala [24] ao invés da representação tempo-frequência da transformada de Fourier. Além disso ela introduz o conceito de multiresolução em que, basicamente, pode se considerar [27] que é possível obter o sinal ou imagem em um nível superior da decomposição através da soma do sinal suavizado somado aos detalhes (ou componentes wavelets) do nível anterior.

Existem diversos tipos de transformada wavelet (contínua, discreta, binária...) que podem ser escolhidas dependendo da aplicação. Aplicações na área de processamento de sinais e imagens foram trazidas a foco depois dos artigos de

Daubechies [28] e Mallat [25] [26], e, desde então, a transformada wavelet tem sido explorada em diversos tópicos dessas áreas.

Diversas aplicações para a transformada wavelet podem ser citadas. Em [27] são destacados: o sistema de digitalização das impressões digitais do FBI, as aplicações em compressão de imagem, vídeo, áudio e voz, e em sistemas de filtragem de ruídos, entre outros.

Em nosso caso, no processo para uma segmentação eficiente utilizando a transformação watershed, é necessário a obtenção de uma imagem marcadora para evitar a sobre-segmentação inerente ao seu processamento como foi visto no capítulo 4. Esse processo deve ser automático, deve localizar os objetos de importância que devem ser mantidos conectados, com precisão, além de ser capaz de tirar elementos espúrios, tais como ruídos (veja a seção 4.4). O detetor escolhido por esse trabalho foi a transformada wavelet módulo máximo que possui características não só de detecção como também de filtragem que podem ser exploradas nesse sentido.

A transformada wavelet módulo máximo pode ser considerada como um detetor de borda multiescalar [12] [13] [29]. A maioria dos detetores de borda multiescalares suaviza o sinal em várias escalas e detecta variações bruscas através de suas derivadas de primeira ou segunda ordem. Para uma classe particular de wavelets, as bordas podem ser detectadas a partir do máximo local da transformada wavelet módulo máximo. Essa característica tem sido explorada com sucesso não só para detetores de singularidades de sinais ou imagens [12] [30], como também em métodos de compressão de imagens [31] [32].

5.2 - Definição da Transformada Wavelet

A partir da função mãe ($\psi(t)$) a transformada wavelet para uma função f pode ser definida formalmente por [25] [26]:

$$Wf(s, u) = \int_{-\infty}^{\infty} f(x) \sqrt{s} \psi(s(x - u)) dx$$

E, se a dilatação de $\psi(x)$ por um fator s for definida por:

$$\psi_s(x) = \sqrt{s} \psi(sx)$$

Pode-se rescrever a wavelet como um produto interno em $L^2(R)$ - espaço de Hilbert das funções $f(x)$ mensuráveis e quadrado-integráveis [25] [26]:

$$Wf(x) = \langle f(x), \psi_s(x-u) \rangle$$

O que corresponde a decomposição de $f(x)$ pela família de funções $(\psi_s(x-u))_{(s,u) \in R^2}$.

Para a reconstrução da função $f(x)$ a partir de sua transformada wavelet, a transformada de Fourier de $\psi(x)$ ($\hat{\psi}(w)$) deve satisfazer algumas condições [25] [26]:

$$C_\psi = \int_0^\infty \frac{|\hat{\psi}(w)|^2}{w} dw < +\infty$$

O que implica que $\psi(w)$ é pequeno nas vizinhanças de $w=0$ e é zero quando $w=0$. Assim, a função $\psi(x)$ pode ser interpretada como a resposta ao impulso de um filtro passa-banda. Se considerarmos $\tilde{\psi}_s(x) = \psi_s(-x)$ então pode se rescrever a transformada wavelet em um ponto u e na escala s como [25] [26]:

$$Wf(s,u) = f * \tilde{\psi}_s(u)$$

Assim a transformada wavelet pode ser vista como a filtragem de $f(x)$ através de um filtro passa-banda cuja resposta ao impulso é $\psi_s(x)$

A expressão para a reconstrução de $f(x)$ a partir de $Wf(s,u)$ é a seguinte:

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_0^{+\infty} Wf(s,u) \psi_s(x-u) ds du$$

Para o caso de duas dimensões, que inclui imagens, a transformada wavelet pode ser rescrita da seguinte maneira [25] [26]:

$$Wf(s,(u,v)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) s \Psi(s(x-u), s(y-v)) dx dy$$

Considerando, da mesma maneira que no caso unidimensional, que $\Psi(x, y) \in L^2(R^2)$ é uma função cuja transformada de Fourier ($\hat{\Psi}(w_x, w_y)$) satisfaz [25] [26]:

$$\forall(w_x, w_y) \in R^2 \int_0^{+\infty} \frac{|\hat{\Psi}(sw_x, sw_y)|^2}{s} ds = C_\Psi < +\infty$$

Então, se $\Psi_s(x, y) = s\Psi(sx, sy)$ e $\tilde{\Psi}_s(x, y) = \Psi_s(-x, -y)$:

$$Wf(s, (u, v)) = f * \tilde{\Psi}(u, v)$$

Que pode ser compreendido como a filtragem de $f(x, y)$ através de um filtro passa-banda bidimensional.

A expressão para reconstrução de $f(x, y)$ é [25] [26]:

$$f(x, y) = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_0^{+\infty} Wf(s, (u, v)) \Psi_s(x - u, y - v) ds du dv$$

5.2.1 - Análise por multiresolução

Outra abordagem de wavelets com vista a sua utilização em processamento de imagens, é a análise por multiresolução desenvolvida por Mallat [25].

Considerando o espaço L^2 de todas as funções quadrado integráveis, a análise por multiresolução de Mallat pode ser compreendida como uma seqüência crescente de subespaços próximos que estão contidos em $L^2(R)$.

A partir da escolha de uma função de escalamento ϕ que satisfaça certas necessidades de continuidade e suavidade, e que, mais importante, permita que a família, $\phi\{(x-k), k \in Z\}$ forme uma base ortogonal para o espaço de referência V_0 , as seguintes relações descrevem a análise.

$$1) \dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots$$

Os espaços V_j são encadeados. A união de todos os V_j está contida no espaço $L^2(R)$.

$$2) f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}, j \in Z$$

Os espaços V_j e V_{j+1} são similares. Se o espaço V_j é atravessado por $\phi_{j,k}(x), k \in Z$ então o espaço V_{j+1} é atravessado por $\phi_{j+1,k}(x), k \in Z$. O espaço V_{j+1} é gerado pelas funções $\phi_{j+1,k}(x) = \sqrt{2}\phi_{j,k}(2x)$.

Para considerar o uso deste tipo de análise aplicada em imagens, já que $V_0 \subset V_1$, tem-se que qualquer função em V_0 pode ser descrita como uma combinação linear de funções base $\sqrt{2}\phi(2x - k)$ de V_1 . Em particular [25]:

$$\phi(x) = \sum_k g(k)\sqrt{2}\phi(2x - k)$$

Coefficientes $g(k)$ são definidos, em termos de produto interno, como $\langle \phi(x), \sqrt{2}\phi(2x - k) \rangle$. Considerando, agora, o complemento ortogonal W_j de V_j até V_{j+1}

(isto é $V_{j+1} = V_j \oplus W_j$). Pode-se definir [25]:

$$\psi(x) = \sqrt{2}\sum_k (-1)^k g(-k + 1)\phi(2x - k)$$

Pode ser mostrado que $\{\sqrt{2}\psi(2x - k), k \in Z\}$ é uma base ortonormal para W_1 .

Recorrendo-se a propriedade de similaridade, já descrita, é possível afirmar que $\{2^{j/2}\psi(2^j x - k), j, k \in Z\}$ é uma base para W_j . Uma vez que, tal como ocorre com V_j , a união de todos os W_j está contida em $L^2(R)$, a família $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), j, k \in Z\}$ é uma base para $L^2(R)$.

Para uma função $f \in L^2(R)$ pode-se encontrar N tal que $f_N \in V_N$ aproxime f até uma precisão preestabelecida. Se $g_i \in W_i$ e $f_i \in V_i$ então :

$$f_N = f_{N-1} + g_{N-1} = \sum_{i+1}^M g_{N-M} + f_{N-M}$$

Em outras palavras é possível obter uma imagem em uma resolução (ou escala) maior a partir da soma da mesma imagem em uma escala menor com os detalhes dessa imagem nessa escala menor.

A representação por multiresolução fornece uma estrutura simples para a análise da informação contida na imagem. Se, em uma resolução inferior somos capazes de ter uma visão geral, quanto mais aumentamos essa resolução somos capazes de ver mais detalhes. Pode ser compreendido, então, que os detalhes da imagem em uma resolução inferior contém o “contexto” da imagem enquanto detalhes de maior qualidade correspondem a “particularidades” dessa imagem. Por exemplo, é difícil entender que um retângulo na imagem é uma janela se não estiver estabelecido que o contexto da imagem é uma casa. Assim é natural iniciar a análise de uma imagem do ponto de vista de uma resolução inferior até uma superior.

5.2.2 - Filtros em quadratura

A análise por multiresolução desenvolvida por Mallat [25] [26] permite também um outro tipo de interpretação, baseada em filtros que vão realizar a transformada wavelet.

Já foi estabelecido que:

$$\phi(x) = \sum_k g(k)\sqrt{2}\phi(2x - k)$$

e

$$\psi(x) = \sqrt{2}\sum_k h(k)\sqrt{2}\phi(2x - k), \text{ onde } k \in Z.$$

As seqüências $\{h(k), k \in Z\}$ e $\{g(k), k \in Z\}$ são os chamados filtros espelhados em quadratura e sua relação é dada por [25] [26]:

$$h(n) = (-1)^n g(1 - n)$$

A seqüência $g(k)$ representa um filtro passa baixas, ou de escalamento, e a $h(k)$ um filtro passa altas ou wavelet. Usando a transformada de Fourier e a ortogonalidade pode-se provar que $\sum g(k) = 2$ e $\sum h(k) = 0$.

A maneira mais simples e rápida de executar a transformada wavelet em um dado sinal ou imagem é a de se utilizar desses filtros em quadratura como operadores.

Para se aplicar a decomposição via wavelet em sinais multidimensionais (como imagens...), extensões multidimensionais de wavelets são necessárias. Uma

maneira de se fazer isto é a de se usar “wavelets separáveis” obtidas através do produto de filtros wavelet de uma dimensão e funções de escalamento.

No caso de duas dimensões, considerando a função de escalamento (ou passa baixas) como $g_c(x)$ e a wavelet (ou passa altas) $h_c(x)$, pode-se construir funções bidimensionais:

$$\begin{aligned} g_c(x,y) &= g_c(x).g_c(y) \\ h_c^{(1)}(x,y) &= g_c(x).h_c(y) \\ h_c^{(2)}(x,y) &= h_c(x).g_c(y) \\ h_c^{(3)}(x,y) &= h_c(x).h_c(y) \end{aligned}$$

Estas funções ou filtros são ortogonais cada um em relação ao outro com respeito a deslocamentos inteiros. A função $g_c(x,y)$ é a função de escalamento (que corresponde a um filtro passa baixas bidimensional), enquanto que as funções $h_c^{(i)}(x,y)$ são wavelets.

O conjunto $\{h_c^{(i)}(2^i x - k, 2^j y - l), i = 1,2,3; j,k,l \in Z\}$ forma uma base ortonormal para funções quadrado integráveis em R^2 . Essa solução corresponde a um banco de filtros bidimensional, separável com subamostragem de 2 em cada dimensão.

A partir da filtragem utilizando esses filtros obtém-se as imagens $A_{2^i}^d f$ (imagem principal, ponto de partida das decomposições em cada nível) e $D_{2^i}^1 f, D_{2^i}^2 f$ e $D_{2^i}^3 f$ (imagens detalhe) de cada nível de decomposição (vide figura 42). Essas imagens em decomposição podem ser arrumadas de forma a compor uma espécie de pirâmide Laplaciana conforme a figura 43.

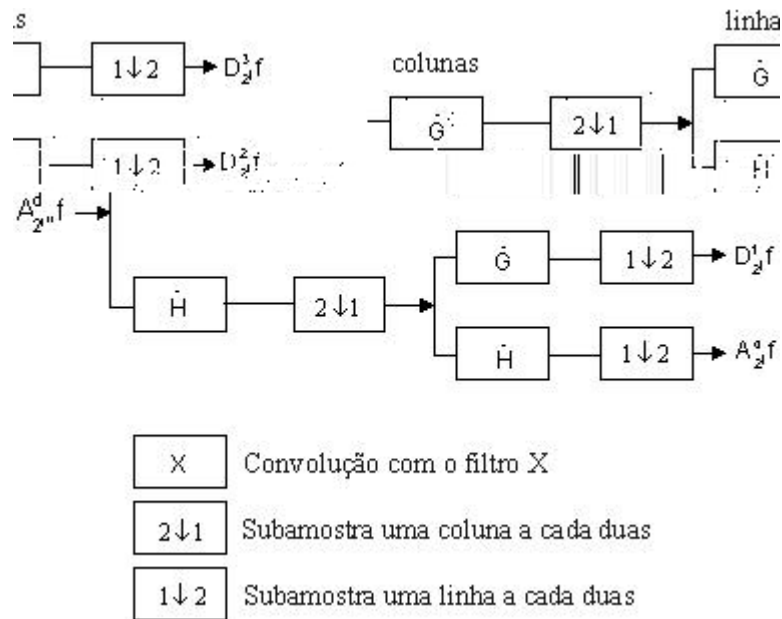


Figura 42: Algoritmo para decomposição da imagem em componente da transformada wavelet.

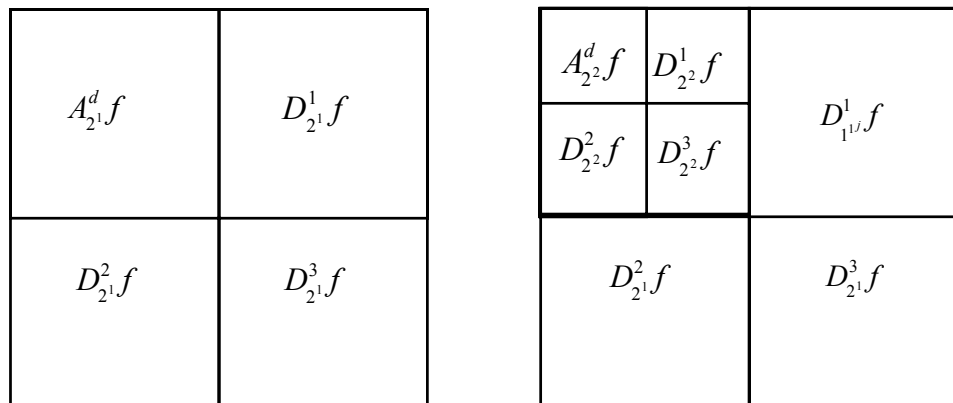


Figura 43: Representação wavelet tradicional para uma (a esquerda) e duas (a direita) camadas ou níveis.

Esse tipo de decomposição pode ser interpretada como uma decomposição de imagens num conjunto de subbandas de frequência independentes e orientadas espacialmente. A imagem $A_{2^j}^d f$ corresponde às baixas frequências em ambas as dimensões (filtragem por $g_c(x,y)$), $D_{2^j}^1 f$ às frequências altas na vertical (filtragem por $h_c^1(x,y)$), $D_{2^j}^2 f$ às frequências altas na horizontal (filtragem por $h_c^2(x,y)$), e $D_{2^j}^3 f$ às frequências altas na horizontal e na vertical (filtragem por $h_c^3(x,y)$).

5.2 – Transformada Wavelet Binária

Na transformada wavelet binária, a função que realiza a “suavização” do sinal ($\theta(x,y)$) é qualquer função em que a sua integral dupla (ao longo de x e y) é igual a 1 e que converge para 0 no infinito. Isso corresponde a resposta ao impulso de um filtro passa baixas. Se forem definidas duas funções que serão, respectivamente, as derivadas parciais de $\theta(x,y)$ ao longo de x e y :

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{e} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y}$$

Essas funções podem ser chamadas de wavelets. Se cada uma dessas funções for dilatada por um fator de escala 2^j , ter-se-á:

$$\begin{aligned} \theta_{2^j}(x,y) &= \frac{1}{4^j} \theta\left(\frac{x}{2^j}, \frac{y}{2^j}\right), \\ \psi_{2^j}^1(x,y) &= \frac{1}{4^j} \psi^1\left(\frac{x}{2^j}, \frac{y}{2^j}\right) \quad \text{e} \\ \psi_{2^j}^2(x,y) &= \frac{1}{4^j} \psi^2\left(\frac{x}{2^j}, \frac{y}{2^j}\right) \end{aligned}$$

Para qualquer função $f(x,y) \in L^2(R^2)$, a transformada wavelet definida com respeito a $\psi^1(x,y)$ e $\psi^2(x,y)$ terá dois componentes fornecidos pelas convoluções:

$$W_{2^j}^1 f(x,y) = f * \psi_{2^j}^1(x,y) \quad \text{e} \quad W_{2^j}^2 f(x,y) = f * \psi_{2^j}^2(x,y)$$

A seqüência de funções :

$$\left(W_{2^j}^1 f(x,y), W_{2^j}^2 f(x,y) \right)_{j \in \mathbb{Z}}$$

é a chamada transformada wavelet binária de $f(x,y)$ (vide figura 44).

Já que a transformada wavelet é definida por convoluções, a transformada de Fourier de cada componente pode ser dada por:

$$\begin{aligned} \hat{W}_{2^j}^1 f(\omega, \xi) &= \hat{f}(\omega, \xi) \hat{\psi}^1(2^j \omega, 2^j \xi) \\ \hat{W}_{2^j}^2 f(\omega, \xi) &= \hat{f}(\omega, \xi) \hat{\psi}^2(2^j \omega, 2^j \xi) \end{aligned} \quad \text{e}$$

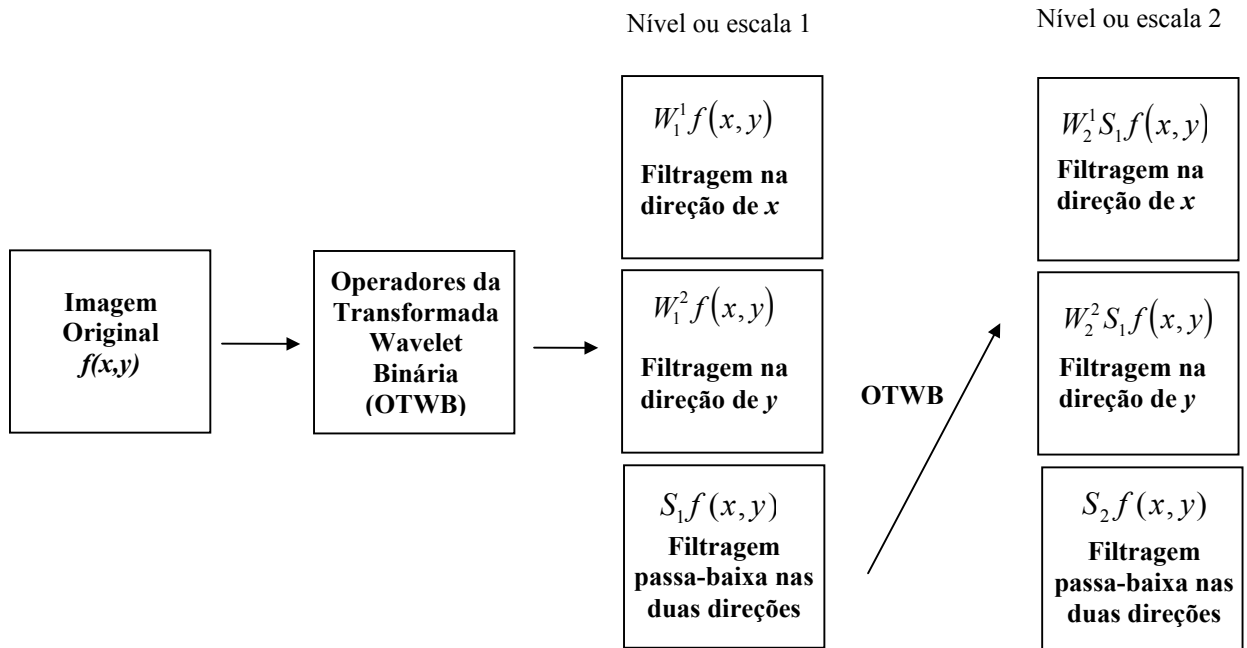


Figura 44 : Diagrama em blocos esquemático da transformada wavelet binária.

A transformada wavelet binária é uma representação completa de um sinal, se e somente se, as dilatações da transformada de Fourier de cada wavelet cobrirem completamente o plano de Fourier. Isso quer dizer que existem duas constantes $A > 0$ e $B > 0$ que:

$$\forall (\omega, \xi) \in R^2, A \leq \sum_{j \in Z} (|\hat{\psi}^1(2^j \omega, 2^j \xi)| + |\hat{\psi}^2(2^j \omega, 2^j \xi)|) \leq B$$

Assim pode-se provar que existe um par não-único de wavelets de reconstrução $\chi^1(x,y)$ e $\chi^2(x,y)$ o qual sua transformada de Fourier satisfaça:

$$\sum_{j \geq 1} (\hat{\psi}^1(2^j \omega, 2^j \xi) \hat{\chi}^1(2^j \omega, 2^j \xi) + \hat{\psi}^2(2^j \omega, 2^j \xi) \hat{\chi}^2(2^j \omega, 2^j \xi)) = 1$$

Qualquer função $f(x,y)$ pode ser reconstruída através de sua transformada wavelet binária com a seguinte soma :

$$f(x,y) = \sum_{j \in Z} (w_{2^j}^1 f * \chi_{2^j}^1(x,y) + w_{2^j}^2 f * \chi_{2^j}^2(x,y)) + s_{2^j}^j$$

A transformada wavelet binária é uma representação redundante do sinal.

Em geral, a análise de um sinal em multiresolução é feita através de um banco de filtros num algoritmo como o da figura 42. No entanto, para atender à condição de invariância por translação, a transformada wavelet binária é calculada através de um algoritmo correspondente a uma adaptação do banco de filtros usual, sem subamostragem e que pode ser visto no apêndice A.

5.3 – Transformada Wavelet Módulo Máximo

Como as wavelets da transformada wavelet binária foram definidas como derivadas parciais da função de suavização $\theta(x,y)$ podemos dizer que:

$$\begin{pmatrix} w_{2^j}^1 f(x,y) \\ w_{2^j}^2 f(x,y) \end{pmatrix} = \begin{pmatrix} 2^j \frac{\partial}{\partial x} (f * \theta_{2^j})(x,y) \\ 2^j \frac{\partial}{\partial y} (f * \theta_{2^j})(x,y) \end{pmatrix} = 2^j \nabla (f * \theta_{2^j})(x,y)$$

Portanto os componentes da transformada wavelet em determinada escala 2^j definem um vetor que é proporcional ao vetor gradiente da imagem suavizada por $\theta_{2^j}(x,y)$. Então o módulo e a fase do vetor gradiente são proporcionais ao módulo e fase da imagem definidos por:

$$M_{2^j} f(x,y) = \sqrt{|w_{2^j}^1 f(x,y)|^2 + |w_{2^j}^2 f(x,y)|^2}$$

$$A_{2^j} f(x,y) = \arctan \left| \frac{w_{2^j}^2 f(x,y)}{w_{2^j}^1 f(x,y)} \right| \quad \text{e}$$

O vetor gradiente aponta na direção em que a derivada parcial de $f * \theta_{2^j}(x,y)$ tem uma amplitude máxima. Então, pode-se definir em cada escala 2^j , o módulo máximo da transformada wavelet como os pontos (x,y) da imagem módulo onde $M_{2^j} f(x,y)$ é localmente máximo ao longo da direção do gradiente fornecido por $A_{2^j} f(x,y)$. Esses módulos máximos são pontos de inflexão de $f * \theta_{2^j}(x,y)$. Esse tipo de abordagem para detecção de bordas, considerando os módulos máximos locais do gradiente da função, foi inicialmente introduzida por Canny [33]. A versão com a transformada wavelet módulo máximo é, portanto, sua versão multiescalas.

As imagens são medidas numa resolução finita. Assim não é possível computar a transformada wavelet abaixo de uma dada escala que é normalizada para 1. Afim de modelar essa limitação de escala, é introduzida uma função real $\phi(x,y)$ cujo módulo da transformada de Fourier satisfaz:

$$|\hat{\phi}(\omega, \xi)|^2 = \sum_{j \geq 1} \hat{\psi}^1(2^j \omega, 2^j \xi) \hat{\chi}^1(2^j \omega, 2^j \xi) + \hat{\psi}^2(2^j \omega, 2^j \xi) \hat{\chi}^2(2^j \omega, 2^j \xi)$$

Pode ser provado então que $\phi(x,y)$ é uma função de suavização e podemos definir um operador de suavização [8] [9]:

$$S_{2^j} f(x, y) = f * \phi_{2^j}(x, y)$$

onde $\phi_{2^j}(x, y)$ é uma dilatação de $\phi(x,y)$ por 2^j . Por causa da resolução finita das medidas, considera-se que a imagem a ser transformada é a suavização, na escala 1, de uma certa imagem original $f(x,y)$ igual a $S_1 f(x, y)$. Chamamos, então, de transformada wavelet binária finita de $S_1 f(x, y)$ como a seqüência:

$$\left\{ (w_{2^j}^1 f(x, y), w_{2^j}^2 f(x, y))_{1 \leq j \leq J}, S_{2^j} f(x, y) \right\}$$

Pode ser provado que $S_1 f(x, y)$ pode ser recuperado a partir de sua transformada wavelet binária finita inversa [8] [9].

Então define-se os módulos máximos da transformada wavelet binária como os pontos onde $M_{2^j} f(x, y)$ é localmente máximo na direção de $A_{2^j} f(x, y)$ em cada escala 2^j , e representam, assim as bordas em cada escala.

5.4 – Vantagens e aspectos práticos da transformada wavelet módulo máximo para a técnica desenvolvida

5.4.1 – Introdução

Na seção 4.4 foram estabelecidas três características relevantes para o método de extração do conjunto de marcadores para a transformação watershed: ser automático, ter imunidade ao ruído ou permitir um tratamento diferenciado deste e fornecer marcadores conectados para os objetos semânticos. A transformada wavelet módulo máximo foi escolhida, como veremos, principalmente devido à possibilidade de tratamento diferenciado do ruído e de fornecer marcadores conectados. De fato, existem muitos métodos automáticos de extração de bordas (o gradiente morfológico é um deles) e por conta só dessa característica a escolha não seria relevante.

Muitas vezes, uma alta variação entre pixels pode não representar, necessariamente, um contorno de um objeto. A textura pode ser definida como um padrão de preenchimento de objetos na imagem. Por exemplo, numa mesma imagem de uma camisa listrada, a textura corresponde às faixas de cores diferentes, se estivermos interessados no contorno da camisa e não em cada uma de suas listras. Os ruídos contidos numa imagem também são caracterizados por variações bruscas de pixel para pixel e que não estão relacionados com as bordas do objeto na imagem. Como foi visto na seção 5.2.1 a transformada wavelet introduz o conceito de multiresolução, que permite analisar um sinal ou imagem em várias escalas, detectando variações em níveis diferentes determinados pelo suporte da função de análise. Dessa forma a representação de uma imagem permite diferenciar arestas e texturas em imagens e a discriminação de ruídos de acordo com o aumento da escala. A transformada wavelet módulo máximo realiza a busca dos módulos máximos da imagem a partir de uma imagem de módulo (seção 5.3). Essas bordas determinadas em cada um dos níveis da transformada conduz a imagens de borda multiescalares. Note na figura 45 que à medida que o nível da transformada aumenta, menos ruídos aparecem.

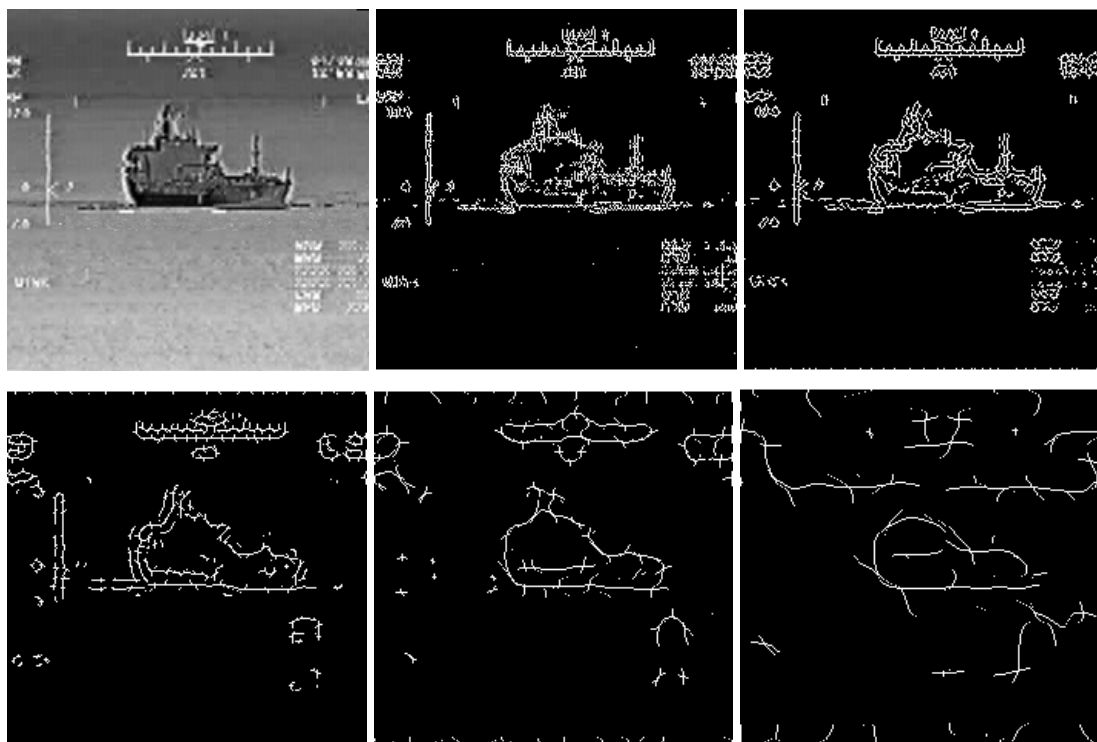


Figura 45 : Módulos máximos ao longo da escala. Da esquerda para a direita e de cima para baixo: Imagem original, módulo máximo para o nível 1, módulo máximo para o nível 2, módulo máximo para o nível 3, módulo máximo para o nível 4 e módulo máximo para o nível 5.

Além disso, a transformada wavelet módulo máximo permite a filtragem de duas maneiras diferentes. Uma é relacionada ao processamento da transformada wavelet binária (seção 5.2), onde a imagem é suavizada em cada escala através de uma filtragem linear, o que resultará, como consequência, em variações bruscas extraídas ou suavizadas de um nível para outro. Além disso é possível realizar uma seleção dessas bordas de maneira não linear (vide seção 6.2.2).

5.4.2 – Extraíndo os mapas de bordas

A transformada wavelet módulo máximo fornece os módulos máximos da imagem a partir da imagem de módulo (vide seção 5.3). Esses módulos máximos da imagem em cada nível correspondem ao mapa de bordas da imagem em cada nível. De maneira diversa a [12] [13] [30] [31] e [32], onde o módulo máximo é pesquisado na direção apontada pelo ângulo fornecido pela imagem de fase $A_2^j f(x, y)$ (vide seção

5.3), foi implementada uma pesquisa nas direções vertical e horizontal. Diferente de [12] [13] [30] [31] e [32], onde o objetivo era a compressão de imagens e era importante se ter o menor número de módulos máximos possível, em nosso caso o objetivo é o de ter os conjuntos de objetos semânticos o mais fechados possível. Essa pesquisa bidimensional traz mais funcionalidade ao método desenvolvido porque aumenta o número de módulos máximos extraídos e a possibilidade do fechamento do conjunto de bordas dos objetos extraídos como pode ser visto na figura 46.

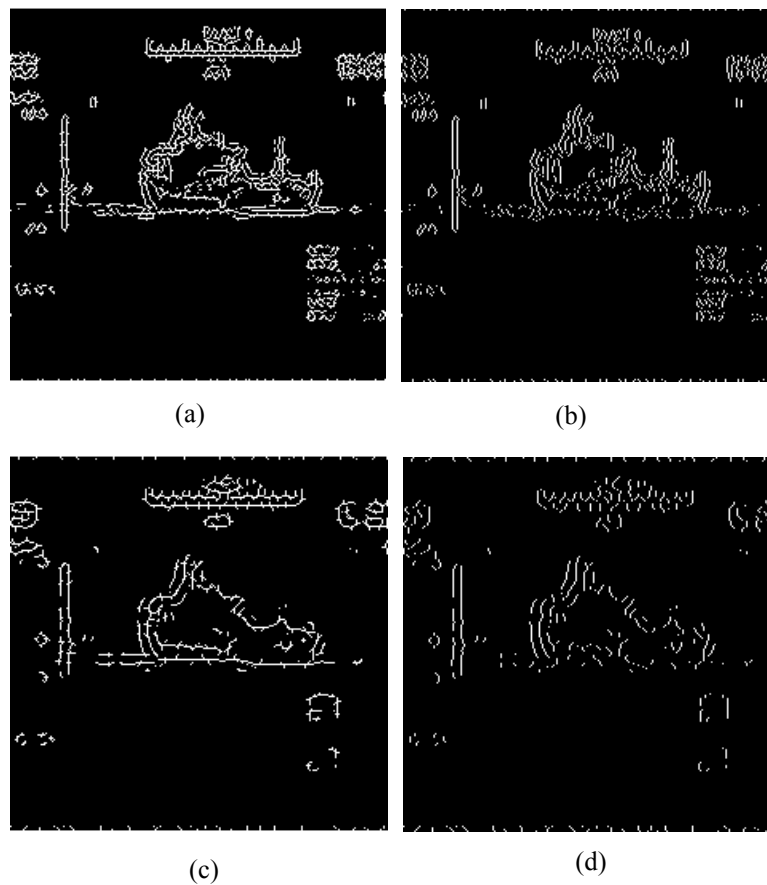


Figura 46: Mapas de módulos máximos com busca bidimensional e tradicional para a imagem original da figura 44. Na coluna da esquerda e de cima para baixo mapas de módulos máximos para os níveis 2 e 3 (a e c) da transformada wavelet binária com busca bidimensional. Na coluna da direita e de cima para baixo mapas de módulos máximos para os níveis 2 e 3 (b e d) da transformada wavelet binária com busca na direção indicada pela imagem de fase.

Além da pesquisa realizada entre vizinhos, os módulos escolhidos como máximos são comparados a um determinado limiar. No processamento da transformada wavelet binária (vide seção 5.2) é realizada uma suavização da imagem. Essa suavização pode ser interpretada como o cálculo de uma média entre valores de pixels.

Se um ruído é um ponto isolado de variação brusca a tendência é que seu valor diminua ao longo dos níveis da transformada. A inclusão de um limiar para comparação tende a suprimir os valores de módulo máximo correspondentes ao ruído. Além disso esse limiar também suprime valores de máximo que apesar de serem maiores que seus vizinhos correspondem a máximos não representativos do conjunto de bordas da imagem. Pode ser visto na figura 47 como a não introdução desse limiar sobrecarrega a imagem de bordas.

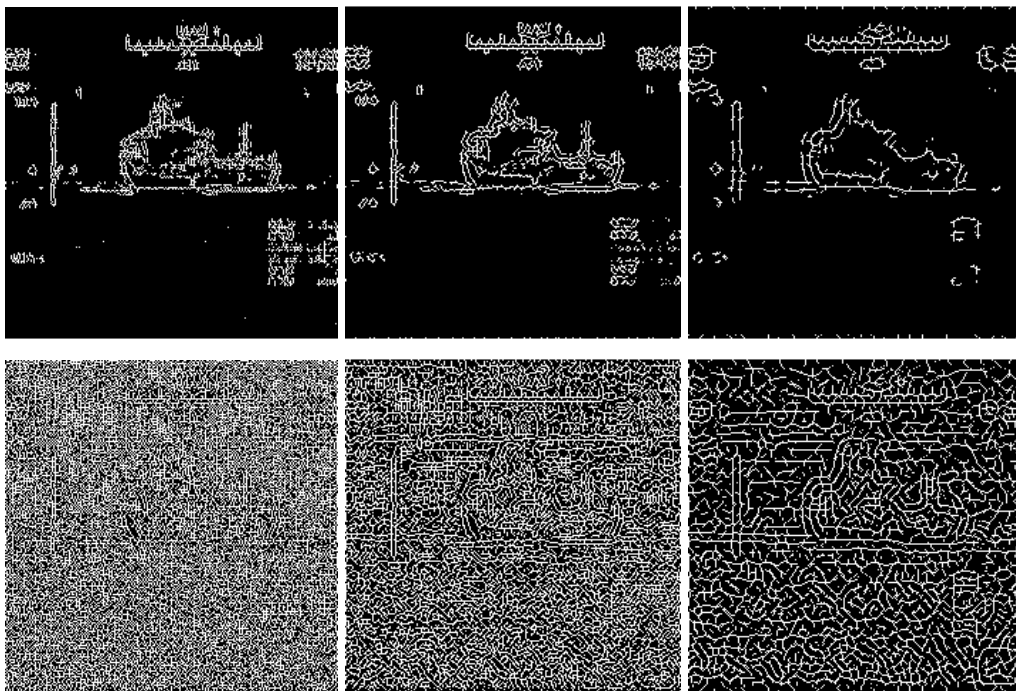


Figura 47 : Mapas de módulos máximos realizados com a comparação com limiar (linha superior) e sem a comparação (linha inferior) para a transformada wavelet módulo máximo realizada para os níveis 1, 2 e 3 respectivamente a partir da imagem original da figura 45.

A determinação desse limiar é totalmente experimental e não há qualquer referência em [12] [13] [30] e [32] sobre sua obtenção, a não ser que ele é baseado no valor máximo da imagem de módulo. Na verdade, ele representa a faixa de valores segundo os quais os módulos máximos serão aceitos como linhas de borda.

Em nossa abordagem foram testados dois tipos básicos de limiar. Um baseado na média e outro no valor máximo do pixel.

A imagem de módulo é uma imagem em que as transições são ressaltadas (vide figura 48), ou seja estão localizadas nos pixels mais elevados. As transições numa imagem, via de regra ocupam uma região menor da imagem. Com o limiar baseado na

média, ou que, em outras palavras, considera a média dos pixels da imagem como o valor mínimo para aceitação de um pixel como módulo máximo, se esperava que ocorresse o efeito de se manter somente os módulos máximos dessas regiões de transição, normalmente com valores de módulo mais elevados que essa média.

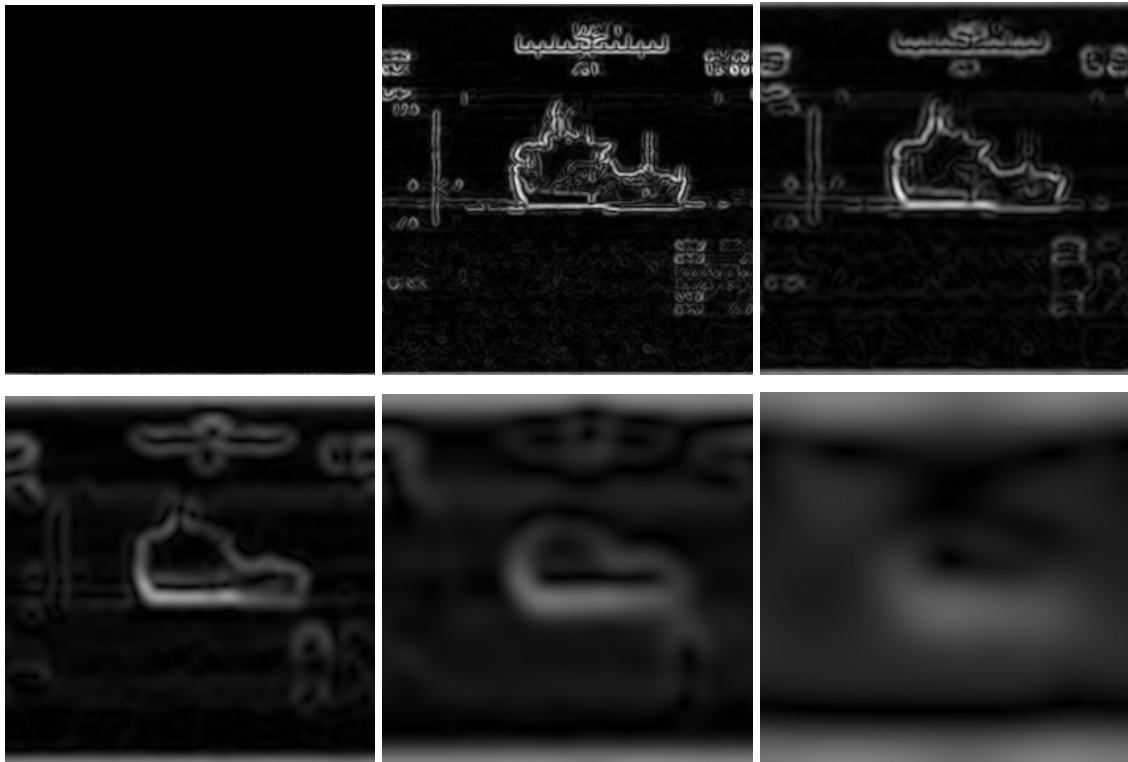


Figura 48 : Imagens de módulo para os níveis 1, 2, 3, 4, 5 e 6 da esquerda para direita e de cima para baixo a partir da imagem original da figura 44.

O limiar baseado no valor máximo dos pixels da imagem de módulo é fundamentado também no fato de que as regiões de transição estão concentradas nos valores de pixel mais elevados. Foi realizada uma divisão desse valor máximo para ajustar seu resultado ao conjunto de imagens de teste. Essa divisão foi realizada com os seguintes divisores 2,5; 5 e 10.

Nas figuras 49 e 50 temos os resultados desses testes para a imagem original da figura 45, nos três primeiros níveis da transformada wavelet módulo máximo. Note que o resultado para o limiar igual à média (figura 50) e ao valor máximo do pixel da imagem de módulo dividido por dez (linhas 3 da figura 49), fornecem um resultado sobrecarregado: são muitos os pixels validados Com o limiar igual ao valor máximo do

pixel da imagem de módulo dividido por 2.5 aconteceu o contrário: poucos pixels foram validados, o que comprometeu a conectividade de alguns conjuntos de marcadores (linha 1 da figura 49). O resultado escolhido foi o igual ao valor máximo do pixel da imagem de módulo dividido por 5 (linha 2 da figura 49) que apresentou coerência de resultado para todas imagens de teste.

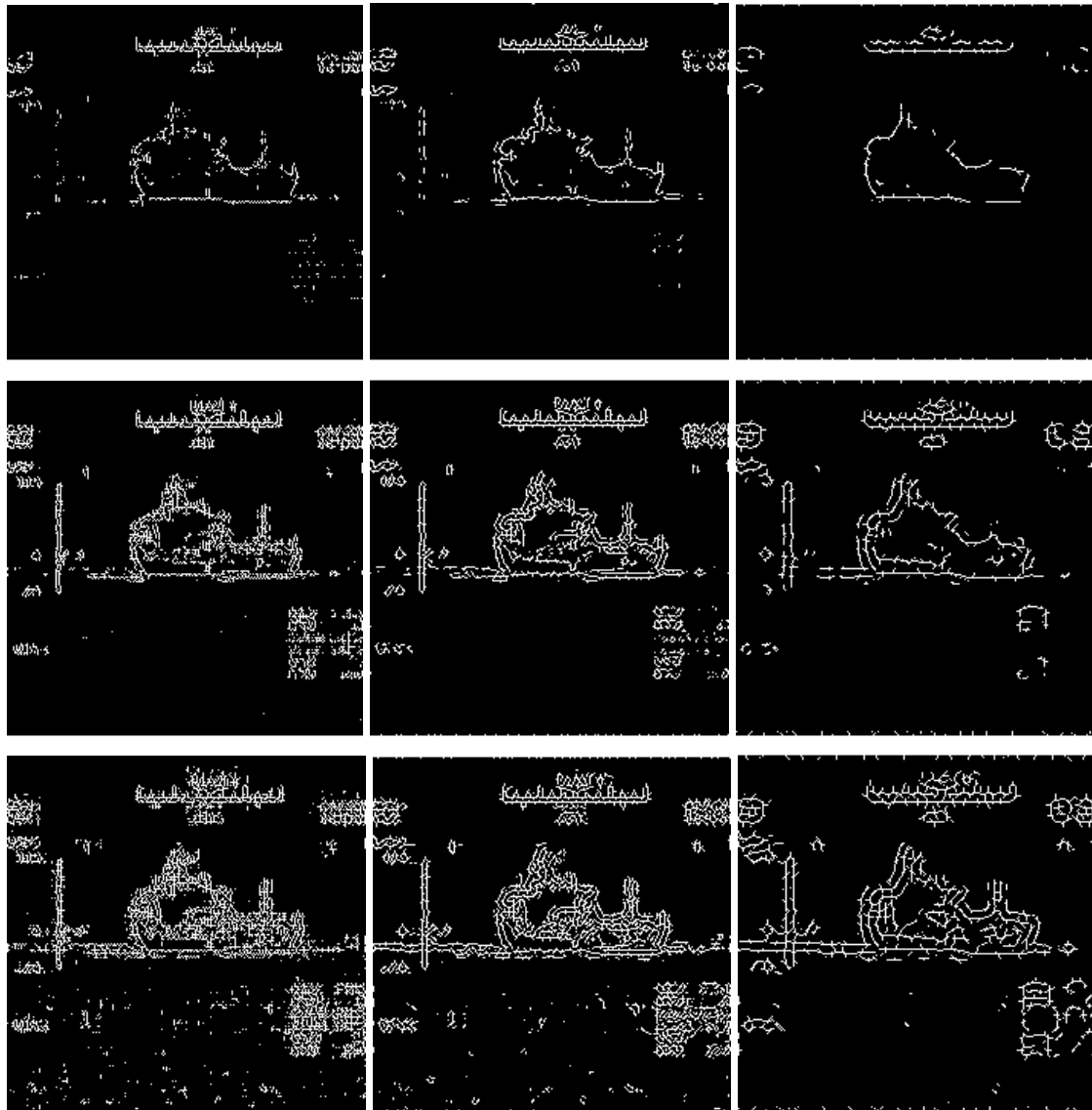


Figura 49 : Teste para determinar o limiar para validar os módulos máximos. Linhas 1 ,2 e 30.02 0 0 10.02 474.2158119

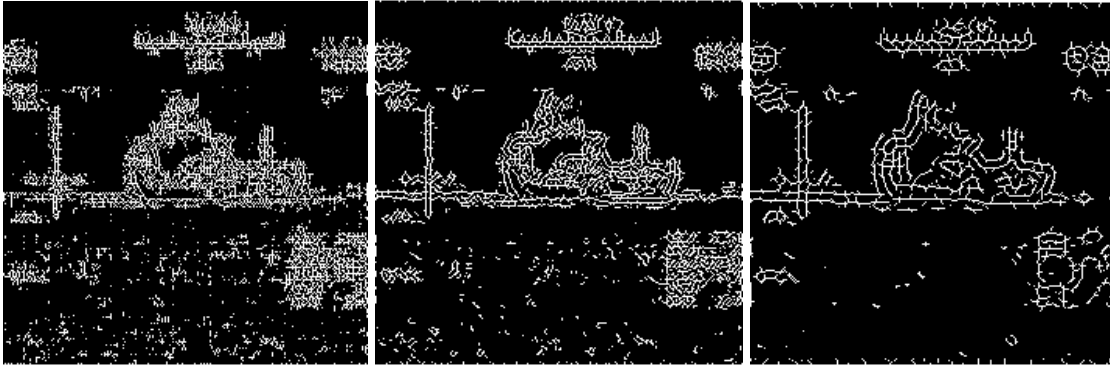


Figura 50 : Teste para determinar o limiar para validar os módulos máximos. Limiar baseado na média para os níveis 1,2 e 3 (da esquerda para a direita).

Essa escolha representa um compromisso entre validar o conjunto de bordas do objeto semântico fechado e o menor número possível alvos falsos. É importante observar que o valor desse limiar deve ser ajustado visando o objetivo da utilização da transformação wavelet módulo máximo. Por exemplo, num sistema de supressão de ruído em que esses valores selecionados servirão à reconstrução da imagem, pode ser importante que seja fornecida uma faixa maior de validação dos módulos para os níveis maiores e menos impregnados de ruído.

5.4.3 – Escolhendo os mapas de borda

Cada nível da transformada wavelet módulo máximo produz um mapa de bordas da imagem. O primeiro passo para a extração do objeto semântico através da transformação watershed consiste em determinar qual desses mapas de bordas deveria ser escolhido para, depois de processado, constituir-se na imagem marcadora. Pode ser observado na figura 45 que, a partir do nível 4, os mapas de borda se mostram ruins no que diz respeito à delimitação precisa do contorno dos objetos semânticos.

A escolha, então, fica restrita aos mapas de bordas dos níveis 1, 2 e 3 da transformada wavelet módulo máximo. O mapa de bordas do primeiro nível está sempre mais contaminado por ruído por conta de que, nesse nível, a imagem ainda não foi suavizada pelo processo da transformada wavelet binária (seção 5.2). O critério dessa escolha é subjetivo e é condicionado ao fato que o mapa de bordas deve conter o objeto semântico com suas bordas formando um conjunto o mais fechado possível e com menor nível de ruído afim de evitar a detecção de alvos falsos.

Na figura 51, 52 ,53 e 54 são vistos alguns resultados dessa escolha. Note que não é uma escolha monótona: cada imagem tem um mapa de borda de um nível diferente associado a essa escolha . Na figura 51 a escolha recai sobre o mapa de bordas no nível 3, por ser a única que apresenta as bordas do objeto semântico conectadas; na figura 52 temos um resultado incomum: o mapa de bordas do nível 1 é o melhor, porém o resultado do mapa de bordas do nível 2 também é muito bom; na figura 53 a escolha recai sobre o mapa de bordas no nível 2, que apresenta mais elementos do objeto semântico e não valida o reflexo do navio como no mapa de bordas 3; por fim, na figura 54 tanto faz utilizar os mapas de borda 2 ou 3 já que apresentam resultados muito próximos.

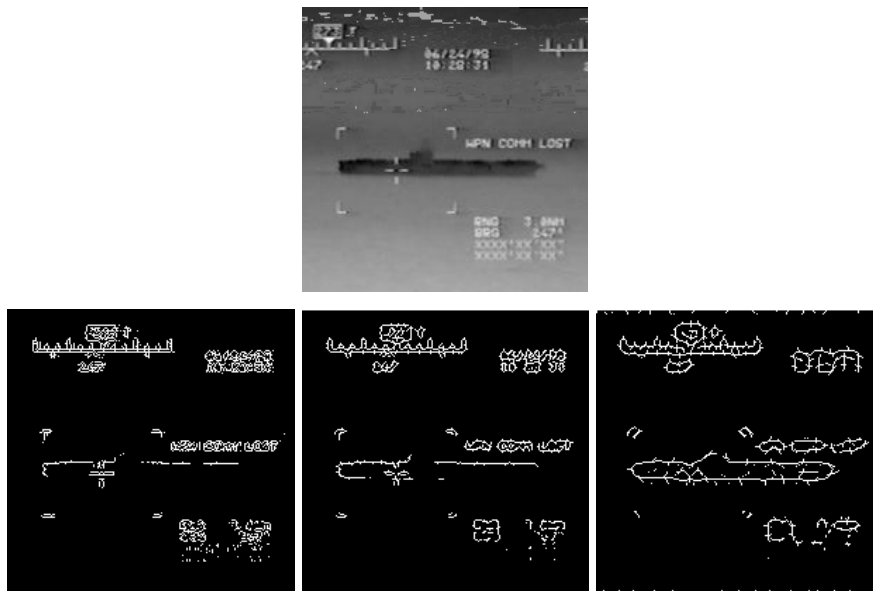


Figura 51 : Mapas de borda para os níveis 1,2 e 3 (da esquerda para a direita) da figura original (linha sobre os mapas de borda).

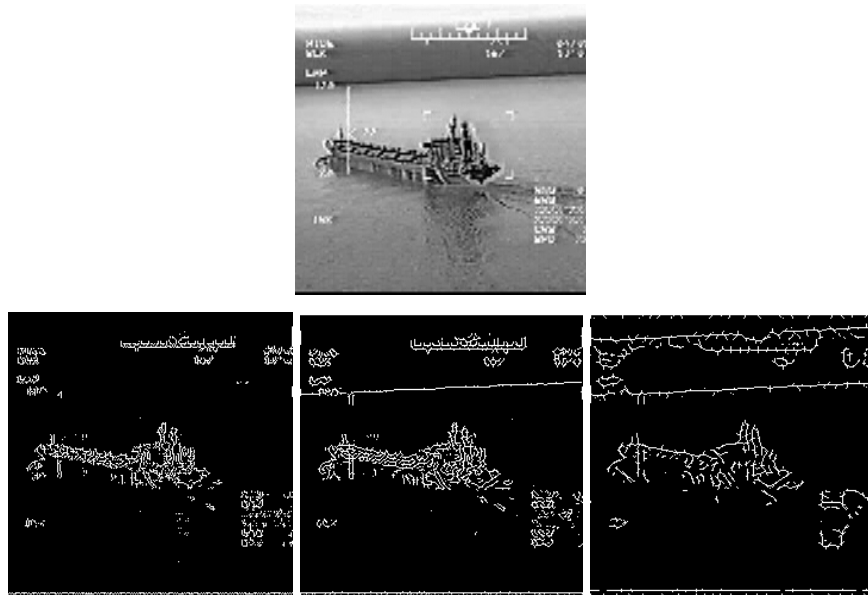


Figura 52 : Mapas de borda para os níveis 1,2 e 3 (da esquerda para a direita) da figura original (linha sobre os mapas de borda).

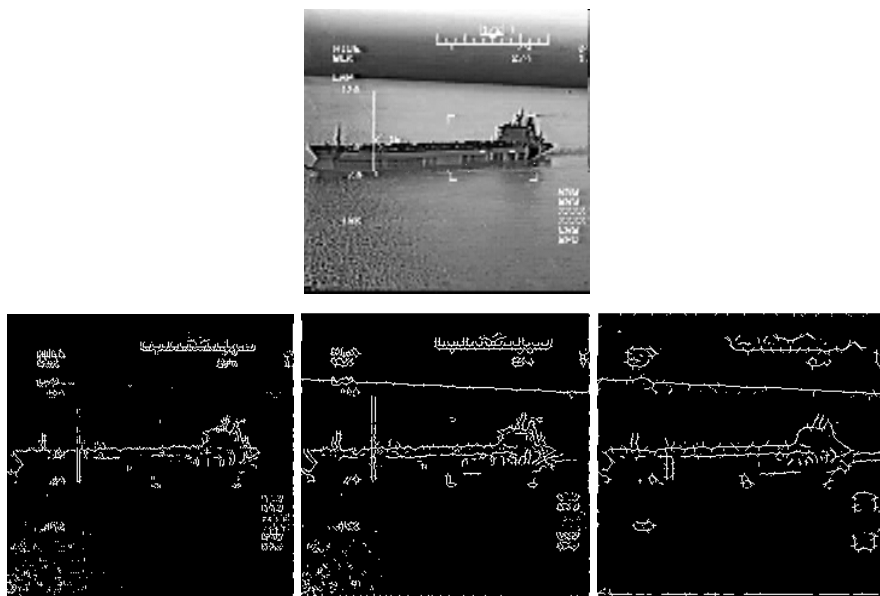


Figura 53 : Mapas de borda para os níveis 1,2 e 3 (da esquerda para a direita) da figura original (linha sobre os mapas de borda).

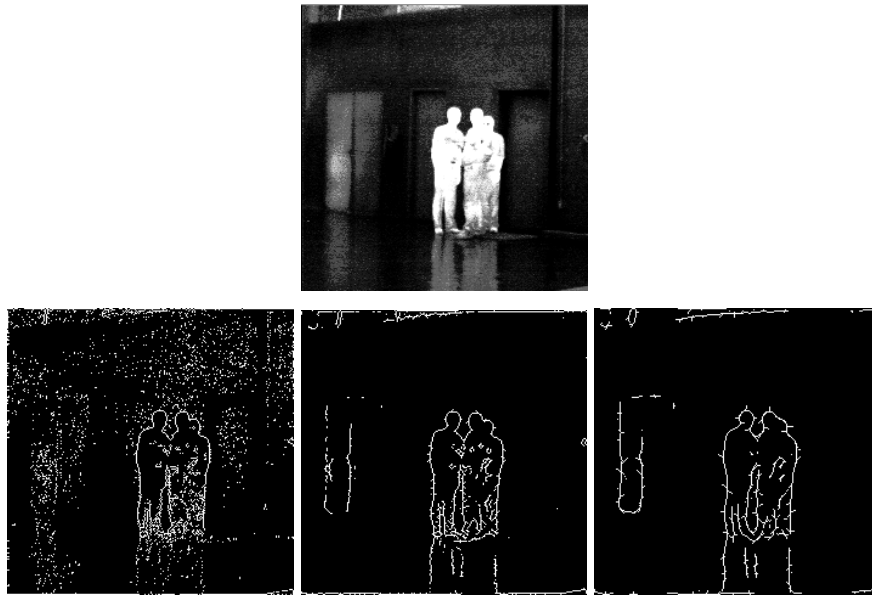


Figura 54 : Mapas de borda para os níveis 1,2 e 3 (da esquerda para a direita) da figura original (linha sobre os mapas de borda).

A grande desvantagem dessa abordagem de se escolher o melhor mapa de imagem para cada imagem, é que esta escolha só pode ser realizada com a supervisão humana. Isso impede o método de segmentação que está sendo criado de ser automático. A multiresolução foi utilizada para resolver esse problema: já que a maior incidência de resultados da escolha recaiam nos mapas de borda dos níveis 2 e 3, e essa escolha era quase sempre definida pelo mapa de bordas que continha o conjunto de bordas do objeto semântico fechado, estes dois mapas foram combinados de uma operação de OU lógico. Isso garante ao processo automação e conjuntos de borda fechados para os objetos semânticos. Com essa combinação, no entanto, é de grande importância a seleção das bordas de importância desse mapa de bordas, pois a combinação desses mapas de borda de dois níveis pode conter elementos indesejados que levem a alvos falsos. Note, por exemplo que a combinação dos mapas de borda dos níveis 2 e 3 das figuras 52 e 53 conduz a mapas de borda mais carregados (figura 55). Uma solução para esse problema será apresentada na seção 6.2.2 com a seleção das bordas consideradas “importantes” para a segmentação.

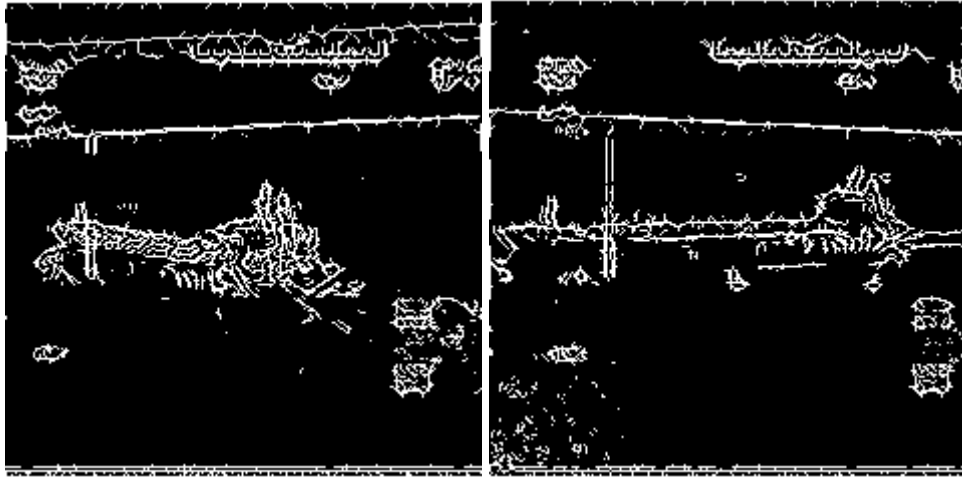
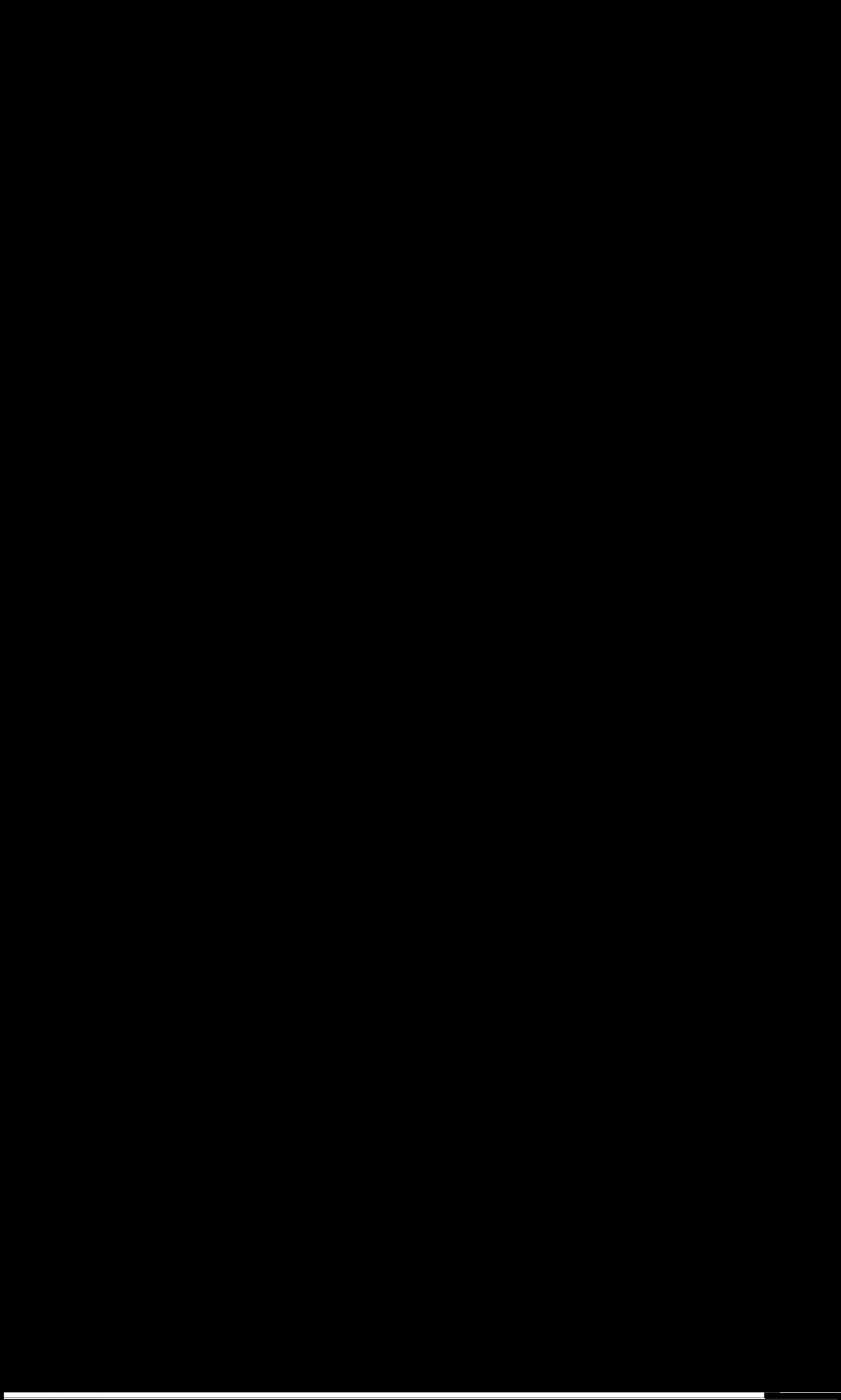


Figura 55 : Mapas de borda combinando os níveis 2 e 3 das imagens 52 e 53.

No próximo capítulo será descrito todo o método desenvolvido.



S

implementação de uma solução para o problema da linha do horizonte (veja a seção 6.3.2). Por fim a dilatação das linhas de borda dos marcadores escolhidos e uma esqueletonização desse resultado são computados e combinados para a obtenção do conjunto de marcadores.

O processo de simplificação da imagem consiste na computação do gradiente morfológico espesso da imagem (vide seção 4.3).

Tendo sido obtidos o marcador e a imagem simplificada é executada a transformação watershed. Uma filtragem final, não linear, é executada para a retirada de eventuais pequenos objetos espúrios que tenham permanecido durante todo o processo de segmentação.

O método foi desenvolvido utilizando o programa MATLAB® 5.3 e o “toolbox” de morfologia desenvolvido pela SDC Information Systems versão 0.14 obtido em <http://www.mmorph.com>. No apêndice B as funções morfológicas utilizadas em nosso método estão relacionadas com as ferramentas do “toolbox” de morfologia.

6.2 – O primeiro passo: detectando os objetos semânticos e construindo o conjunto de marcadores

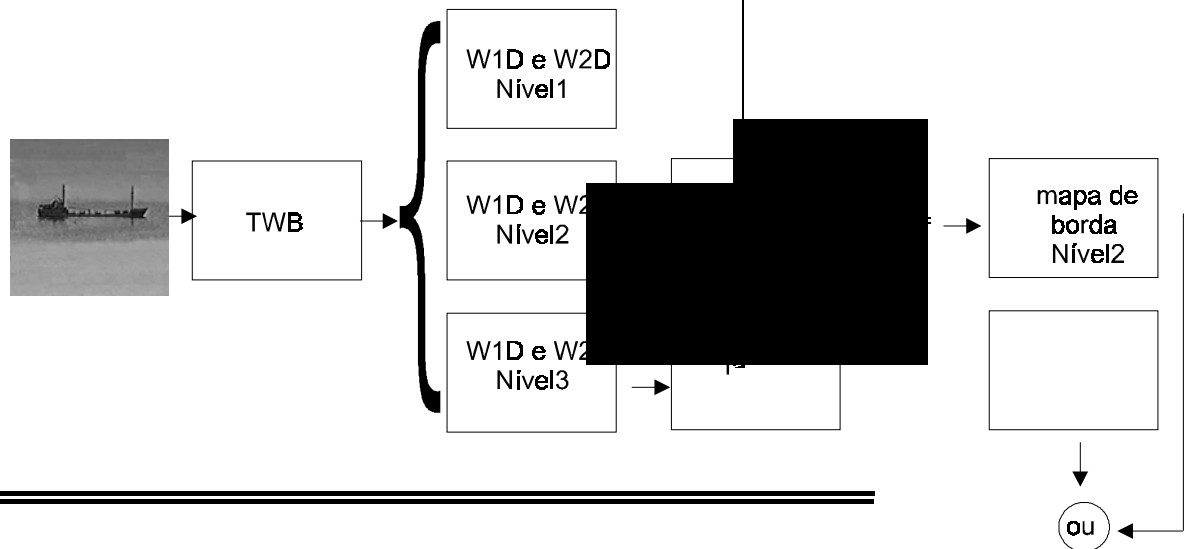
6.2.1- Detectando as bordas dos objetos semânticos

Como o “mecanismo de segmentação” humano, a extração automática de objetos semânticos por computador precisa de informações a priori destes objetos. O conhecimento a priori para a extração de objetos em infravermelho é que estes se encontram em regiões de transientes, destacados por conta da sua diferença de temperatura com o ambiente. Então o problema de detectá-los se transforma no problema de detectar suas bordas. Infelizmente, em grande parte das imagens infravermelho o nível de ruído é alto, o que pode conduzir à segmentação de elementos não semânticos que podem ser tratados como alvos falsos, se pensarmos em aplicações de vigilância e classificação de alvos.

O detetor de bordas escolhido nessa abordagem foi o que se utiliza da transformada wavelet módulo máximo [12] [13]. O uso e o motivo da escolha deste tipo de detetor foi detalhada na seção 5.4.1. Em linhas gerais podemos afirmar que o uso da informação contida no conceito da transformada wavelet módulo máximo é feito através

da validação do pixel por um limiar que foi definido em 5.4.2 e a combinação dos mapas de borda dos níveis 2 e 3 (seção 5.4.3) através de uma operação OU lógico. É importante ressaltar, ainda, que de maneira diversa a [12], [13], [30] e [32] onde o módulo máximo é pesquisado só na direção indicado pela imagem de fase (vide seção 5.4.2), nesse trabalho, já que o objetivo é o de tantos conjuntos de borda de objetos semânticos fechados quanto possível, os módulos máximos são escolhidos nas direções horizontal e vertical. Essa pesquisa em duas direções traz mais funcionalidade ao método desenvolvido pois aumenta a possibilidade do fechamento do conjunto de bordas dos objetos extraídos (vide figura 46).

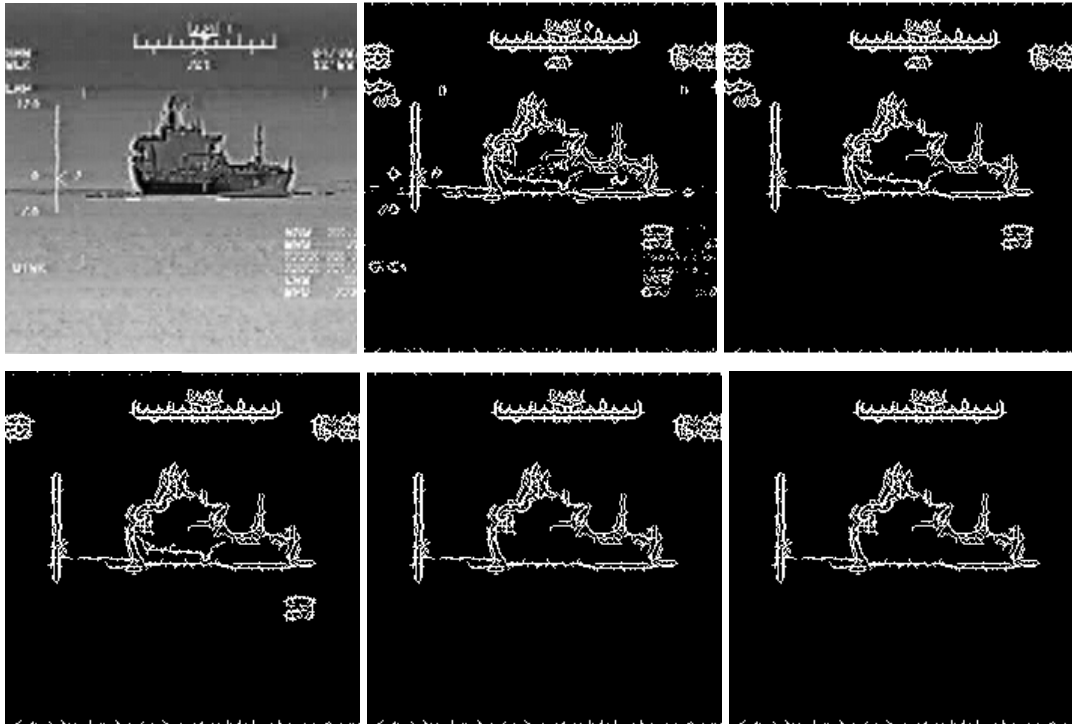
Um diagrama em blocos da detecção de bordas pode ser visto na figura 57.



partir dos componentes da transformada wavelet binária nos níveis dois e três. Agora é necessário escolher dentre essas bordas quais as que pertencem a objetos significativos.

Geralmente, as fronteiras das estruturas coerentes e importantes dentro de uma imagem originam as linhas de borda maiores [13] [31] [32]. Assim foi assumido que as linhas de borda menores correspondem a ruído ou estruturas irrelevantes. Esse objetivo é comum às tarefas de compressão como em [13] [31] e [32]. Porém nesses trabalhos a tarefa de busca dessas estruturas é realizada de maneira diferente: cada pixel de módulo máximo válido da imagem de bordas em cada nível é classificado de acordo com sua vizinhança (quantos vizinhos e qual a direção). A partir dessa classificação e da conseqüente determinação dos pixels de extremo de cada linha de borda, é calculado o tamanho desta linha de borda (número de pixels encadeados) e esta é descartada se menor que um determinado tamanho. Em nosso trabalho, afim de eliminar linhas menores que um determinado tamanho, deixando as outras intactas, o operador morfológico abertura por área foi executado. A escolha desse operador morfológico se deve a sua menor complexidade e menor tempo de processamento. Como foi visto na seção 3.13.1 a abertura por área é um filtro não linear que remove todos os componentes conectados cuja a área da superfície seja menor que um valor dado de limiar.

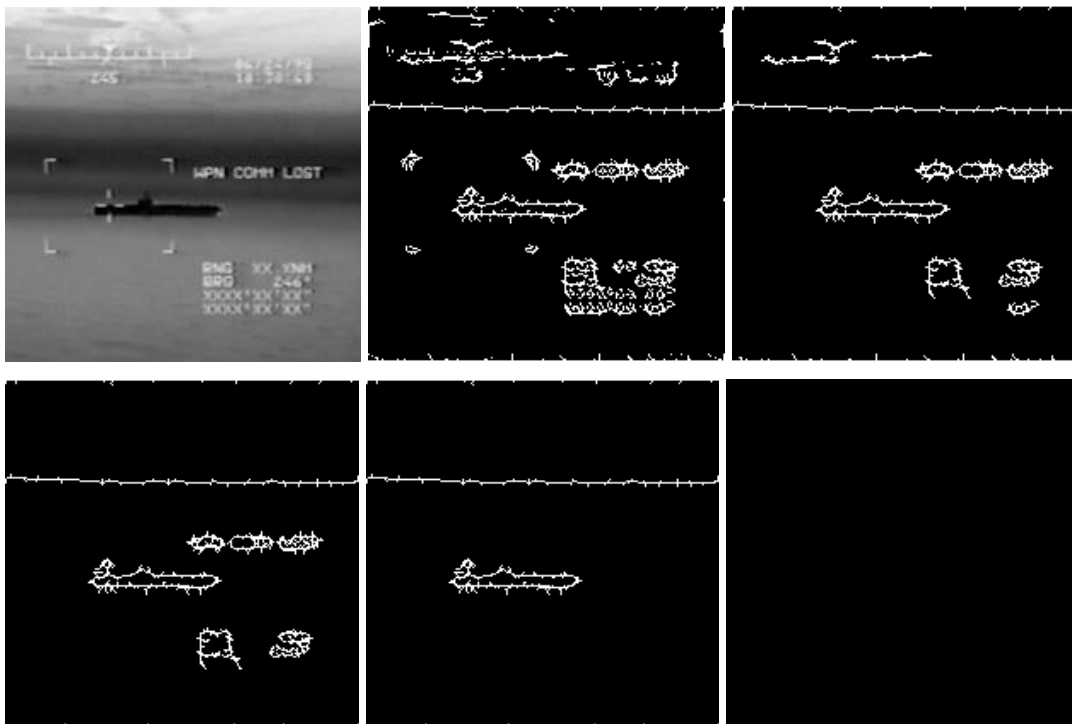
Esse limiar cujo significado, nesse caso, é o número de pixels encadeados a partir do qual vai se considerar uma linha de borda como pertencente a um alvo verdadeiro, foi determinada empiricamente e o seu valor é de 100 pixels encadeados. A sua determinação foi realizada usando 50, 100, 200 e 300 pixels encadeados a cada teste. O valor determinado mostrou-se eficiente ao longo de todas as imagens testadas. Na figura 58 são apresentados alguns resultados dessa determinação. Observe que em alguns casos (exemplo 1) seria possível aumentar o valor desse limiar, entretanto este perderia objetos semânticos muito pequenos (caso do exemplo 2) ou com bordas compostas por pequenos segmentos de linhas. É conveniente observar que esse valor foi ajustado para nosso conjunto de imagens de teste que tinha em vista ser o mais genérico possível e retirar o maior número de linhas de borda não significativas possível. Para aplicações com imagens de distância e tipo específicos esse valor pode ser reajustado.



Exemplo 1

Linha 1: imagem original, mapa de bordas obtido pela combinação dos níveis 2 e 3 da TWMM e abertura por área executado sobre esse mapa de bordas com valor 50.

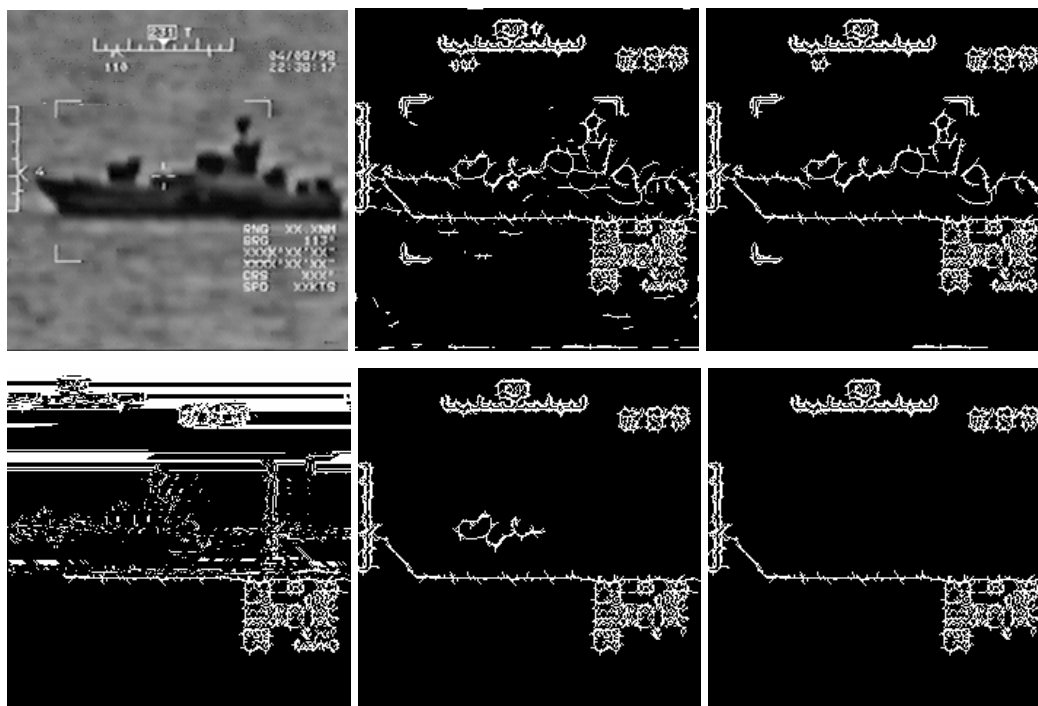
Linha 2: abertura por área executado sobre o mapa de bordas da linha 1 com valores 100, 200 e 300 respectivamente



Exemplo 2

Linha 1: imagem original , mapa de bordas obtido pela combinação dos níveis 2 e 3 da TWMM e abertura por área executado sobre esse mapa de bordas com valor 50

Linha 2: abertura por área executado sobre o mapa de bordas da linha 1 com valores 100, 200 e 300 respectivamente



Exemplo 3

Linha 1: imagem original , mapa de bordas obtido pela combinação dos níveis 2 e 3 da TWMM e abertura por área executado sobre esse mapa de bordas com valor 50

Linha 2: abertura por área executado sobre o mapa de bordas da linha 1 com valores 100, 200 e 300 respectivamente

Figura 58 : Exemplos de teste para determinar o limiar da abertura por área que simplifica o mapa de bordas obtido pela combinação dos níveis 2 e 3 da TWMM.

6.2.3 – O problema do horizonte

Uma das principais aplicações de um sistema de segmentação de imagens em infravermelho é a vigilância. Nesse caso é vital o delineamento preciso do formato do alvo para uma classificação correta, seja ele um alvo marinho, aéreo ou terrestre. Em casos de visões panorâmicas que contenham navios, aviões ou veículos terrestres, o horizonte pode aparecer na imagem e causar o efeito indesejado de ser detectado e considerado como alvo, ou compor um conjunto de bordas errado para a classificação. Em [14] é relacionado um trabalho que, para fugir desse efeito, simplesmente rejeita imagens com horizonte! Em nossa abordagem, já que o horizonte compõe uma borda suficientemente grande para passar pela operação de abertura por área (seção 6.2.2), esta pode compor um marcador com as margens da imagem e ser detectada pela transformação watershed. Para a solução desse problema foi idealizado um processo

baseado em morfologia matemática cujo diagrama em blocos pode ser visto na figura 59.

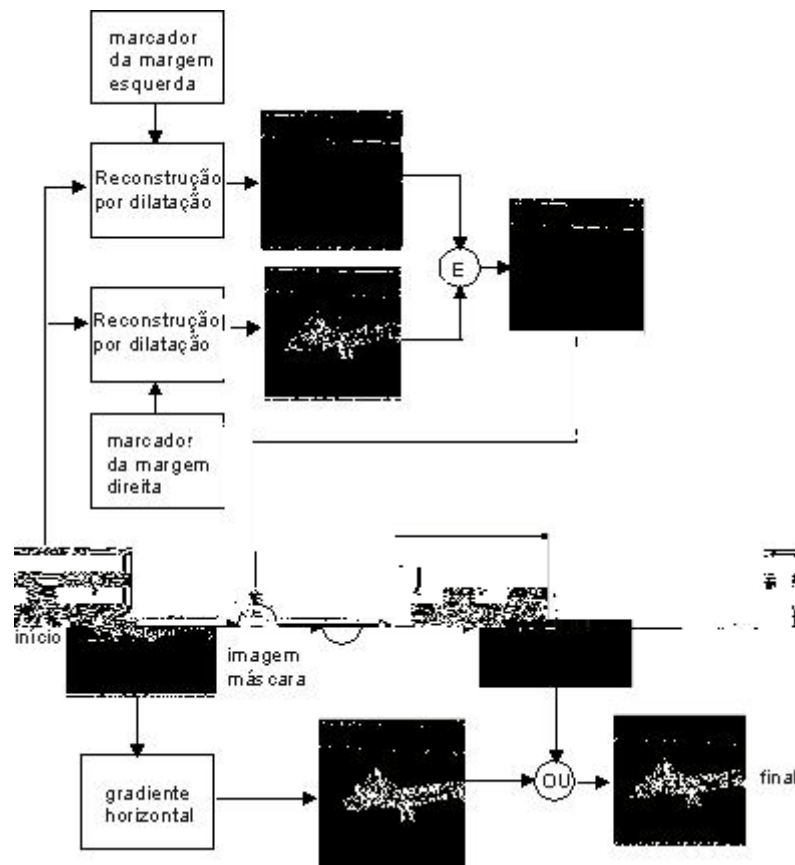


Figura 59 : Diagrama em blocos do processo de retirada da linha do horizonte.

Primeiro, antes da descrição desse processo, é preciso se ter a noção do que é o horizonte para uma imagem. O horizonte pode ser considerado como uma linha horizontal, ou com alguma inclinação, que toca as duas margens verticais da imagem. Portanto, testar se a imagem tem elementos que tocam simultaneamente as duas margens é o primeiro passo para uma solução eficaz. A partir do resultado positivo deste teste, esses elementos devem ser retirados. Isso é realizado com a utilização de uma operação lógica E de duas reconstruções morfológicas por dilatação (vide seção 3.12.3.1). As imagens marcadoras para essas reconstruções são construídas da seguinte maneira: uma com as duas colunas de pixel mais próximas à margem esquerda e outra com as duas colunas de pixel mais próximas à margem direita. Essas duas imagens marcadoras tem como imagem máscara a própria imagem das bordas escolhidas, que é

aquela que resultou da combinação dos níveis 2 e 3 da transformada wavelet módulo máximo (veja diagrama da figura 57) depois de passar pelo operador morfológico abertura por área (seção 6.2.2). A operação de reconstrução por dilatação com o marcador da margem esquerda seleciona todos os elementos que tocam a margem esquerda, enquanto que com o marcador da margem direita são selecionados os elementos que tocam a margem direita (vide figura 60).

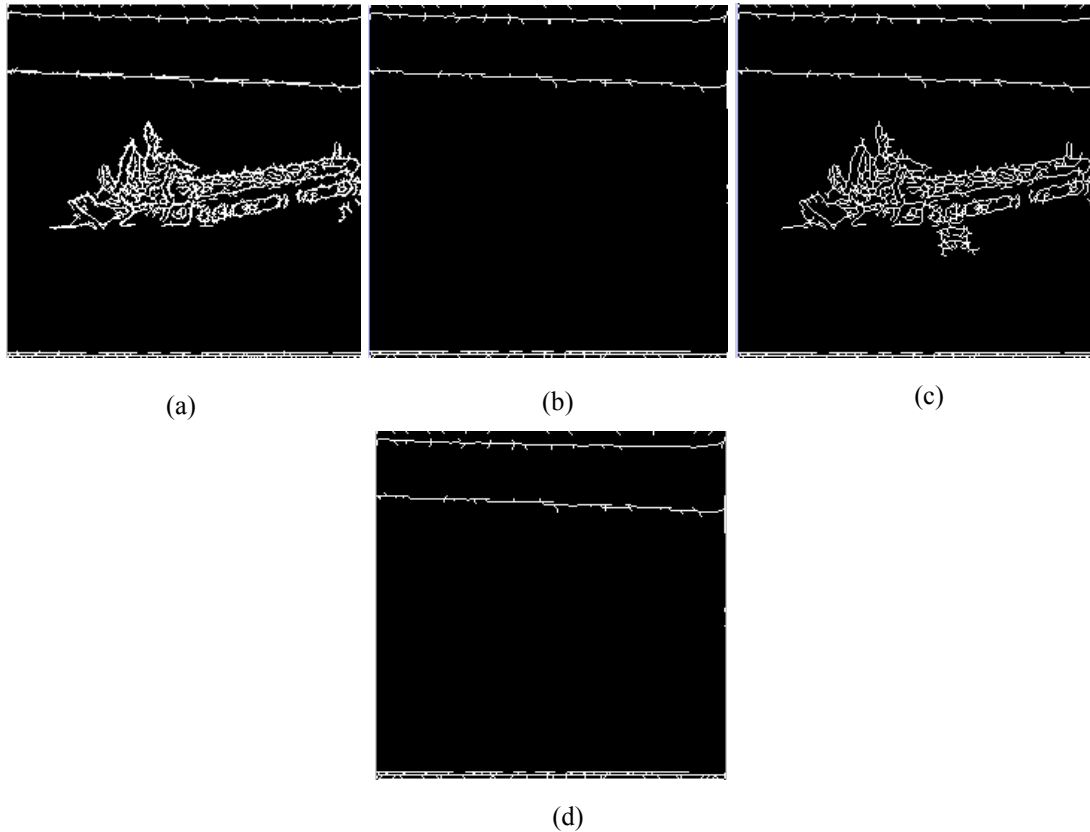


Figura 60 : Retirada de objetos que tocam simultaneamente as duas margens da imagem imagem máscara (a), reconstrução por dilatação com uma imagem marcadora (semente) constituída pelas duas colunas mais próximas a margem esquerda (b), reconstrução por dilatação com uma imagem marcadora (semente) constituída pelas duas colunas mais próximas a margem direita (c) e E lógico dessas duas reconstruções.(d).

A operação E lógico entre estas duas imagens resulta em uma imagem com os elementos que tocam as duas margens ao mesmo tempo. Bastaria diminuir a imagem das bordas escolhidas (imagem máscara da figura 59) por essa das bordas que tocam simultaneamente as duas margens para eliminar o horizonte, se apenas a linha do horizonte tocasse as duas margens. No entanto, esse procedimento retiraria todas as

estruturas que tocam as duas margens, bem como objetos que tocassem o horizonte, apesar de serem objetos semânticos (vide figura 61).

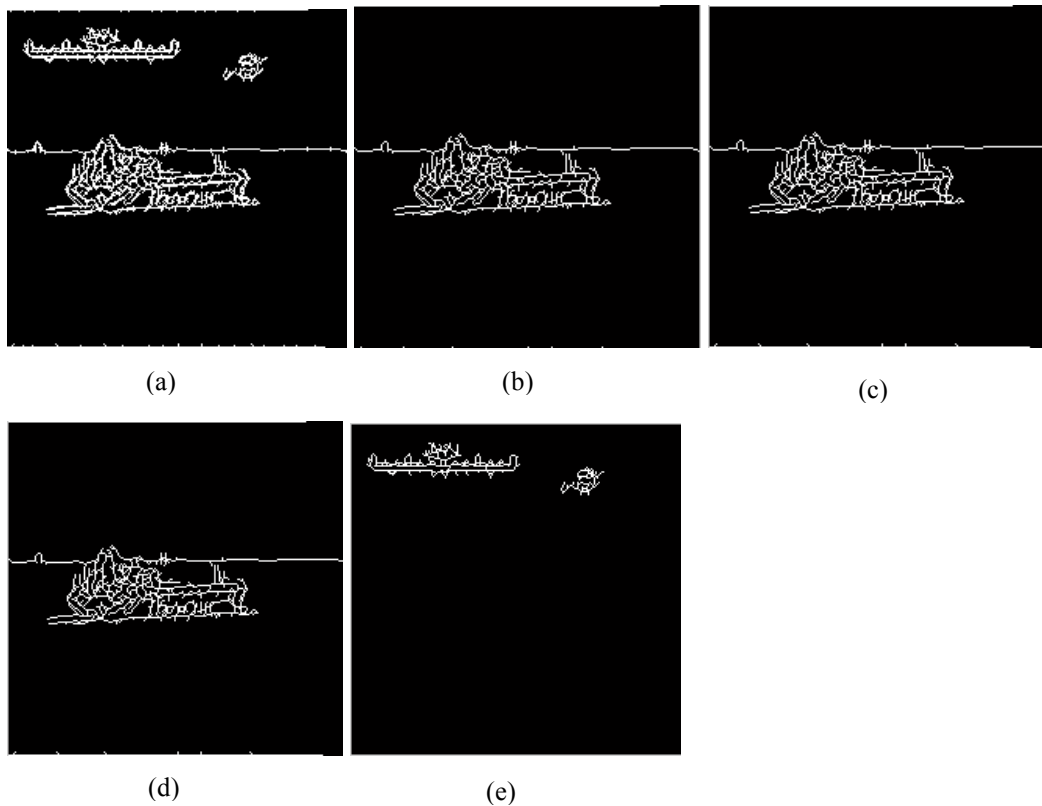


Figura 61 : Problema da linha do horizonte: imagem máscara (a), reconstrução por dilatação com uma imagem marcadora (semente) constituída pelas duas colunas mais próximas a margem esquerda (b), reconstrução por dilatação com uma imagem marcadora (semente) constituída pelas duas colunas mais próximas a margem direita (c) , E lógico dessas duas reconstruções (d) e imagem máscara menos a imagem formada pelo E lógico das duas reconstruções por dilatação(e).

Esse problema foi evitado com a realização de um gradiente morfológico horizontal (seção 3.3). Esse tipo de gradiente preserva elementos que tocam a linha do horizonte ou as duas margens mas não são completamente horizontais. É preciso atentar-se ao fato de que estamos obtendo o marcador para a segmentação. Assim, mesmo que esse processo de retirada dos efeitos da linha do horizonte desconecte conjuntos fechados em pequenos segmentos a dilatação a ser realizada na fase seguinte (seção 6.2.4) voltará a fechar esses conjuntos. Na figura 62 é possível ver o efeito desse gradiente horizontal.

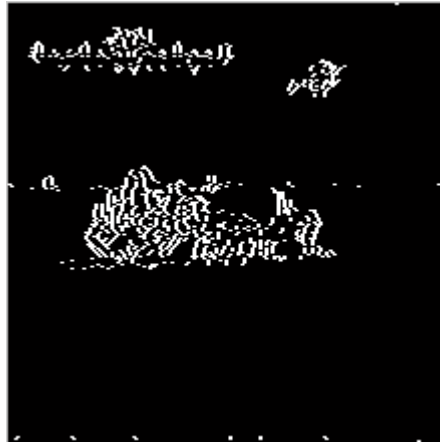


Figura 62 : Gradiente horizontal como parte do processo para a solução do problema da linha do horizonte.

A imagem máscara é diminuída daquela composta pela combinação das reconstruções morfológicas por dilatação e combinada a esta composta pelo gradiente direcional horizontal através de um OU lógico, o resultado comporá o conjunto final das bordas escolhidas como marcadores (vide figura 63).

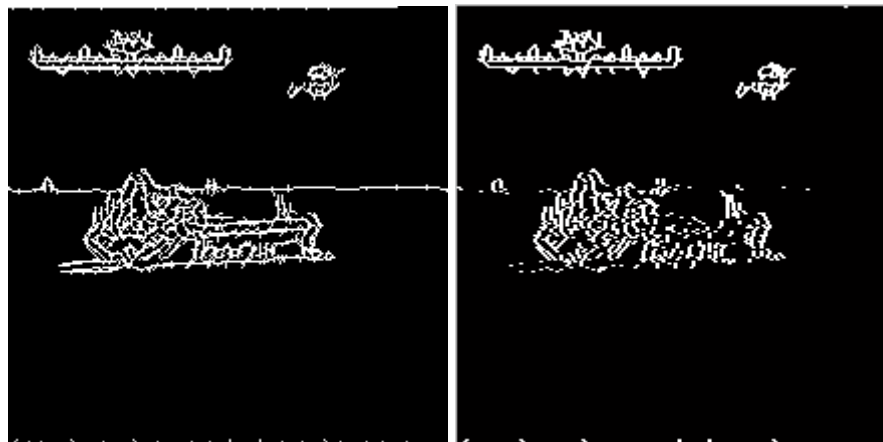


Figura 63 : Mapa de bordas antes (esquerda) e depois (direita) do tratamento do problema do horizonte.

6.2.4 – Finalizando o marcador

O procedimento para a finalização do marcador é mostrado no diagrama em blocos da figura 64.

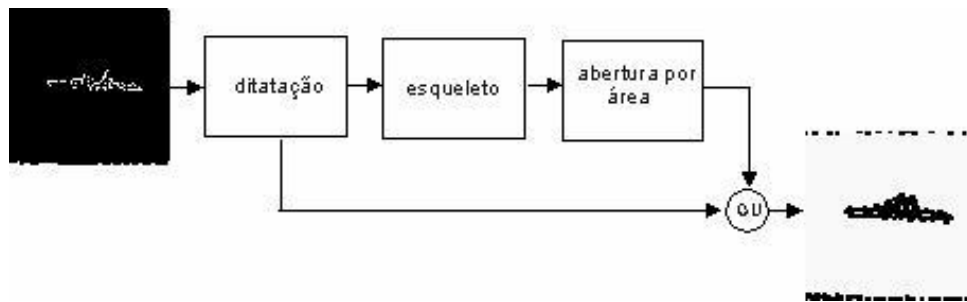


Figura 64 : Diagrama em blocos do processo de finalização do marcador para a transformação watershed.

Nesse ponto, a imagem binária obtida, que representa o conjunto das bordas escolhidas como marcadores, é dilatada usando um elemento estruturante quadrado de tamanho 6X6 pixels . Essa dilatação, como pode ser visto na seção 4.3, aumenta a região da imagem em que será realizada a imposição de mínimo para a transformação watershed, garantindo assim que as bordas do objeto semântico estarão contidas nessa região. O processo da transformação watershed considera como marcador válido um conjunto binário que possua, pelo menos, um pixel branco cercado por uma região de pixels pretos. Essa dilatação realizada pode fazer com que isso não aconteça para pequenas estruturas de alvos verdadeiros. Para garantir que mesmo essas pequenas estruturas possuam um pixel branco dentro de uma região preta uma operação de esqueletonização da imagem dilatada seguida por uma abertura por área nesse esqueleto é realizada. A imagem dilatada e este esqueleto filtrado são combinados com uma operação lógica OU (vide figura 65). Essa abertura por área, realizada após a esqueletonização, é executada de forma a diminuir os pixels brancos inseridos dentro das regiões negras, tentando, assim, evitar que pequenas estruturas que tenham atravessado todo o processo até aqui e que são objetos irrelevantes sejam segmentados. Seu tamanho determinado empiricamente é de 10 pixels encadeados.

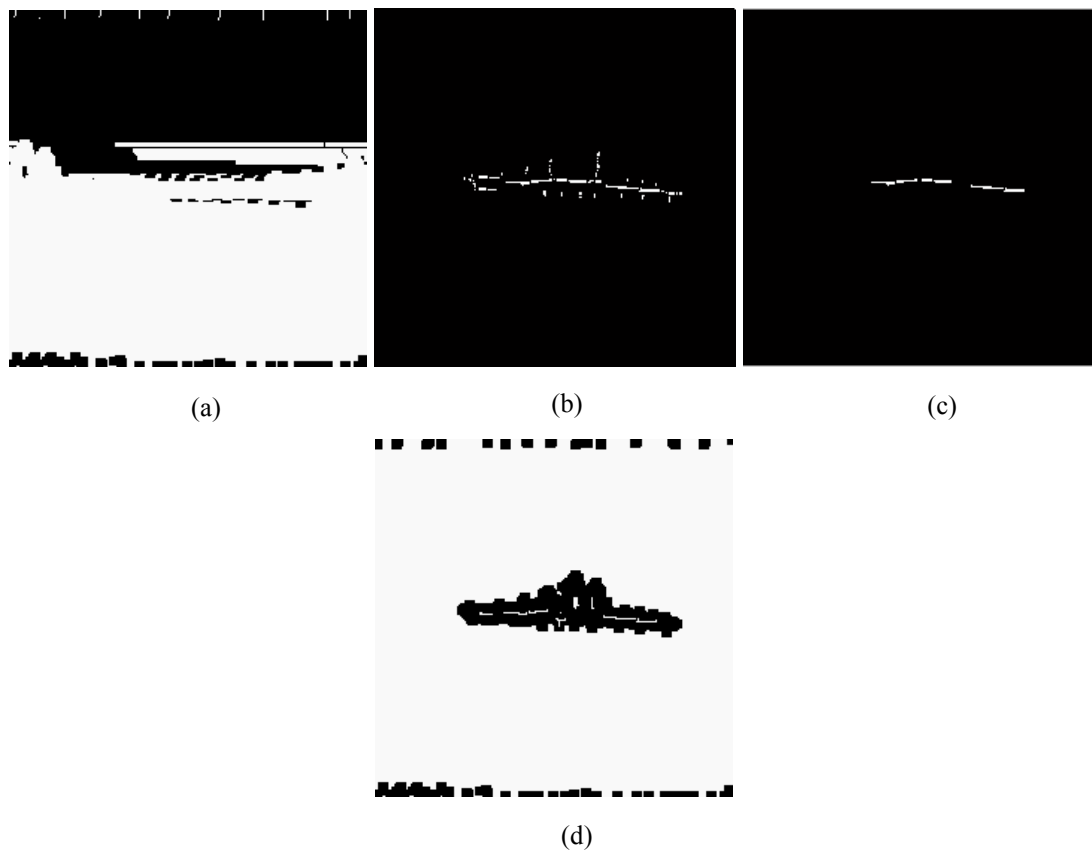


Figura 65 : Esqueletonização do marcador. Marcador antes da esqueletonização (a), esqueleto do marcador (b), abertura por área do esqueleto (c), marcador final representando uma operação de E lógico entre o marcador inicial e o esqueleto reduzido(d)

Na figura 66 pode ser visto um marcador com e sem o uso desse artifício baseado no esqueleto, bem como o produto final relacionado a este.

Convém observar que para imagens muito impregnadas de ruído, tais como vídeo sintético, que formam estruturas que são detectadas pelo procedimento descrito até aqui, a esqueletonização/abertura por área deve ser abandonada, afim de manter-se bons resultados de segmentação (veja imagens da figura 74 – página 110).

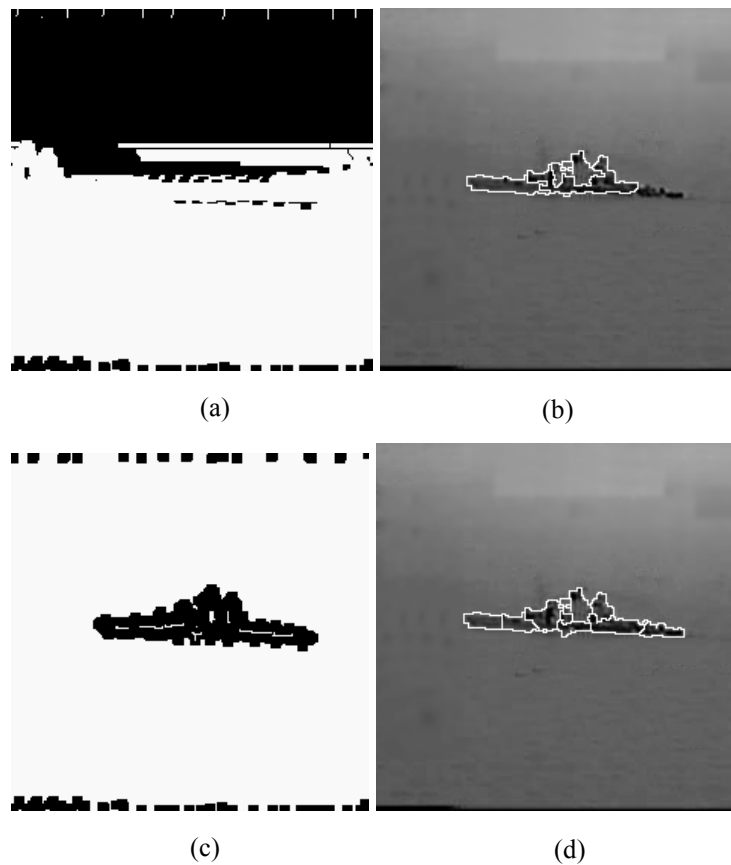


Figura 66 : Marcadores propostos. Da esquerda para direita e de cima para baixo: marcador sem a composição com o esqueleto (a) e resultado final da segmentação (b); marcador com a composição com o esqueleto (c) e resultado final da segmentação (d)

6.3 – Segundo passo: a simplificação da imagem

A aplicação da transformação watershed diretamente na imagem, sujeita o resultado a sobre-segmentação devido à presença de mínimos locais não significativos. A maneira de evitar essa segmentação excessiva já foi abordada por Meyer e Beucher [11] que estabeleceram o uso de marcadores e a simplificação da imagem como procedimentos chave. No método aqui proposto o procedimento para a simplificação já foi descrito (seção 4.3). Esta é realizada com um gradiente morfológico espesso (seção 4.3). Na figura 67 vemos o efeito da simplificação.

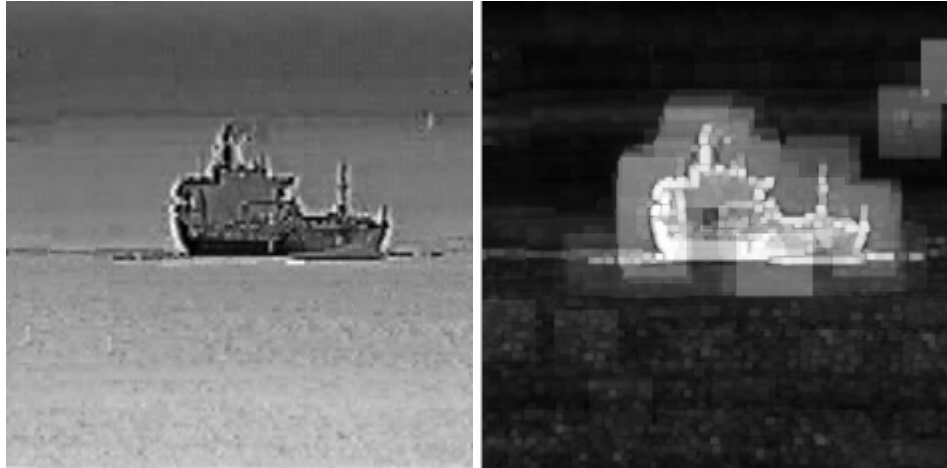


Figura 67 : Imagem original e imagem simplificada.

6.4 – Último passo: a transformação watershed

Já que o marcador foi obtido e a imagem original foram simplificados, a transformação watershed pode ser executada. Depois da computação desta, uma abertura por área é executada para a retirada de qualquer objeto segmentado que possua menos de 150 pixels encadeados. Essa abertura por área segue a mesma filosofia de considerar que elementos de borda maiores (ou com mais pixels encadeados) são os mais importantes na cena, e, portanto, pequenas estruturas tendem a ser alvos falsos que permaneceram mesmo depois de todo o processo de segmentação e também foi determinada empiricamente. Na figura 68 é possível ver a influência dessa abertura por área no resultado final. Note que apesar do limiar dessa abertura por área ser maior do que aquele desenvolvido na seção 6.2.2 para a retirada de bordas não significativas na imagem de bordas obtida pela combinação de dois mapas de borda selecionados da transformada wavelet módulo máximo (seção 5.4.3), não há prejuízo para o contorno do objeto semântico pois a tendência da transformada watershed é conectar essas bordas, aumentando, assim, o número de pixels encadeados pelo contorno do objeto e pelos alvos falsos.

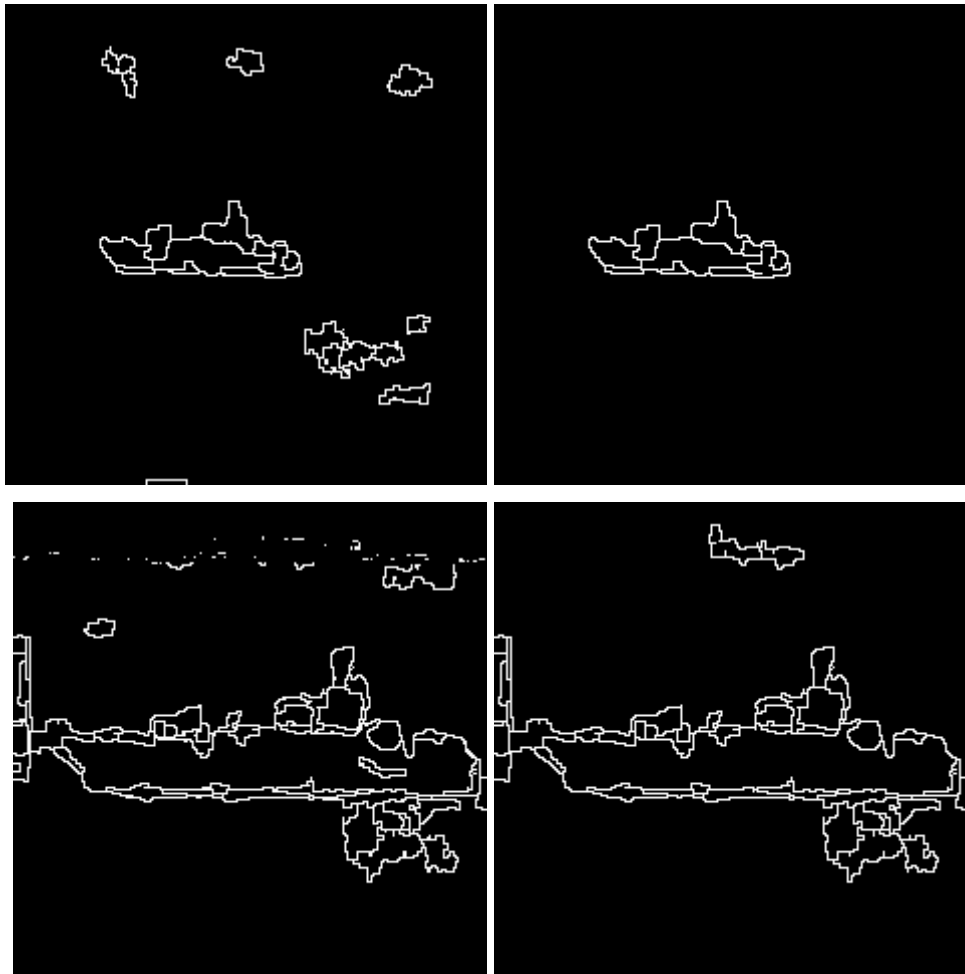


Figura 68 : Watersheds antes (esquerda) e depois do filtro final de abertura por área.

7 - RESULTADOS

7.1 - Introdução

Afim de avaliar a eficiência do método proposto, um conjunto bastante diverso de imagens em infravermelho foi segmentado com o uso deste. Essas imagens foram obtidas através de uma camera AN/AAS-44V “Forward-Looking Infrared” (FLIR), montada no nariz de um helicóptero SH-60B e de uma camera FLIR Prism DS fixa em um laboratório e em locações. Fazem parte deste conjunto de imagens navios, carros de combate, pessoas e objetos. Essas imagens possuem objetos semânticos acima e abaixo da temperatura ambiente indiscriminadamente, em cenários próximos ou em vistas panorâmicas.

A avaliação dos resultados foi realizada de maneira subjetiva, a partir da observação da presença ou não de alvos falsos e da exatidão da borda final extraída quando esta é sobreposta à imagem original.

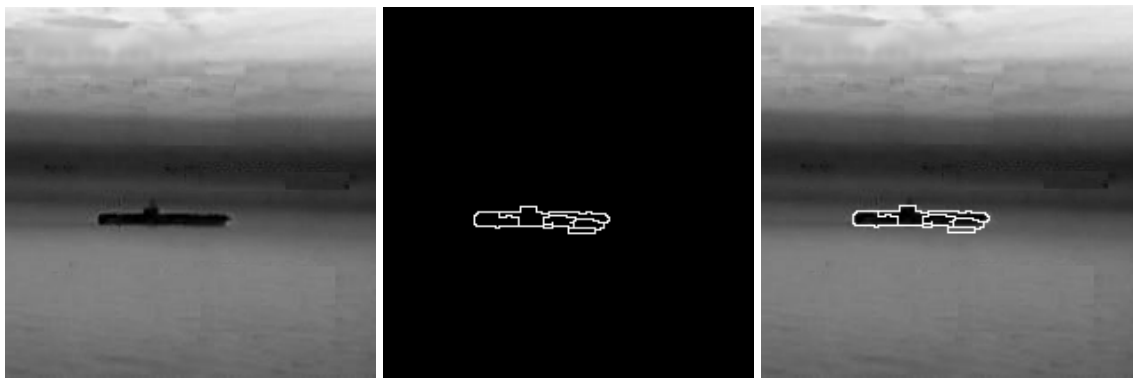
Ao longo do método algumas variáveis foram obtidas empiricamente e validadas através do conjunto de imagens de teste. São estas:

- Limiar para a validação do módulo máximo da imagem de módulo obtida através dos componente da transformada wavelet binária (seção 5.2).
- Linhas de borda “relevantes” (número de pixels encadeados) do mapa de bordas obtido através da combinação dos mapas de borda dos níveis 2 e 3 (seção 6.2.2).
- Tamanho dos objetos irrelevantes que serão desprezados pela filtragem final (seção 6.4).

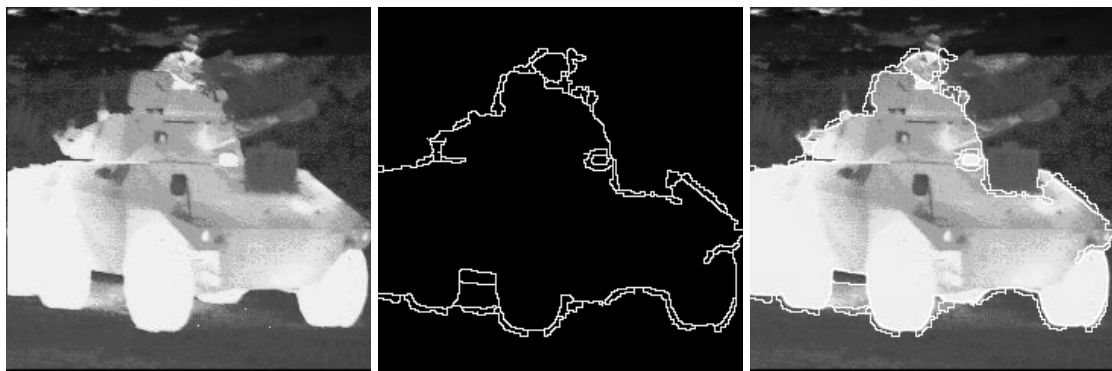
Deve ser observado que a escolha de um conjunto de imagens bastante diverso foi proposital, com a intenção de se ajustar essas variáveis para uma solução o mais genérica possível. No caso do uso desse método para solução de algum problema específico, é possível o ajuste dessas variáveis para se aumentar ainda mais a eficácia dessa abordagem.

7.2 – Generalidade do resultado

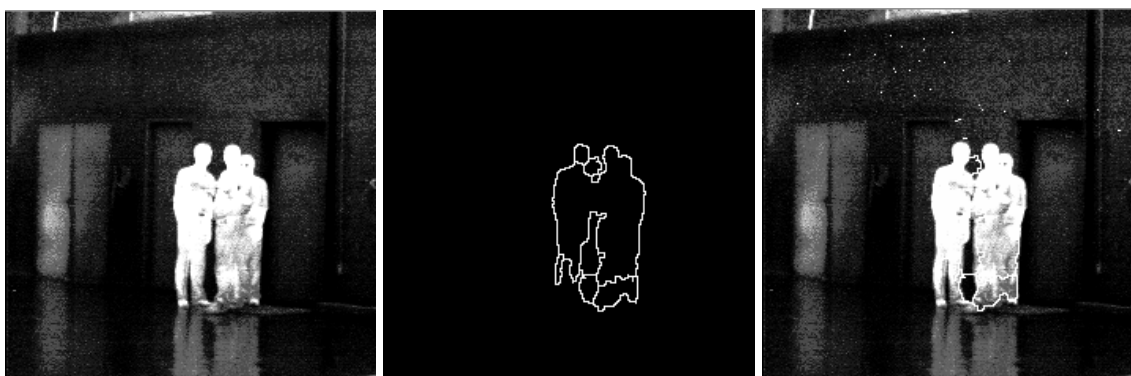
Na figura 69 são apresentados quatro resultados de segmentação para quatro tipos de objetos semânticos diferentes. Além disso estes se encontram em temperaturas superiores (b e c) ou inferiores (a e d) ao meio ambiente, e em várias distâncias.



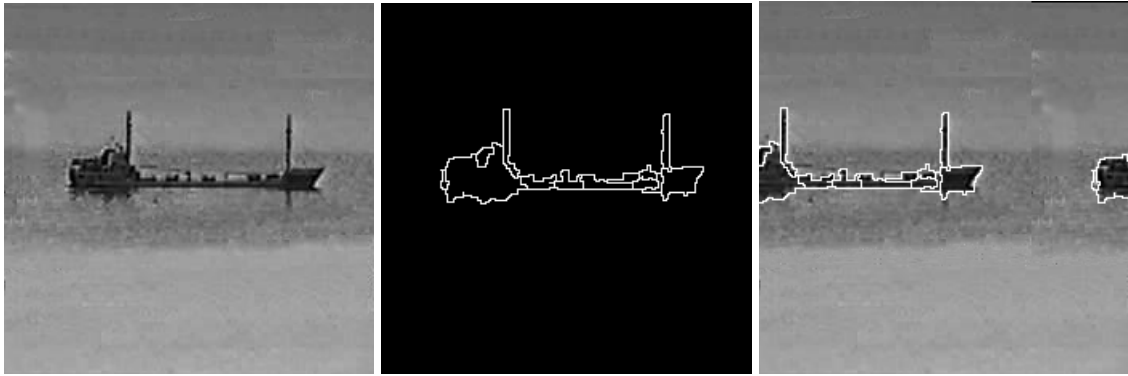
(a)



(b)



(c)



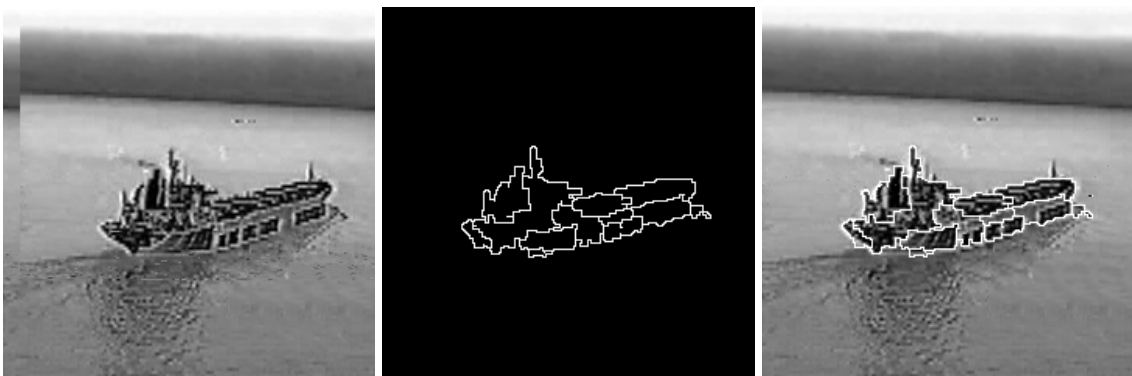
(d)

Figura 69 : Resultados da segmentação. Da esquerda para a direita em todos os quatro exemplos: imagem original, bordas extraídas e bordas extraídas sobrepostas às imagens.(a) navio distante e mais frio que o ambiente, (b) carro de combate próximo e mais quente que o ambiente, (c) grupo de pessoas a média distância e mais quente que o ambiente e (d) navio a média distância e mais frio que o ambiente.

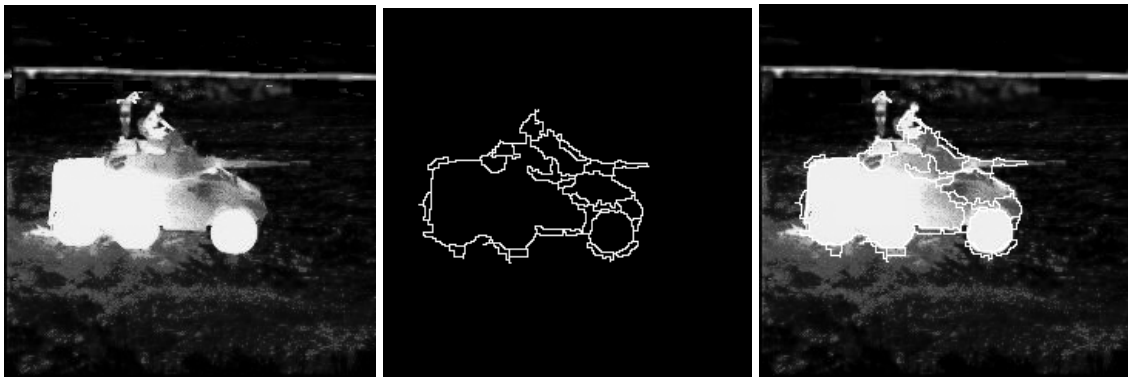
Nesse caso é observado que a extração apresenta bons resultados nos quatro exemplos não havendo alvos falsos, não importando a distância do alvo ou se este está a uma temperatura superior ou inferior ao meio ambiente, atestando a robustez do procedimento proposto para esses parâmetros.

7.3 – Solução para o problema da linha do horizonte

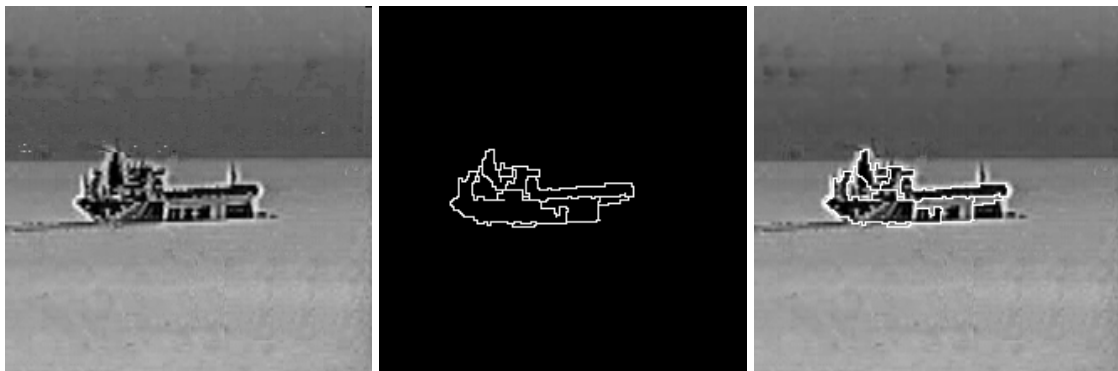
O problema da linha do horizonte é o de detectar o objeto semântico sem detectar a linha do horizonte (seção 6.2.3). Este é solucionado com eficácia com a utilização do método aqui proposto, como pode ser visto na figura 70.



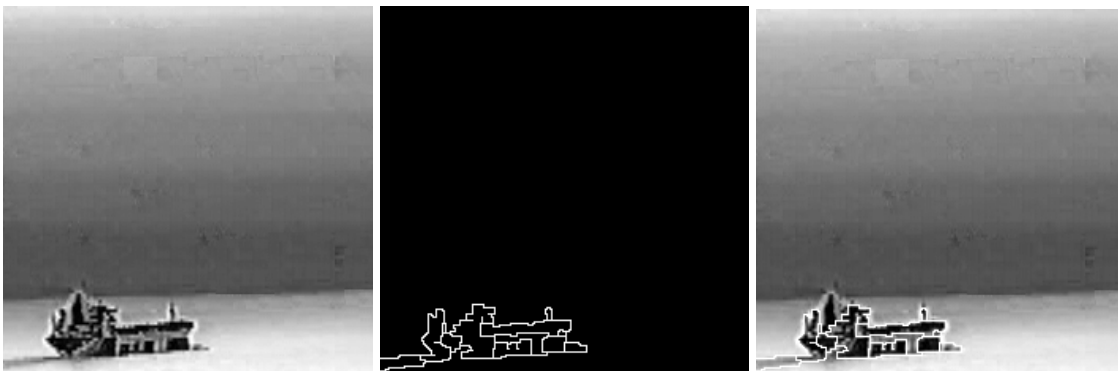
(a)



(b)



(c)

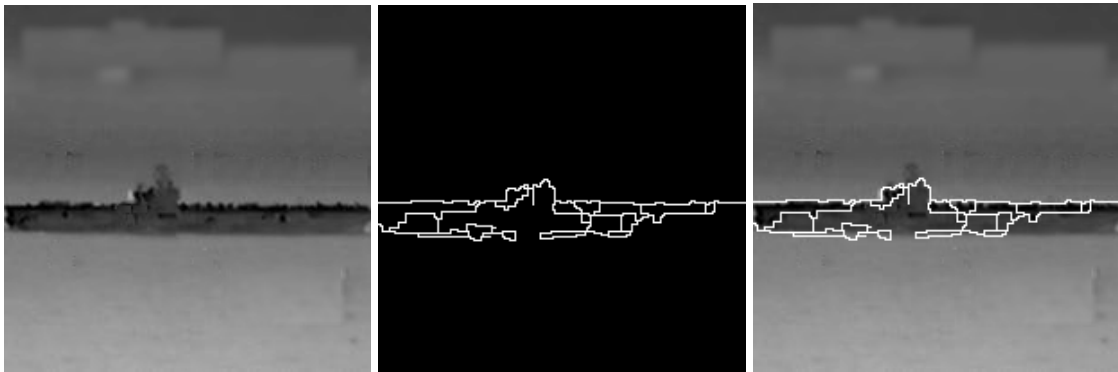


(d)

Figura 70 : Resultados da segmentação de imagens com a linha do horizonte. Da esquerda para a direita em todos os quatro exemplos: imagem original, bordas extraídas e bordas extraídas sobrepostas às imagens.(a) navio (b) carro de combate (c) navio e (d) navio.

Note que em todas as imagens apresentadas a segmentação suprime a linha do horizonte, mesmo naquelas em que o objeto semântico toca a linha do horizonte (figuras 70 (c) e (d)).

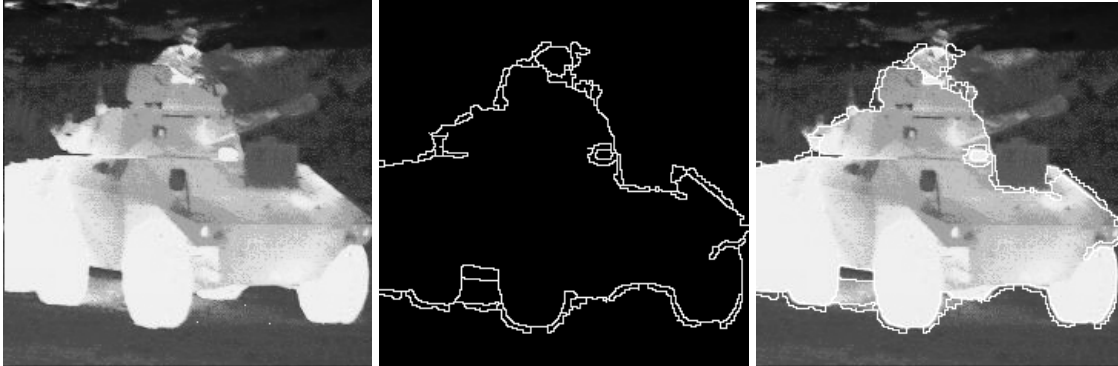
Também foram testados objetos que tocam as duas margens, ou possuem elementos que tocam as duas margens e que poderiam ser confundidos com a linha do horizonte (figura 71).



(a)



(b)



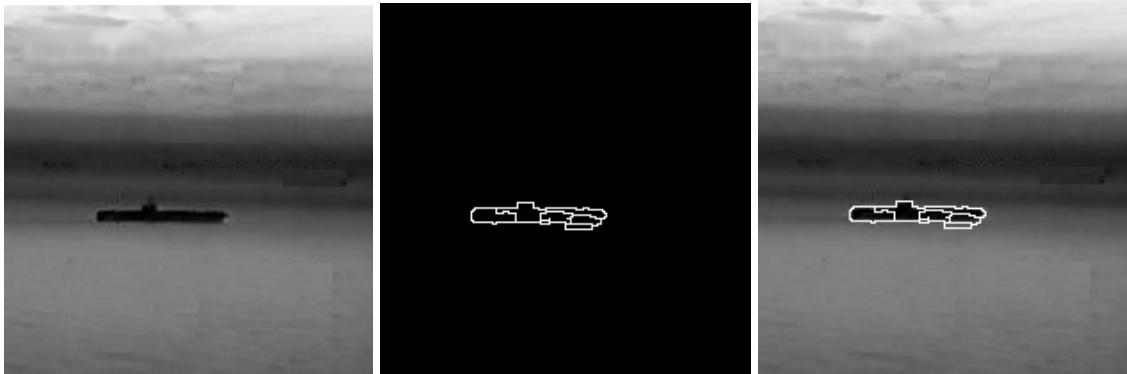
(c)

Figura 71 : Resultados da segmentação de imagens. Da esquerda para a direita em todos os quatro exemplos: imagem original, bordas extraídas e bordas extraídas sobrepostas às imagens.(a) navio (b) navio e (c) carro de combate.

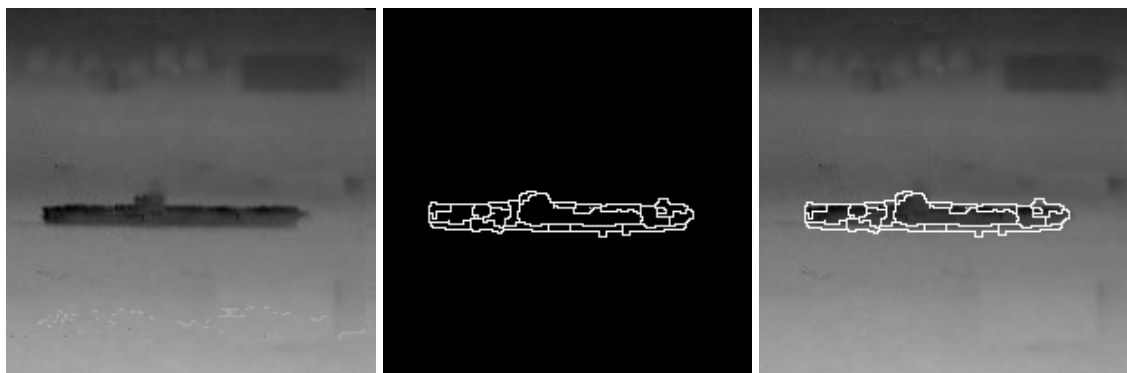
Todos os objetos são detectados comprovando assim a eficiência do procedimento adotado.

7.4 – Pequenas estruturas e reflexos

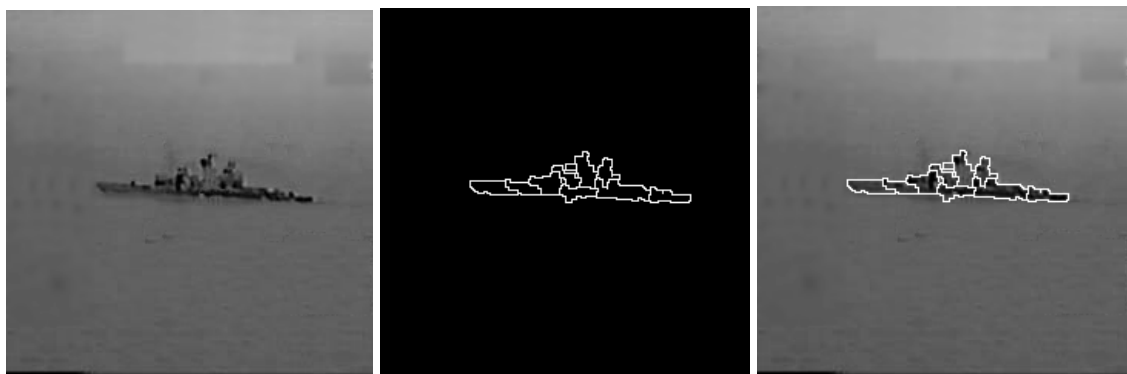
Na figura 72 pode ser visto que pequenas e finas estruturas de imagens são extraídas.



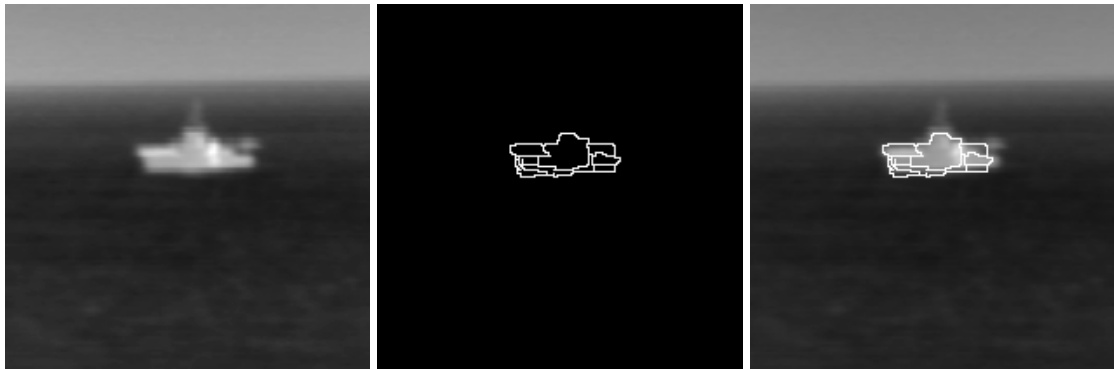
(a)



(b)



(c)

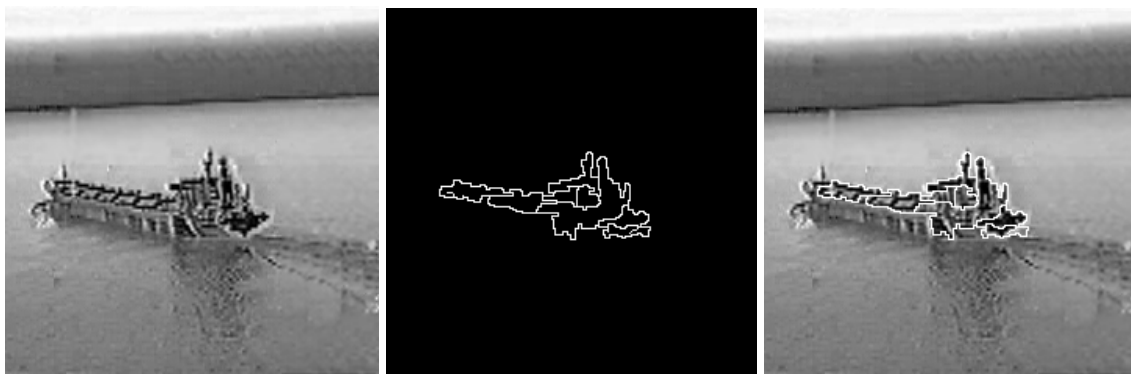


(d)

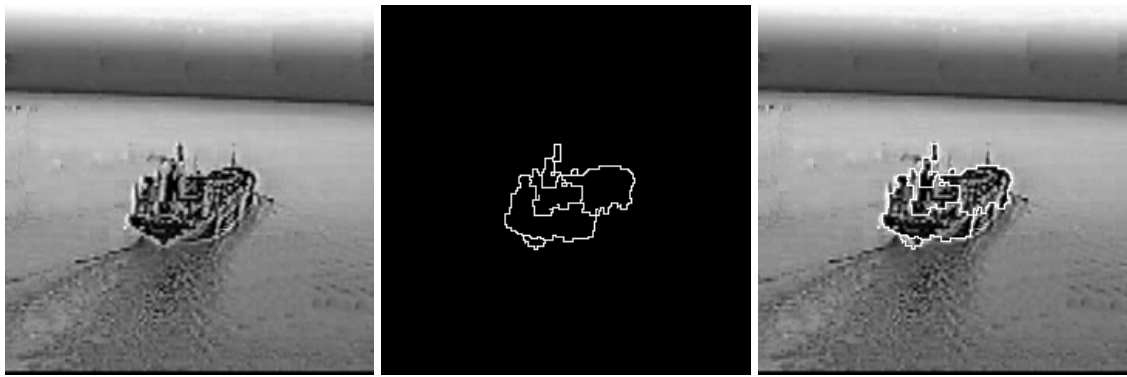
Figura 72 : Resultados da segmentação de imagens com estruturas pequenas ou finas. Da esquerda para a direita em todos os quatro exemplos: imagem original, bordas extraídas e bordas extraídas sobrepostas às imagens.(a) navio (b) navio (c) navio e (d) navio.

A detecção dessas estruturas poderia ser comprometida pelo processo de dilatação das bordas extraídas, porém o processo de esqueletonização seguido pela abertura por área, descrito na seção 6.2.4, garante a detecção correta.

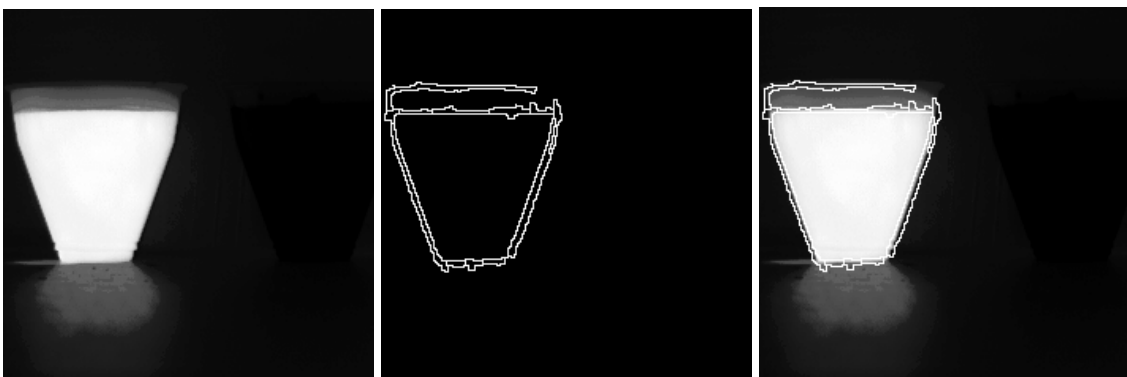
Na figura 73 pode ser observado que estruturas tais como o reflexo do navio em (a) e (b), do copo em (c) acertadamente não são detectadas pelo método proposto. Esse efeito também pode ser observado na figura 69 (c) com relação ao o reflexo do grupo e na figura 70 (a) com relação ao reflexo do navio e se deve ao valor do limiar escolhido para a validação de um ponto de máximo como módulo máximo (vide seção 5.4.2). No caso, apesar desses reflexos serem transições percebidas pela imagem de módulo não são maiores que aquele limiar.



(a)



(b)



(c)

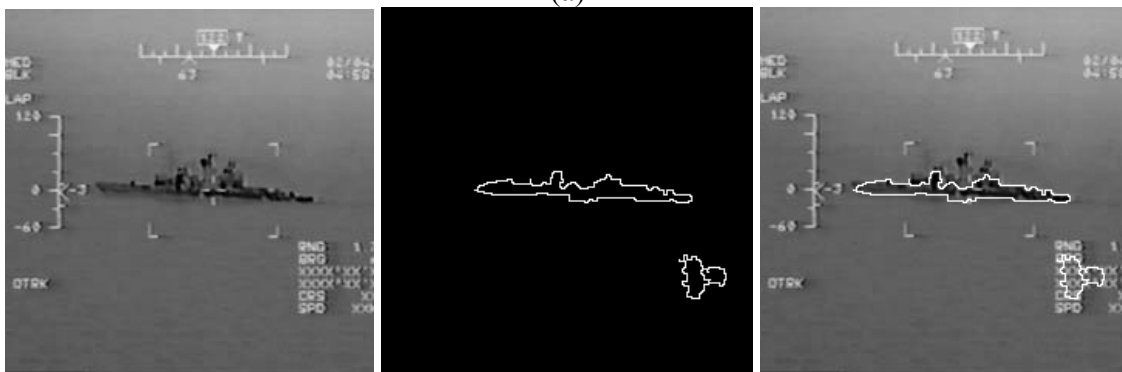
Figura 73 : Resultados da segmentação de imagens. Da esquerda para a direita em todos os quatro exemplos: imagem original, bordas extraídas e bordas extraídas sobrepostas às imagens.(a) navio (b) navio e (c) copo

7.5 – Imagens com vídeo sintético

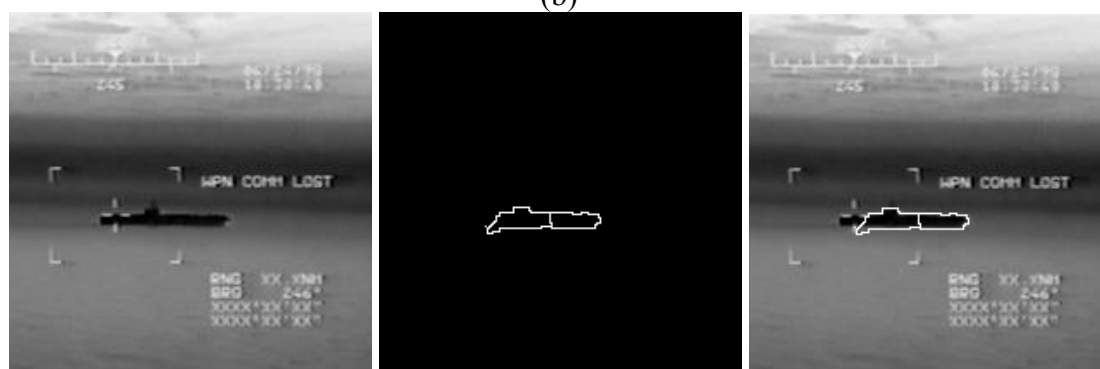
O problema de imagens com ruídos convencionais inerentes à aquisição das imagens é esgotado na detecção e construção dos marcadores. Entretanto, cameras infravermelhas relacionadas com operações de vigilância podem oferecer à extração de objetos semânticos imagens compostas com vídeo sintético. O vídeo sintético em questão compreende um conjunto de informações sobre a imagem tais como hora, temperatura de determinados pontos, entre outras. Essas informações são sempre formadas por variações bruscas e podem fornecer conjuntos fechados quando do processo de extração dos marcadores. Isso faz com que o processo de segmentação tenha que ser ajustado sob pena de aumento na detecção de alvos falsos. Uma maneira de ajuste possível é a de não executar a esqueletonização sugerida na seção 6.2.4.



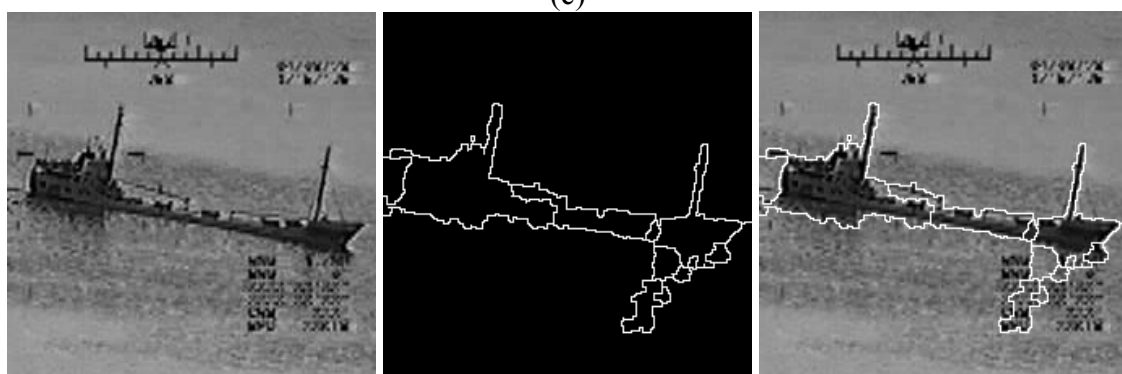
(a)



(b)



(c)



(d)

Figura 74 : Resultados da segmentação de imagens com vídeo sintético. Da esquerda para a direita em todos os quatro exemplos: imagem original, bordas extraídas e bordas extraídas sobrepostas às imagens.

Na figura 74 pode ser visto que o resultado é bom , se levarmos em conta a profunda impregnação da imagem com alvos falsos artificiais provenientes do vídeo sintético.

7.6 - Conclusão

Os resultados obtidos ao longo do conjunto de imagens de teste demonstraram a robustez e a generalidade da solução sugerida nesse método. Problemas relacionados tais como o da linha do horizonte e dos reflexos dos objetos semânticos são contornados. Para objetivos específicos de detecção pequenos ajustes ou adaptações podem ser realizadas com aumento da qualidade do resultado.

8 – CONCLUSÕES E SUGESTÕES

A área de processamento de imagens está ligada à manipulação e análise de imagens via computador. Existem, basicamente, dois grupos de aplicação para o processamento de imagens: melhoria da qualidade da imagem para interpretação humana e processamento da imagem para visão artificial. A segmentação de imagens sem supervisão está associada ao segundo grupo e é um dos mais desafiadores problemas para a visão artificial. A extração de objetos de imagens em infravermelho é de grande interesse tanto para aplicações no setor militar como no comercial, e a transformação watershed tem sido pouco explorada para a segmentação deste tipo de imagem, apesar de ser uma poderosa ferramenta morfológica.

Uma nova abordagem para a detecção automática de objetos semânticos em imagens infravermelhas foi proposta neste trabalho. Este método encontra sucesso utilizando a combinação das técnicas da transformada wavelet para extração de bordas com a morfologia matemática. Como pode ser visto ao longo dos capítulos 5 e 6 o processo é inteiramente automático, com resultados robustos. Além disso foi encontrada uma solução para o problema da linha do horizonte, que atinge, especialmente, a detecção de alvos. Outra vantagem do método é que este dispensa uma fase de filtragem preliminar da imagem antes da fase de escolha do marcador, dado que ela é realizada pelo processo da transformada wavelet módulo máximo.

Um sistema de classificação, tipicamente, tem três componentes principais: segmentação da imagem, extração do objeto e classificação do objeto. A precisão da segmentação é crucial num processo de classificação. Em nossa abordagem, a segmentação e a extração do objeto são uma coisa só. Isso foi possível graças a o conhecimento das características intrínsecas do objeto a ser extraído e à aplicação desse conhecimento no processo de construção dos marcadores.

O objetivo seminal desse trabalho estava na criação de um método eficiente e preciso para a segmentação de navios em imagens em infravermelho, fazendo parte de um projeto, desenvolvido no Instituto de Pesquisas da Marinha, de um classificador baseado em assinaturas infravermelhas de navios. O resultado da segmentação serviria de dado de entrada para um método de classificação baseado em uma rede neural [34]

treinada com uma biblioteca de assinaturas de navios em infravermelho. Essas assinaturas correspondem à imagens em infravermelho de navios tomadas a uma distância fixa e a partir de oito ângulos de observação. No decorrer do desenvolvimento do método de segmentação foi perseguido o objetivo de torná-lo o mais genérico possível sem perder a precisão para as imagens de navios. A partir dos resultados é possível a comprovação de que esses objetivos foram alcançados, já que os objetos semânticos são extraídos em qualquer posição da cena, em temperaturas acima ou abaixo do ambiente e conectadas ou não à linha do horizonte.

Como sugestão para a continuidade deste trabalho, alguns tópicos de pesquisa podem ser abordados:

- . A efetiva integração deste método de segmentação com um sistema de classificação para alvos específicos. Existe uma previsão para a integração dessa abordagem de segmentação com o método de classificação desenvolvido em [34] no segundo semestre do ano corrente.

- . A utilização desta técnica de segmentação em outros tipos de imagem que possuam como conhecimento a priori para a sua segmentação a localização da região de interesse em transientes, ou seja, nas bordas do objeto. Estão em curso estudos iniciais para a utilização do método proposto em imagens com baixa profundidade de campo, com boas perspectivas quanto ao resultado.

- . A classificação prévia da imagem de entrada por alguma técnica automática pode auxiliar numa posterior modificação adaptativa dos parâmetros, adequando a técnica proposta a outros tipos de imagem.

APÊNDICE A

Algoritmo para realização da transformada wavelet binária e sua inversa

Supondo que as wavelets $\psi^1(x,y)$ e $\psi^2(x,y)$ são separáveis e podem ser caracterizados pelos filtros discretos H , G , K e L cujos coeficientes estão descritos na tabela 1, e denominando H_p , G_p , K_p , e L_p , os filtros discretos obtidos a partir da colocação de $(2^p - 1)$ zeros entre os coeficientes dos filtros H , G , K e L , e definindo D como o filtro em que sua resposta ao impulso corresponda ao pulso de Dirac, ou seja, é igual a 1 em 0 e a 0 em qualquer outro valor, o algoritmo seguinte realiza a transformada wavelet binária de duas dimensões de uma imagem $S_1^d f$ conforme descrito em [13]. Em cada escala 2^j o algoritmo decompõe $S_{2^j}^d f$ em $S_{2^{j+1}}^d f$, $W_{2^{j+1}}^{1,d} f$ e $W_{2^{j+1}}^{2,d} f$:

```
 $j=0$   
 $while (j < J)$   
 $W_{2^{j+1}}^{1,d} = \frac{1}{\lambda_j} S_{2^j}^d f * (G_j, D)$   
 $W_{2^{j+1}}^{2,d} = \frac{1}{\lambda_j} S_{2^j}^d f * (D, G_j)$   
 $S_{2^{j+1}}^d f = S_{2^j}^d f * (H_j, H_j)$   
 $j = j + 1$   
 $end$ 
```

Assim ao convoluirmos o filtro G_j com a imagem original ou a já suavizada por H_j (dependendo da escala ou nível), na direção da abscissa x estaremos obtendo $W_{2^{j+1}}^{1,d} f$ e, ao convoluirmos na direção da ordenada y teremos $W_{2^{j+1}}^{2,d} f$ através das escalas

j. É conveniente lembrar que todos os componentes obtidos através desse processo de transformada wavelet binária possuem o mesmo tamanho da imagem original.

As provas para a validade desse algoritmo podem ser encontradas em [13].

O algoritmo de reconstrução utiliza os filtros K_p e L_p bem como o filtro \tilde{H}_j cuja função de transferência é o complexo conjugado do filtro H_j e obtém $S_1^d f$ através da recuperação, em cada escala 2^j , do sinal $S_{2^{j-1}}^d f$ a partir de $S_{2^{j+1}}^d f$, $W_{2^{j+1}}^{1,d} f$ e $W_{2^{j+1}}^{2,d} f$:

```

j = J
while (j > 0)
     $S_{2^{j+1}}^d f = \lambda_j \cdot W_{2^j}^{1,d} f * (K_{j-1}, L_{j-1}) + \lambda_j \cdot W_{2^j}^{2,d} f * (L_{j-1}, K_{j-1}) + S_{2^j}^d f * (\tilde{H}_{j-1}, \tilde{H}_{j-1})$ 
    j = j - 1
end

```

Tabela 1: Coeficientes dos filtros separáveis que realizam a transformada wavelet binária

<i>N</i>	<i>H</i>	<i>G</i>	<i>K</i>	<i>L</i>
-3			0.0078125	0.0078125
-2			0.054685	0.046875
-1	0.125		0.171875	0.1171875
0	0.375	-2.0	-0.171875	0.65625
1	0.375	2.0	-0.054685	0.1171875
2	0.125		-0.0078125	0.046875
3				0.0078125

Devido a discretização o módulo máximo da wavelet de um degrau não tem a mesma amplitude em todas as escalas como deveriam em um modelo contínuo. As constantes λ_j que são dadas na tabela 2 compensam esse problema.

Tabela 2: Coeficientes de normalização

j	λ
1	1.5
2	1.12
3	1.03
4	1.01
5	1.00
$j > 5$	1.00

Os problemas de borda são solucionados através da utilização de colunas e linhas simétricas à imagem com respeito as suas bordas. Isso deve ser levado em conta quando da execução da convolução para a obtenção precisa das linhas e colunas relevantes ao processo.

APÊNDICE B

Funções utilizadas para a implementação dos operadores morfológicos

Para a implementação do método desenvolvido nesse trabalho foi utilizado o programa MATLAB® versão 5.3 e o “toolbox” de morfologia desenvolvido pela SDC Information Systems versão 0.14 disponível no endereço eletrônico <http://www.mmorph.com>.

A primeira função utilizada é a abertura por área que realiza a escolha das linhas de borda (seção 6.2.2). A função do MATLAB é:

$Y = \mathit{mmareaopen}(F, A)$ onde:

F = Imagem binária ou em tons de cinza

A = Número de pixels conectados

Y = Imagem resultante da operação

Essa operação é ainda utilizada na redução do esqueleto (seção 6.2.4) e depois da transformação watershed (seção 6.4).

Para resolver o problema do horizonte (seção 6.2.3), primeiro é utilizada a reconstrução morfológica por dilatação onde é utilizada iterativamente a função:

$Y = \mathit{mmcdil}(F, G, B, N)$, onde:

F = Imagem binária ou em tons de cinza (imagem marcadora)

G = Imagem binária ou em tons de cinza (imagem máscara)

B = Elemento estruturante

N = Número de iterações

Y = Imagem resultante da operação

Também é utilizada para a solução do problema do horizonte a operação gradiente morfológico:

$Y = \mathit{mmgradm}(F, \mathit{BDIL}, \mathit{BERO})$, onde:

F = Imagem binária ou em tons de cinza

BDIL = Elemento estruturante para a dilatação (nesse caso é uma linha de tamanho 2 pixels encadeados)

BERO = Elemento estruturante para a erosão (nesse caso é uma linha de tamanho 2 pixels encadeados)

Y = Imagem resultante da operação

Para a finalização do marcador, primeiro é necessário a realização de uma dilatação por um elemento estruturante quadrado de tamanho 6X6 (seção 6.24):

Y = *mmdil* (F , B), onde:

F = Imagem binária ou em tons de cinza

B = Elemento estruturante que, no caso foi realizado com a função *mmsebox*(6)

Y = Imagem resultante da operação

Para a finalização do marcador também é utilizada a função que compõe o esqueleto:

Y = *mmskelm* (F), onde:

F = Imagem binária

Y = Imagem resultante da operação

O gradiente morfológico espesso que simplifica a imagem (seção 6.3) é realizada com a mesma função de gradiente morfológico descrito para o gradiente horizontal, contudo BDIL e BERO são compostos por um quadrado de 10X10. Isso garante a espessura maior das bordas da imagem de gradiente (seção 3.3).

A transformação watershed (seção 6.4) é realizada a partir do marcador e a função do MATLAB é:

Y = *mmwatershed* (F ,G, BC), onde:

F = Imagem em tons de cinza

G = Imagem binária ou em tons de cinza (imagem marcadora)

BC = Elemento estruturante que representa a conectividade da watershed (em nosso caso foi utilizado *mmsebox*)

Y = Imagem resultante da operação

BIBLIOGRAFIA

- [1] JAIN, ANIL K., *“Fundamentals of Digital Image Processing”*, Prentice-Hall, 1989.
- [2] MILLER, JOHN LESTER, *“Principles of Infrared Technology”*, Chapman and Hall, 1994.
- [3] KAMGAR-PARSI, BEHROOZ e KAMGAR-PARSI, BEHZAD, *“Improved Image Thresholding for Object Extraction in IR Images”*, Proceedings of International Conference on Image Processing 2001, pp.758-761, Outubro 2001.
- [4] LEE, J. S. e YANG, M.C.K., *“Threshold Selection using Estimates from Truncate Normal Distribution”*, IEEE Transactions on Systems, Man and Cybernetics, Vol. 19, pp. 422-429, Março 1989.
- [5] DEER, S. Z. e CHELLAPA, R., *“Probe-based Automatic Target Recognition in Infrared Imagery”*, IEEE Transactions on Image Processing, Vol. 6, pp. 92-102, Janeiro 1997.
- [6] SOILLE, P., *“Morphological Image Analysis - Principle and Applications”*, Springer, 1999.
- [7] FACON, J., *“Morfologia Matemática: Teoria e Exemplos”*, Editor Jacques Facon, 1996.
- [8] GIARDINA, C. R. e DOUGHERTY E. R., *“Morphological Methods in Image and Signal Processing”*, Prentice-Hall, 1988.
- [9] SERRA J., *“Image Analysis and Mathematical Morphology”*, New York: Academic Press, 1983.
- [10] BEUCHER, S., *“Segmentation Tools in Mathematical Morphology”*, SPIE Image Algebra and Morphological Image Processing, Vol. 1350, pp. 70-84, 1990.
- [11] MEYER, F., BEUCHER, S., *“Morphological Segmentation”* Journal of Visual Communication and Image Representation, pp. 21-46, Setembro 1990.
- [12] MALLAT, S., HWANG, W. L., *“Singularity Detection and Processing with Wavelets”*, IEEE Transactions on Information Theory, Vol. 38, pp. 617-643, Março 1992.

- [13] MALLAT, S., ZHONG, S., ***“Characterization of Signal from Multiscale Edges”***, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 710-732, Julho 1992.
- [14] GARG A., ***“Object Extraction from Infrared Images”***, Dissertação de MSc., University of South Australia, School of Electronic Engineering, Abril 1996.
- [15] BEUCHER, S. e LANTUEJOUL, C., ***“Use of Watersheds in Contour Detection”***, Int. Workshop on Image Processing, Real-time Edge and Motion Detection/Estimation, CCETT/IRISA, Rennes, França, Setembro 1979.
- [16] MEYER, F., ***“Integrals, Gradients and Watershed Lines”***, Proceedings of Mathematical Morphology and its Applications to Signal Processing, pp. 70-75, Barcelona, Maio 1993.
- [17] NAJMAN, L. e SCHMITT, M., ***“Definition and some Properties of the Watershed of a Continuous Function”***, Proceedings of Mathematical Morphology and its applications to Signal Processing, pp. 75-81, Barcelona, Maio 1993.
- [18] VINCENT L., e SOILLE P., ***“Watersheds in Digital Spaces: An Efficient Algorithm based on Immersion Simulations”***, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 583-598, Vol. 13, No. 6, Junho 1991.
- [19] GU, CHUANG e LEE, MING-CHIEH, ***“Semiautomatic Segmentation and Tracking of Semantic Video Objects”***, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 5, Setembro 1998.
- [20] GAO, H.; SIU W. e HOU C., ***“Improved Techniques for Automatic Image Segmentation”***, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 12, pp.1273-1280, Dezembro 2001.
- [21] HADDON, J. F. e BOYCE, J. F., ***“Image Segmentation by Unifying Region and Boundary Information”***, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 929-948, Vol. 12, No.10, Outubro de 1990.
- [22] HADDON, J. F. e BOYCE, J. F., ***“Unification of Image Segmentation and Edge Detection”***, IEE Proceedings, pp. 129-135, Vol. 137, No. 3, Junho 1990.
- [23] GAUCH, J. M., ***“Image Segmentation and Analysis via Multiscale Gradient Watershed Hierarchies”***, IEEE Transactions on Image Processing, pp. 69-79, Vol. 8, No. 1, Janeiro 1999.
- [24] RIOUL, O. e VETTERLI, M., ***“Wavelets and Signal Processing”***, IEEE Signal Processing Magazine, Vol. 8, pp. 14-38, Outubro 1991.

- [25] MALLAT, S., “*A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, pp. 674 - 693, Julho 1989.
- [26] MALLAT, S., “*Multifrequency Channel Decompositions of Images and Wavelet Models*”, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 37, pp. 2091-2110, Dezembro 1989.
- [27] STRANG, G. e NGUYEN, T., “*Wavelets and Filter Banks*”, Wellesley-Cambridge Press, 1996.
- [28] DAUBECHIES, I., “*The Wavelet Transform, Time-Frequency Localization and Signal Analysis*”, IEEE Transactions on Information Theory, Vol. 36, No. 5, pp. 961-1005, Setembro 1990.
- [29] MALLAT, S., “*A Wavelet Tour of Signal Processing*”, Academic Press, 1998.
- [30] MACIEL, B. C., “*Detecção de Arestas em Multiresolução e Reconstrução de Imagens a partir destas*”, Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática, Fevereiro 2000.
- [31] NEVES, S. R., “*Uma Abordagem para Codificação de Imagens a partir das Bordas e Texturas Via Transformada Wavelet*”, Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 1997.
- [32] MALLAT, S., FROMENT, J., “*Second Generation Compact Image Coding with Wavelets*” em Wavelets - A Tutorial in Theory and Applications, C. Chui, Ed., New York Academic Press, pp. 655-678, Janeiro 1993.
- [33] CANNY, J., “*A Computational Approach to Edge Detection*”, IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 8, pp. 679-698, 1986.
- [34] ALVES, J. A., “*Recognition of Ship Types from a Infrared Image using Moment Invariants and Neural Networks*”, Dissertação de Mestrado, Naval Postgraduate School, Monterey, Califórnia, EUA, 2001.