

# Segmentação de Alvos em Imagens Infravermelhas Utilizando Watersheds, Wavelets e Morfologia Matemática

Sergio R. Neves<sup>1</sup>, Gelson V. Mendonça<sup>2</sup>, Eduardo A. B. da Silva<sup>2</sup>

<sup>1</sup>Instituto de Pesquisas da Marinha, Grupo de Guerra Eletrônica, Rua Ipiru, 2- CEP: 21930-090 - Rio de Janeiro- RJ-Brasil

<sup>2</sup>Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica, Caixa Postal 68504 – CEP: 21945-970 – Rio de Janeiro – RJ – Brasil

**Resumo** — Para um sistema de classificação de alvos é fundamental que se possa extrair características precisas destes, com o objetivo de ter-se o maior número de acertos possível. Uma imagem infravermelha pode ser considerada como um mapeamento em duas dimensões da radiação térmica de uma cena, e apresenta como uma característica importante para extração de objetos, as bordas desses elementos da cena. Este trabalho descreve um método automático para a segmentação de imagens infravermelhas utilizando morfologia matemática e a transformação watershed a partir de um marcador que usa a transformada wavelet módulo máximo, com resultados precisos e com baixo número de detecções falsas.

**Palavras-chaves** — Segmentação, infravermelho, morfologia, watersheds, wavelets.

## I. INTRODUÇÃO

Por segmentação de uma imagem, são compreendidas tarefas tais como a extração de objetos de interesse de uma determinada cena ou a fragmentação da imagem em regiões de textura semelhante ou aproximada. Para a extração de objetos contidos numa imagem, é essencial que se possa separar esses objetos de interesse do “resto”. Essa distinção realizada via computador, entretanto, pode conduzir a processos complexos, por várias razões: a) a definição de objeto de interesse nem sempre é clara, b) diversas técnicas de segmentação são bastante sensíveis a ruído, c) os mecanismos de segmentação são normalmente específicos e limitados. A separação da imagem em regiões também é um processo sujeito a ruído e que procura simplificar a imagem para tarefas tais como a compressão ou análise.

Ao pensar no “mecanismo de segmentação” humano, no sentido da extração de objetos semânticos, 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 voltada, principalmente, para o jogador que está com a bola; quando lemos, “separamos” da página a palavra na sequência que se está lendo. Em outras palavras, o cérebro precisa de algum conhecimento a priori para realizar a tarefa de segmentação. Para realizar a segmentação via computador, também é necessário estabelecer uma motivação ou regra a priori, para o que se deseja separar de uma cena. Em uma cena, a cor, formato ou tamanho também são informações representativas para a segmentação.

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

Uma imagem infravermelha 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 se juntam a outros introduzidos artificialmente, especialmente por câmeras infravermelhas militares, como vídeo sintético. Todos esses ruídos podem conduzir à detecção de alvos falsos.

O conhecimento a priori para a segmentação de imagens infravermelhas, especialmente nas tarefas de vigilância e identificação de alvos, é o seguinte: o objeto de interesse está sempre a uma temperatura acima ou abaixo da temperatura ambiente. As regiões de interesse estão, então, caracterizadas por transientes como picos e bordas.

O método que é apresentado aqui para segmentação de imagens infravermelhas, com o objetivo de extração de objetos semânticos, utiliza como modelo o paradigma de Beucher-Meyer [1] que, basicamente, consiste em duas etapas: a primeira que seleciona marcadores para os objetos de interesse (considerada como etapa “inteligente”) e um passo que realiza a transformada watershed para a extração destes. Na etapa de seleção dos marcadores é proposta a utilização da transformada wavelet módulo máximo [2] [3] combinada com operadores morfológicos, o que, como será visto, funciona como uma pré-segmentação e auxilia na retirada de ruídos. Esse marcador oferece à transformada watershed uma região aproximada da localização do contorno do objeto de interesse, que esta transformada vai extrair com precisão.

Esse artigo está organizado da seguinte maneira: na seção II é introduzida a segmentação morfológica, bem como alguns de seus conceitos mais importantes; na seção III tem-se a transformada wavelet módulo máximo que será útil para a realização do marcador para o método proposto; na seção IV

é descrito o método proposto; na seção V são mostrados os resultados e na seção VI temos as conclusões e a sua aplicabilidade em outros tipos de imagem.

## II SEGMENTAÇÃO MORFOLÓGICA

### A. Introdução

A segmentação quando tratada do ponto de vista da morfologia matemática, envolve essencialmente duas abordagens, dependendo de seu objetivo: a detecção de bordas ou de regiões. Em ambos os casos a ferramenta morfológica chave para sua realização é a transformada watershed [4] [5]

### B. Transformada 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 esses “vales” são 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” [6]. Existem vários algoritmos desenvolvidos para sua realização, mas muitas vezes estes conduzem à sobre-segmentação, que consiste na divisão da imagem de maneira exagerada, devido ao fato de mínimos locais não significativos originarem bacias de segmentação. A forma de evitar, ou diminuir esse efeito é a escolha de marcadores para a região onde se encontra o objeto de interesse. A simplificação da imagem a partir de algum processo de filtragem ou redução de mínimos locais e, conseqüentemente, bacias de segmentação, também podem conduzir a uma diminuição do efeito de sobre-segmentação.

## III. TRANSFORMADA WAVELET MÓDULO MÁXIMO

A transformada wavelet módulo máximo [2] [3] que é utilizada para a detecção de bordas e realização do marcador, pode ser considerada como pertencente à classe dos detectores baseados no gradiente. A partir da escolha de uma função de “suavização”  $\theta(x, y)$  que possui a resposta ao impulso de um filtro passa-baixas, as funções wavelet podem ser definidas como as derivadas parciais de  $\theta(x, y)$  ao longo de  $x$  e  $y$ :

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

Para qualquer função  $f(x, y) \in L^2(R^2)$  a transformada wavelet binária terá dois componentes fornecidos pelas convoluções:

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

É possível realizar essa decomposição wavelet de modo multiescalar: para cada escala ou nível, teremos que realizar a filtragem passa-baixas  $\theta(x, y)$  e utilizar essa imagem “suavizada” para extrair os componentes  $W^1 f$  e  $W^2 f$  do nível seguinte. Para tanto é necessário, também, dilatar  $\theta, \psi^1$  e  $\psi^2$  de um fator de escala  $2^j$  para obter os operadores de cada nível:

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

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

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

É possível, então construir uma imagem de módulo e uma imagem de fase para cada nível da transformada:

$$\begin{aligned} M_{2^j} f(x, y) &= \sqrt{|W_{2^j}^1 f(x, y)|^2 + |W_{2^j}^2 f(x, y)|^2} \\ A_{2^j} f(x, y) &= \arctan \left| \frac{W_{2^j}^2 f(x, y)}{W_{2^j}^1 f(x, y)} \right| \end{aligned}$$

O módulo máximo pode ser extraído através dos pixels que, na imagem de módulo, são maiores que seus vizinhos. Esses pixels compõem as linhas de borda da imagem.

## IV. MÉTODO PROPOSTO

O método proposto pode ser visto no diagrama em blocos da Fig. 1. Pode ser considerado que este possui três passos principais: extração dos objetos semânticos e determinação do marcador, simplificação da imagem e transformação watershed e filtragem final.

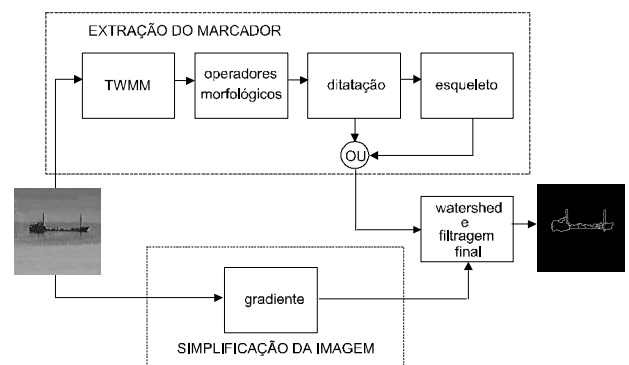


Fig. 1: Diagrama em blocos do método proposto.

A. Primeiro passo: detectando os objetos semânticos e extraindo seus marcadores

1) Detectando as bordas: O primeiro desafio para um método de segmentação automático é determinar onde estão os objetos semânticos em uma cena. Essa tarefa deve levar em consideração um conhecimento a priori da imagem. O conhecimento a priori a ser utilizado na extração de objetos semânticos numa imagem infravermelha é a localização

destes objetos em regiões de transição como bordas. Infelizmente, na maioria das aplicações de segmentação de imagens infravermelhas, o nível de ruído é alto, o que pode conduzir a falsa detecção de alvos. O detector de bordas escolhido para o nosso método foi o baseado na transformada wavelet módulo máximo [2] [3] (veja a seção III). O uso deste tipo de detector apresenta duas vantagens principais em relação à influência de ruído: a filtragem passa-baixa intrínseca ao processamento da transformada wavelet módulo máximo e a possibilidade de escolha das linhas mais relevantes usando a informação de multiresolução.

Diferente dos métodos em [2] e [3], onde os módulos máximos são pesquisados ao longo da direção dada por  $A_2^j f(x, y)$ , foi executada uma busca nas direções horizontal e vertical. O motivo dessa busca bidirecional é o de se obter as bordas dos objetos semânticos o mais próximo possível de contornos fechados. Um limiar foi determinado para empiricamente validar o módulo escolhido por essa busca como um módulo máximo. Como pode ser visto na fig. 2, esse procedimento certifica que só teremos módulos relevantes nos mapas de borda ao longo da escala.

Devido ao fato que, acima do quarto nível de escala, os mapas de borda obtidos não serem de boa qualidade com relação ao contorno do objeto semântico (vide figura 2), nós decidimos usar somente os três primeiros níveis da transformada wavelet binária para obter os mapas de borda desejados. A fim de detectar os alvos da imagem, primeiro foi implementada uma transformada wavelet binária [2] [3], utilizando a arquitetura proposta por Zhong et al [3]. Com esse procedimento foram obtidos os componentes wavelet ao longo das escalas. Com esses dois componentes a imagem de módulos é calculada (veja a seção III). Então essa imagem de módulos é utilizada na pesquisa dos módulos máximos e extração das bordas.

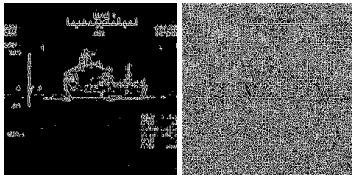


Fig. 2 : Mapas de módulos máximos realizados com a comparação com limiar (esquerda) e sem a comparação (direita) para a transformada wavelet módulo máximo realizada para nível 1

O marcador pode ser interpretado como a região limitada onde a transformada watershed será executada. Isso é realizado através da operação morfológica conhecida como imposição de mínimo [4]. Para garantir que as linhas de borda dos objetos semânticos serão incluídas nessas regiões onde será executada a transformação watershed, é importante escolher o mapa de bordas adequado. Nossa procura é por um mapa de bordas com as seguintes características: a) pequeno número de falsos alvos; b) o conjunto de bordas do objeto semântico ser tão próximo quanto possível de um contorno fechado; c) baixo nível de ruído. Analisando os mapas de borda do nosso conjunto de imagens de teste, concluímos que os melhores resultados foram os obtidos a partir dos mapas de borda referentes ao segundo e terceiro nível da decomposição wavelet. Então os mapas de borda correspondentes ao segundo e terceiro níveis da decomposição pela transformada wavelet são combinados através de uma operação lógica OU. Esse procedimento tem a vantagem de aumentar o número de contornos fechados. Na

fig. 3 temos o diagrama em blocos para o método de detecção de bordas proposto.

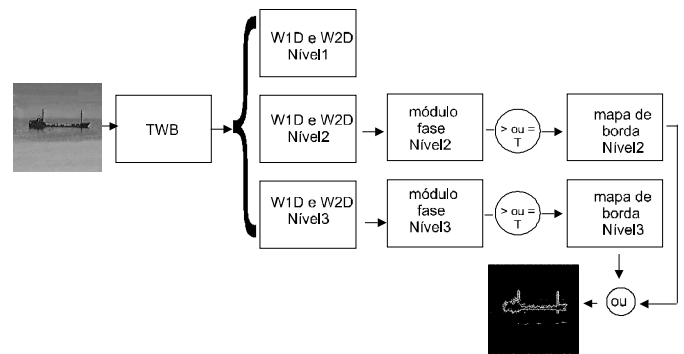


Fig. 3 : Diagrama em blocos da detecção das bordas do objeto semântico. TWB é a transformada wavelet binária. W1D e W2D são os dois dos componentes dessa transformada nos múltiplos níveis. Os mapas de borda são obtidos através da busca bidimensional descrita e comparação com o limiar determinado empiricamente.

2) *Escolhendo as bordas "mais importantes"*: Na maioria dos casos, as fronteiras dos objetos mais significativos em uma imagem geram as linhas de borda maiores. Baseado nisso foi assumido que linhas de borda consideradas abaixo de um determinado número de pixels encadeados correspondem a estruturas irrelevantes. Para eliminar linhas de borda menores que um tamanho determinado e manter as outras intactas, o operador morfológico abertura por área [4] foi utilizado na imagem resultante do processo descrito pelo diagrama em blocos da fig. 3. A abertura por área compara os objetos conectados de uma imagem com seu elemento estruturante e esses objetos são descartados se têm área menor do que este elemento estruturante. A área do elemento estruturante foi determinada empiricamente; os resultados de nossas simulações mostraram que uma área de 100 pixels foi adequada para todo o nosso conjunto de testes. Na fig. 4 é possível observar muitos dos alvos falsos foram eliminados depois da abertura por área.

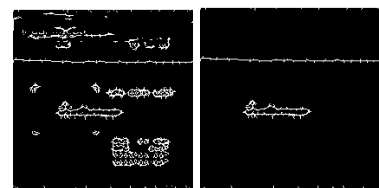


Fig. 4: Mapas de borda antes e depois do filtro de abertura por área.

3) *Resolvendo o problema da linha do horizonte*: Em imagens infravermelhas panorâmicas, pode ocorrer um efeito indesejável: quando aparece, a linha do horizonte tem o tamanho suficiente para ser detectada mesmo depois da execução do filtro de abertura por área. Então, ela compõe um conjunto fechado com as margens da imagem sendo mantida pela transformação watershed, como pode ser visto na fig. 5. Para efeito de extração de alvos, esse efeito é indesejável. A fim de eliminar as linhas do horizonte, um procedimento baseado em morfologia matemática foi desenvolvido.

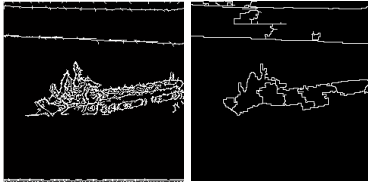


Fig. 5: Esquerda: mapa de bordas escolhido. Direita: segmentação final (após a transformação watershed).

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 [4], tendo como imagem máscara a imagem pre-segmentada. 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. Permanecem como resultado desse processo somente as linhas que tocam as duas margens laterais da imagem simultaneamente. Utilizando uma subtração elas são retiradas da imagem de entrada (veja a fig. 6).

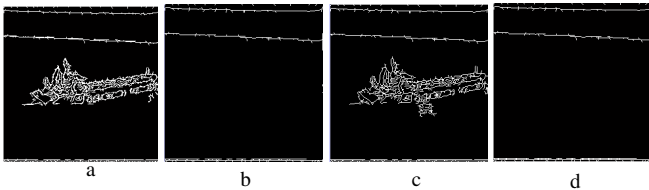


Fig. 6 : Retirada de objetos que tocam simultaneamente as duas margens da imagem imagem máscara (a), reconstrução por dilatação com uma imagem marcadora (semente) constituída pelas duas colunas mais próximas a margem esquerda (b), reconstrução por dilatação com uma imagem marcadora (semente) constituída pelas duas colunas mais próximas a margem direita (c) e E lógico dessas duas reconstruções.(d).

Infelizmente esse procedimento tem a desvantagem de remover qualquer objeto que toque a linha do horizonte ou as duas margens (veja a fig. 7).

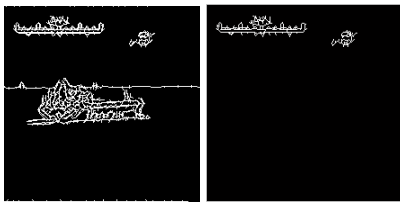


Fig. 7: Problema da linha do horizonte: mapa de bordas em que o objeto semântico toca a linha do horizonte (esquerda) e resultado do E lógico entre as duas reconstruções descritas (direita).

Esse problema foi evitado através do uso de um gradiente direcional [4] na imagem de entrada. Esse gradiente direcional tem como elemento estruturante uma linha horizontal. Através disso é possível preservar elementos que tocam o horizonte ou as duas margens mas não são horizontais (veja a fig. 8).

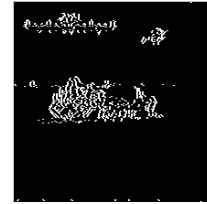


Fig 8 : Gradiente horizontal como parte do processo para a solução do problema da linha do horizonte.

O resultado desejado é obtido através de uma operação OU lógico dos resultados da operação de remoção da linha do horizonte e esse gradiente direcional.

4) *Finalizando o marcador*: Até esse ponto obtivemos uma imagem binária que pode ser considerada como uma pré-segmentação. Essa imagem binária vai ser dilatada por um objeto estruturante que é um quadrado de dimensões 7X7 pixels. O tamanho desse elemento estruturante foi determinado experimentalmente, a partir de testes realizados como conjunto de imagens de teste. A dilatação dessa imagem tem o propósito de deixar os marcadores mais espessos a fim de garantir que o marcador contenha todas as linhas de borda dos objetos semânticos.

O processamento da transformação watershed considera como um marcador válido um objeto, em uma imagem binária, que possua, pelo menos, um pixel branco dentro de uma região composta de pixels pretos. Para garantir que mesmo marcadores de pequenos objetos tenham marcadores válidos, uma esqueletonização [4] é executada nessa imagem dilatada. Depois esse esqueleto e a imagem dilatada são combinados com um E lógico (veja a fig. 9).

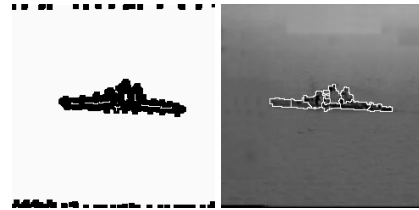


Fig. 9: Marcador (esquerda) e segmentação final (direita).

## B. Segundo passo: simplificação da imagem original

Antes de realizar a transformação watershed, a simplificação da imagem original é necessária. Várias técnicas foram testadas com esse intuito, incluindo limiarização em oito níveis, filtragem passa-baixas e gradiente morfológico. Os melhores resultados foram obtidos com o uso do gradiente morfológico, devido ao fato deste realçar as bordas do objeto semântico.

## C. Último passo: a transformação watershed

Uma vez que o marcador foi determinado e a imagem original foi simplificada, a transformação watershed pode ser executada. Como é necessário evitar a sobre-segmentação, primeiramente usamos o conjunto de marcadores para realizar a imposição de mínimo [4]. Depois as fronteiras dos objetos são obtidas através da computação das watersheds da imagem. Depois da transformação watershed, outra abertura por área é executada a fim de eliminar pequenos alvos falsos

que tenham permanecido depois de todo o processo de segmentação.

## V. RESULTADOS

Um conjunto bastante diverso de imagens em infravermelho foi segmentado com o uso deste método proposto a fim de avaliar sua eficiência. Essas imagens foram obtidas através de uma câmera AN/AAS-44V “Forward-Looking Infrared” (FLIR), montada no nariz de um helicóptero SH-60B e de uma câmera 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.

Pode ser visto na fig. 10 que o método de segmentação apresenta resultados com bastante precisão. O problema da linha do horizonte foi eliminado como pode ser visto nas fig. 10(b), 10(d) e 10(e): a linha do horizonte não é segmentada enquanto objetos que tocam as duas margens laterais da imagem são segmentados (vide fig. 10(f)). Observe que apesar de algumas imagens de navio possuírem estruturas muito finas, como a da figura 10(a), essas estruturas são extraídas com sucesso graças a esqueletonização descrita. Observe, na figura 10(b), que o método proposto tem sucesso em não detectar o reflexo do navio na superfície da água. Isso está relacionado com o limiar escolhido para validar um máximo local como um módulo máximo. Apesar desses reflexos serem transições que podem ser detectadas pela imagem de módulo, elas não são maiores do que aquele limiar. Os bons resultados mostrados na fig. 10 são consistentes com os resultados oferecidos por todas as imagens de nosso conjunto de imagens de teste, atestando, assim, a robustez do processo proposto.

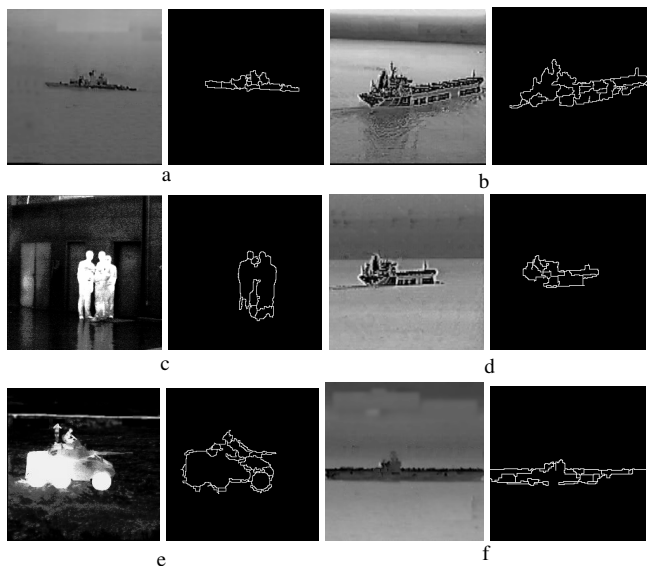


Fig. 10 (a) (b) (c) (d) (e) (f): Esquerda: imagem original. Direita: imagem segmentada.

## VI CONCLUSÃO

Versões anteriores desse método foram publicados em [7] e [8]. Esse método de segmentação faz parte de um sistema de classificação de assinaturas em infravermelho de navios em desenvolvimento no Instituto de Pesquisas da Marinha. Para esse sistema, classificadores baseados em redes neurais e em

curvas principais vem sendo testados e avaliados integrados a este método de segmentação em busca do melhor resultado. Recentemente, baseado em resultados de métodos de segmentação que utilizam wavelets e watersheds [9] [10], o método proposto vem sendo testado, com pequenas modificações, em imagens digitais com o objetivo de separação em regiões homogêneas. Os primeiros resultados tem sido bastante promissores, especialmente para imagens impregnadas de ruído, como pode ser visto na fig. 11.

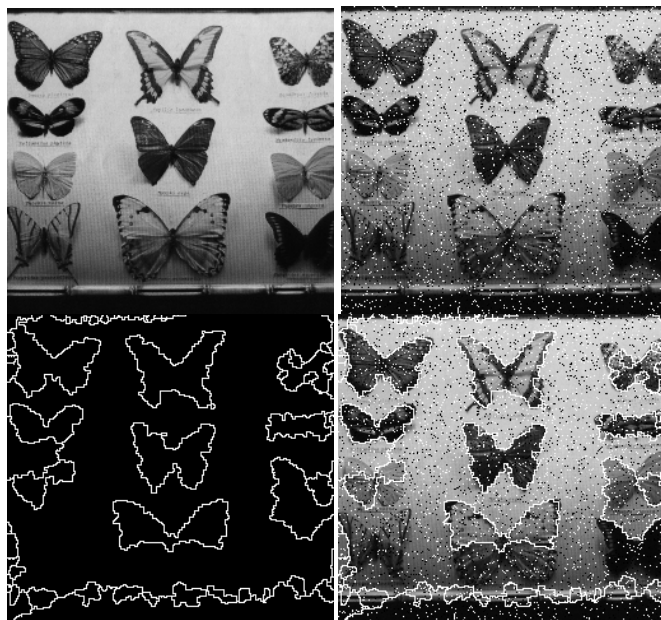


Fig. 11: Linha superior: Imagem original e imagem impregnada de ruído. Linha inferior segmentação realizada e segmentação sobreposta à imagem segmentada.

## REFERÊNCIAS

- [1] F. Meyer e S. Beucher, “Morphological Segmentation”, J. of Visual Communication and Image Representation, pp. 21-46, 1990.
- [2] S. Mallat. e W. L. Hwang, “Singularity Detection and Processing with Wavelets”, IEEE Transactions on Information Theory, Vol. 38, pp. 617-643, March 1992.
- [3] S. Mallat e S. Zhong, “Characterization of Signal from Multiscale Edges”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 710-732, July 1992.
- [4] P. Soille, Morphological Image Analysis – Principle and Applications, Springer Ed., Germany, 1999.
- [5] J. Serra., “Image Analysis and Mathematical Morphology”, New York: Academic Press, 1983.
- [6] S. Beucher e C. Lantuejoul, “Use of watersheds in contour detection,” Int. Workshop on Image Processing, real-time edge and motion detection/estimation, CCETT/IRISA, Rennes, France, Sept. 79.
- [7] Neves, S. R.; Silva E. A. B. e Vieira G. V., “Wavelet-Watershed Automatic Infrared Image Segmentation Method”, Electronic Letters, Vol. 39, Issue 12, pp. 903-904, June 2003.
- [8] Neves, S. R.; Silva E. A. B. e Vieira G. V., “Método Automático de Segmentação de Alvos em Infravermelho utilizando Wavelets e Watersheds”, Pesquisa Naval, Vol. 17, 2004
- [9] Chai Y., Gao L. and Lu S., “Wavelet-based Watershed for Image Segmentation Algorithm” , Proceedings of the World Congress on Intelligent Control and Automation, pp. 9595-9599, Dalian, China, June 21-23, 2006.
- [10] Jung, C. R. “Combining wavelets and watersheds for robust multiscale image segmentation”, Image and Vision Computing 25 (2007) pp. 24-33.