

**DAVID MENOTI**

**SEGMENTAÇÃO DE ENVELOPES POSTAIS  
PARA LOCALIZAÇÃO DO BLOCO  
ENDEREÇO:  
UMA ABORDAGEM BASEADA EM  
SELEÇÃO DE CARACTERÍSTICAS NO ESPAÇO *WAVELET***

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná, como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

**CURITIBA**

**2003**

**DAVID MENOTI**

**SEGMENTAÇÃO DE ENVELOPES POSTAIS  
PARA LOCALIZAÇÃO DO BLOCO  
ENDEREÇO:  
UMA ABORDAGEM BASEADA EM  
SELEÇÃO DE CARACTERÍSTICAS NO ESPAÇO *WAVELET***

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná, como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

Área de Concentração: Metodologia e Técnicas de Computação

**Orientador:** Prof. Dr. Jacques Facon

**Co-Orientador:** Prof. Dr. Díbio Leandro Borges

**Co-Orientador:** Prof. Dr. Alceu de Souza Britto Júnior

**CURITIBA**

**2003**

Menoti, David

Segmentação de Envelopes Postais para Localização do Bloco Endereço: uma abordagem baseada em seleção de características no espaço *Wavelet*. Curitiba, 2003. 94p.

Dissertação – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada.

1. Segmentação. 2. Envelopes Postais. 3. Seleção de Características. 4. *Wavelet*. 5. Bloco Endereço. I. Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Informática Aplicada.

Dedico este trabalho a meus pais Marisa e João.

Dois grandes guerreiros e heróis.

## **Agradecimentos**

Primeiramente a Deus, por permitir que eu perceba que existo.

Aos meus pais, Marisa e João, e familiares que sempre me incentivam a continuar lutando pelos meus ideais.

Gostaria de agradecer ao Professor Orientador Jacques pela orientação, paciência, e amizade desenvolvida durante o mestrado que já se estende desde a iniciação científica na graduação. Aos Professores Co-orientadores Díbio e Alceu pela ajuda na fundamentação teórica do trabalho, por questionamentos e contribuições construtivas e também pelo laço de amizade construído. Ao aluno de iniciação científica Klaus Fuchs, que me auxiliou no processo de segmentação manual da Base de Imagens.

Aos financiadores: a Pró-Reitoria de Pesquisa e Pós-Graduação (PRPPG-PUCPR) e ao Instituto de Ciências Exatas e de Tecnologia (ICET-PUCPR) que tornaram possível a execução de todo o mestrado em tempo integral e dedicação exclusiva; a Fundação Araucária que forneceu equipamento de ponta para o desenvolvimento da dissertação.

Aos meus Colegas de estudos, Professores, e Funcionários do PPGIA.

Aos meus tantos amigos de mesa de Bar. Amém.

# Sumário

Lista de Figuras .....	viii
Lista de Tabelas .....	xiii
Lista de Símbolos .....	xv
Resumo .....	xvi
Abstract.....	xvii
1. Introdução.....	1
1.1    Objetivos.....	5
1.2    Justificativas .....	6
1.3    Contribuições.....	7
1.4    Organização da Dissertação.....	7
2. Trabalhos Relacionados.....	8
3. Método Proposto .....	16
3.1    Decomposição <i>Wavelet</i> .....	18
3.2    Identificação dos Pontos Salientes .....	22
3.3    Rotulação Controlada das Janelas de Pontos Salientes .....	27
3.4    Projeção Reversa dos Pontos Salientes Seleccionados .....	32
3.5    Perseguição de Contorno através dos Pontos Salientes.....	34
3.6    Comentários Finais .....	41
4. Experimentos.....	43
4.1    Abordagem " <i>ground truth</i> " .....	43
4.2    As Bases de Imagens .....	45
4.2.1    Base de imagens reais de envelopes postais brasileiros .....	46
4.2.2    Peculiaridades das imagens de envelopes postais .....	46
4.2.3    Base de imagens sintetizadas de envelopes postais brasileiros .....	47
4.2.4    Números sobre as bases de imagens.....	50
4.3    Estratégia de avaliação da metodologia de segmentação .....	51
4.4    Análise de Resultados.....	54
4.4.1    Outras considerações a respeito da Metodologia de Avaliação Proposta ...	65
4.4.2    Comentários Finais sobre a Análise dos Resultados .....	69

4.5	Análise da Sensibilidade dos Parâmetros do Algoritmo de Segmentação ..	69
4.5.1	Sensibilidade ao fator de seleção de coeficientes ( $\lambda_1$ ).....	70
4.5.2	Sensibilidade ao nível de significância ( $\lambda_2$ ) entre distribuições no teste de hipótese.....	73
4.5.3	Sensibilidade ao fator que delimita a região de existência de objetos de segmentação ( $\lambda_3$ ) .....	75
4.5.4	Sensibilidade ao tamanho $k$ da janela de Pontos Salientes.....	77
4.5.5	Sensibilidade ao limitador local ( $LimLocal(\alpha)$ ) gerado pelo arranjo entre os 4 <i>pixels</i> filhos referenciados pelo ponto saliente.....	79
4.5.6	Resumo da Análise da Sensibilidade.....	81
4.6	Considerações sobre a Complexidade Computacional.....	83
4.7	Comentários Finais.....	85
5.	Conclusões e Trabalhos Futuros.....	86
6.	Referências Bibliográficas.....	90
A.	Tipos de fundos de envelopes complexos sem objetos de segmentação.....	93

## Lista de Figuras

Figura 1.1: O TRANSORMA, a primeira máquina de triagem (1931).....	1
Figura 1.2: Exemplos de imagens de envelopes utilizadas no trabalho, apresentando grande variações de diagramação, tipos de fundos, tamanho, espessura e tipos de fontes manuscritas, quantidade, tamanho e posição de selos e carimbos. ....	3
Figura 1.3: Objetivo da Segmentação. (a) Imagem típica; (b) Imagem-resultado que deve ser obtida pelo método proposto tendo como entrada a imagem (a).....	6
Figura 3.1: Diagrama de blocos do método de segmentação proposto, indicando os 5 passos principais e suas respectivas saídas. $I$ tem dimensão $n \times m$ ; $I_W$ tem dimensão $n/2 \times m/2$ ; $I_S$ tem dimensão $n/2 \times m/2$ ; $I_{RC}$ tem dimensão $n/2k \times m/2k$ ; $I_{PRPS}$ tem dimensão $n \times m$ ; $I_{FINAL}$ tem dimensão $n \times m$ , onde $k$ é o tamanho de janela de $k^2$ elementos.....	17
Figura 3.2: Exemplo ilustrativo de imagem de entrada para o algoritmo de segmentação proposto.....	18
Figura 3.3: Diagrama de blocos ilustrando o algoritmo de decomposição <i>Wavelet</i> de Mallat.....	19
Figura 3.4: Primeiro Passo. (a) Imagem original ( $n \times m$ ); (b) A transformada <i>Wavelet</i> da imagem original pelo algoritmo de Mallat utilizando a função de base Haar ( $n/2 \times m/2$ ), gerando $I_W(BB)$ , $I_W(BA)$ , $I_W(AB)$ e $I_W(AA)$ .....	21
Figura 3.5: Gráfico típico de uma distribuição normal “Z” (ou gaussiana) com média zero e desvio padrão unitário; as retas indicam o ponto de corte $(100 - \lambda_1)/2$ ; as duas áreas hachuradas correspondem aos $\lambda_1\%$ dos coeficientes mais significativos.....	23
Figura 3.6: Gráfico da distribuições normal “Z” (pontos contínuos) e dos coeficientes <i>Wavelet</i> (pontilhada) de detalhe de uma imagem típica de envelope postal. (a) Distribuição dos coeficientes horizontais ( $BA$ ); (b) Distribuição dos coeficientes verticais ( $AB$ ). ....	24
Figura 3.7: Coeficientes mais significativos (Imagem $I_{WCS}$ ), com $\lambda_1 = 43\%$ ; (a) Imagem $I_{WCS(BA)}$ ; (b) Imagem $I_{WCS(AB)}$ .....	24
Figura 3.8: Resultado do segundo passo (imagem $I_S$ ), intersecção entre as imagens da Figura 3.7.....	25

Figura 3.9: Janela de pontos salientes. (a) Imagem representativa das janelas de pontos salientes ( $n/2k \times m/2k$ ), agrupados por uma janela de $k \times k$ ( $8 \times 8$ ); (b) Histograma das janelas de pontos salientes de (a).....	28
Figura 3.10: Pseudo-código do algoritmo de rotulação controlada de janelas de pontos salientes. ....	31
Figura 3.11: Saída do terceiro passo ( $I_{RC}$ - dimensão $n/2k \times m/2k$ ). (a) Conjunto $I_{RCW}$ , imagem das janelas de alta densidade de pontos salientes; (b) Conjunto $I_{RC}$ , removidas as janelas sem conexão com nenhuma outra janela vizinha.....	31
Figura 3.12: Ilustração da projeção reversa dos pontos salientes de janelas de alta densidade na imagem original em tons de cinza. ....	32
Figura 3.13: Projeção reversa dos pontos salientes. (a) Conjunto de pontos salientes correspondentes as janelas de alta densidade ( $n/2 \times m/2$ ); (b) Exemplo de uma janela de alta densidade de pontos salientes projetados reversamente na imagem original ( $2k \times 2k$ ); (c) Conjunto $I_{PRPS}$ ( $n \times m$ ). ....	33
Figura 3.14: Informação de Contexto. (a) Histograma de um envelope típico; (b) Histograma suposto ( $N(I_\mu, I_\sigma)$ ) de um envelope, com objetos de segmentação em $\lambda_3\%$ . ....	35
Figura 3.15: Transição entre o objeto de segmentação e o fundo; os pontos salientes. Um corte transversal ilustrativo em uma imagem típica de envelope.....	35
Figura 3.16: Os filhos ( $2 \times 2$ pixels) de pontos salientes de janelas de alta densidade. (a), (b) e (c) Evidência para objeto de segmentação; (d) e (e) Evidência para o fundo do envelope.....	36
Figura 3.17: Um pequeno recorte ilustrando o algoritmo de perseguição de contorno, utilizando $LimLocal(\alpha)$ como sendo a média dos quatro filhos. (a) Inicialização do contorno; (b) Incorporação dos vizinhos; (c)-(d) Evolução e finalização do contorno sem restrição global; (e)-(f) Evolução e finalização do contorno com restrição global. ....	39
Figura 3.18: Pseudo-código do algoritmo de perseguição de contorno. ....	40
Figura 3.19: Resultado do algoritmo de segmentação proposto. (a) $I_{CONTORNO}$ em tons de cinza; (b) Imagem $I_{FINAL}$ binária.....	41
Figura 4.1: Imagens de referência ( <i>ground truth</i> ). (a) Resultado da segmentação desejada; Imagens (b), (c) e (d) com os <i>pixels</i> das classes bloco endereço, selo e carimbo, respectivamente. ....	45

Figura 4.2: Imagens Reais. (a) Fundo branco; (b) Fundo complexo.....	46
Figura 4.3: Envelope postal. (a) Imagem típica; (b) Imagem segmentada desejada de (a). 47	
Figura 4.4: Esquema de construção de imagens sintetizadas através de imagens reais e suas respectivas imagens de referências e tipos de fundos complexos. $I_O$ é a imagem real original; $I_{GT1}$ , $I_{GT2}$ e $I_{GT3}$ são as imagens de referência ( <i>ground truth</i> ) das classes bloco endereço, selo e carimbo, respectivamente; $I_U$ é a união entre $I_{GT1}$ , $I_{GT2}$ e $I_{GT3}$ ; $I_{OS}$ contém todos os objetos de segmentação da imagem $I_O$ ; $I_F$ é o fundo complexo de envelope postal virgem (sem componentes); $I_{FS}$ é a região selecionada para ser o fundo do envelope postal da imagem a ser sintetizada; $I_{SINT}$ é a imagem sintetizada resultante deste processo; ( <i>Bin</i> ) indica que as imagens são marcadores. ....	48
Figura 4.5: Exemplos de imagens sintetizadas.....	49
Figura 4.6: Esquema de avaliação das imagens de envelopes postais. $I_O$ é a imagem original; $I_{GT}$ é a segmentação esperada (ou desejada) para $I_O$ ; $I_{GT1}$ , $I_{GT2}$ e $I_{GT3}$ são as imagens <i>ground-truth</i> (de referência) das classes bloco endereço, selo e carimbo, respectivamente, previamente separadas; $I_{FINAL}$ é o resultado final do algoritmo de segmentação; $I_{Ob1}$ , $I_{Ob2}$ e $I_{Ob3}$ são as imagens que contém os <i>pixels</i> segmentados de cada uma das classes: bloco endereço, selo e carimbo, respectivamente; $I_{Ruído}$ é a imagem que contém os <i>pixels</i> segmentados que não pertencem a nenhuma das três classes. ( <i>Bin</i> ) indica que as imagens são marcadores.....	53
Figura 4.7: Exemplos de imagens utilizadas nos experimentos. (a) e (b) Imagens que pertencem à base de imagens reais; (c) e (d) Imagens que pertencem à base de imagens sintetizadas. ....	55
Figura 4.8: Exemplo de imagens $I_{FINAL}$ obtidas nos experimentos pelo método proposto referente às imagens da Figura 4.7. (a) e (b) Resultados referentes a imagens que pertencem a bases de imagens reais; (c) e (d) Resultados referentes a imagens que pertencem à base de imagens sintetizadas. ....	56
Figura 4.9: Exemplos de imagens intermediárias criadas nos experimentos. Imagens de pontos salientes $I_S$ (dimensão $n/2 \times m/2$ ), no espaço <i>Wavelet</i> , saída do segundo passo do algoritmo de segmentação proposto. ....	57
Figura 4.10: Exemplos de imagens intermediárias criadas nos experimentos. Imagens das janelas rotuladas/selecionadas $I_{RC}$ (dimensão $n/2k \times m/2k$ , onde $k$ é o tamanho da janela de pontos salientes), saída do terceiro passo do algoritmo de segmentação proposto. ....	58

Figura 4.11: Exemplos de imagens intermediárias criadas nos experimentos. Imagens da projeção reversa dos pontos salientes selecionados $I_{PRPS}$ (dimensão $n \times m$ ), saída do quarto passo do algoritmo de segmentação proposto. ....	59
Figura 4.12: Exemplos de imagens intermediárias criadas nos experimentos. Imagens com os <i>pixels</i> que iniciaram o contorno, tarefa inicial do quinto passo do algoritmo proposto. ....	60
Figura 4.13: Gráfico da distribuição dos resultados da base de imagens reais (200 imagens). (a) Distribuição do ruído; (b) Distribuição da taxa de retorno do bloco endereço. ....	63
Figura 4.14: Gráfico da distribuição dos resultados da base de imagens sintetizadas (2000 imagens). (a) Distribuição do ruído; (b) Distribuição da taxa de retorno do bloco endereço. ....	64
Figura 4.15: Influência da taxa de retorno do bloco endereço na análise do resultado (taxa de retorno do bloco endereço de 79,44% e ruído de 0,00%). (a) Imagem original; (b) Imagem desejada; (c) Imagem segmentada (observada) pelo método proposto; (d) <i>Pixels</i> que não foram segmentados pelo algoritmo e que deveriam pertencer aos objetos de segmentação. ....	66
Figura 4.16: Localização do ruído na segmentação. (a) Bloco endereço danificado (ruído de 0,79%); (b) Imagem segmentada com alto ruído (2,28%), porém objetos de segmentação intactos. ....	67
Figura 4.17: Influência do ruído na análise do resultado (taxa de retorno do bloco endereço de 99,82% e ruído de 0,45%). (a) Imagem original; (b) Imagem desejada; (c) Imagem segmentada (observada) pelo método proposto; (d) <i>Pixels</i> que foram classificados como ruídos. ....	68
Figura 4.18: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do parâmetro $\lambda_1$ . ....	72
Figura 4.19: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do parâmetro $\lambda_2$ . ....	75
Figura 4.20: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do parâmetro $\lambda_3$ . ....	77
Figura 4.21: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do tamanho $k$ da janela de pontos salientes. ....	78
Figura 4.22: Posição do limitador local em relação ao relevo da imagem do envelope	

postal.....	79
Figura 4.23: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação aos arranjos do limitador local $LimLocal(\alpha)$ .....	81
Figura 4.24: Resultados dos experimentos utilizados para a análise de sensibilidade do método proposto sobre a base de imagens reais.....	82
Figura 4.25: Resultados dos experimentos utilizados para a análise de sensibilidade do método proposto sobre a base de imagens sintetizadas.....	83
Figura A.1: Tipos de Fundos de envelopes complexos utilizados para a construção da base de imagens sintetizadas.....	94

## Lista de Tabelas

Tabela 4.1: Quantidade de <i>pixels</i> das classes nas bases de imagens de envelopes postais em relação ao número total de <i>pixels</i> da imagem.....	51
Tabela 4.2: Exemplo hipotético para avaliação.....	54
Tabela 4.3: Conjunto de parâmetros que apresentaram melhor resultado nos experimentos. .....	61
Tabela 4.4: Resultados na base de imagens reais.....	62
Tabela 4.5: Resultados na base de imagens sintetizadas.....	62
Tabela 4.6: Valores de $\lambda_1$ usados na análise da sensibilidade.....	70
Tabela 4.7: Resultados obtidos com a variação de $\lambda_1$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.....	71
Tabela 4.8: Resultados obtidos com a variação de $\lambda_1$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.....	71
Tabela 4.9: Valores de $\lambda_2$ usados na análise da sensibilidade.....	73
Tabela 4.10: Resultados obtidos com a variação de $\lambda_2$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.....	74
Tabela 4.11: Resultados obtidos com a variação de $\lambda_2$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.....	74
Tabela 4.12: Valores de $\lambda_3$ usados na análise da sensibilidade.....	75
Tabela 4.13: Resultados obtidos com a variação de $\lambda_3$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.....	76
Tabela 4.14: Resultados obtidos com a variação de $\lambda_3$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.....	76
Tabela 4.15: Tamanho $k$ de janelas de pontos salientes usados na análise da sensibilidade. .....	77
Tabela 4.16: Resultados obtidos com a variação do tamanho $k$ da janela de pontos salientes, na base de imagens reais, mantendo-se todos os outros parâmetros fixos.....	78
Tabela 4.17: Resultados obtidos com a variação do tamanho $k$ da janela de pontos salientes, na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.....	78

Tabela 4.18: Os limitadores locais ( $LimLocal(\alpha)$ ) utilizados nos experimentos.....	80
Tabela 4.19: Resultados obtidos com os arranjos de $LimLocal(\alpha)$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos. ....	80
Tabela 4.20: Resultados obtidos com os arranjos de $LimLocal(\alpha)$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos. ....	80
Tabela 4.21: Tempo de processamento e esforço computacional do algoritmo de segmentação, codificado em C++, em um PIII dual de 1.2GHZ com 512MB de memória RAM, sobre a plataforma RedHat Linux. ....	84

## Lista de Símbolos

$BB$	Coeficientes de aproximação (ou de baixa frequência)
$BA$	Coeficientes de detalhe (ou de alta frequência) na direção horizontal.
$AB$	Coeficientes de detalhe (ou de alta frequência) na direção vertical.
$AA$	Coeficientes de detalhe (ou de alta frequência) na direção diagonal.
$1 \downarrow 2$	Operação de subamostragem ( <i>downsampling</i> ) de linhas.
$2 \downarrow 1$	Operação de subamostragem ( <i>downsampling</i> ) de colunas.
$QMF$	Filtro Espelhado em Quadratura ( <i>Quadrature Mirror Filter</i> )
$\tilde{H}$	Filtro $QMF$ Passa-baixa.
$\tilde{G}$	Filtro $QMF$ Passa-alta.
$Z_\alpha$	Ponto de probabilidade normalizada.
$\lambda_1$	Fator de seleção de coeficientes.
$\lambda_2$	Nível de significância do teste de hipótese
$\lambda_3$	Fator que delimita a região de existência de objetos de segmentação.
$LimLocal(\alpha)$	Limitador local para o algoritmo de crescimento de região
CB	Correios Brasileiros
ECT	Empresa Brasileira de Correios e Telégrafos
PUCPR	Pontifícia Universidade Católica do Paraná

## Resumo

A Automação Postal voltou a fazer parte da agenda de pesquisa das comunidades de Reconhecimento de Padrões e Visão por Computador, devido ao barateamento e à facilidade de aquisição das imagens de envelopes postais. Ao mesmo tempo, as unidades de processamento dos computadores de hoje evoluíram, em termos de frequência/velocidade, viabilizando, desta forma, a utilização de algoritmos complexos. A segmentação dos envelopes postais ainda é o ponto crucial no desenvolvimento de um sistema de Análise de Imagens para a Automação Postal. Neste trabalho é apresentada uma nova abordagem de segmentação de envelopes postais para localização do bloco endereço, baseada na seleção de características no espaço *Wavelet*. O objetivo é separar automaticamente nos envelopes postais as regiões relacionadas ao fundo do envelope, o carimbo, o selo, e o bloco endereço. Primeiro, uma imagem típica do envelope postal é decomposta usando o algoritmo de Mallat e a função de base Haar. Os canais de saída de alta frequência são analisados para localizar pontos de saliência, visando separar o fundo do envelope. Um teste de hipótese estatístico é realizado para decidir quais são as regiões mais consistentes, a fim de remover ruídos remanescentes. Os pontos selecionados são projetados de volta na imagem original em tons de cinza, onde a evidência do espaço *Wavelet* é usada para começar um processo de crescimento de região para incluir os *pixels* mais prováveis de pertencer às regiões do selo, carimbo e áreas escritas alcançando dessa forma a segmentação final. Experimentos foram executados sobre duas bases de imagens – uma original e outra sintetizada – que contém várias diagramações e tipos de fundos. Uma metodologia para avaliação quantitativa da segmentação baseada em imagens de referência (*ground truth*) é apresentada, que compara *pixel a pixel* a segmentação desejada com a segmentação encontrada. Foram alcançados resultados promissores com mais de 97% de taxa de retorno do bloco endereço com menos de 1% de ruído. Ao final, conclusões e direções futuras são indicadas nesta dissertação.

**Palavras-chave:** 1. Segmentação. 2. Envelopes Postais. 3. Seleção de Características. 4. *Wavelets*. 5. Bloco Endereço.

## Abstract

Postal automation has been recently integrated into the research agenda of Pattern Recognition and Computer Vision communities, since acquisition and storage of envelopes images has become easier and cheaper than a decade ago. Simultaneously, actual central processor units evolve in frequency/speed, becoming possible the use of complex algorithms. However, segmentation task of a typical image of a mail piece is still the greatest bottleneck in the development of an automatic postal system. In this work, a new approach of segmentation of postal envelopes for address block location, based on feature selection in *Wavelet* space, is introduced. The aim is to automatically separate in postal envelopes the regions related to background, stamps, rubber stamps, and the address blocks. First, a typical image of a postal envelope is decomposed using Mallat algorithm and Haar basis. High frequency channel outputs are analyzed to locate salient points in order to separate the background. A statistical hypothesis test is taken to decide upon more consistent regions in order to clean out some noise left. The selected points are projected back to the original gray level image, where the evidence from the wavelet space is used to start a growing process to include the pixels more likely to belong to the regions of stamps, rubber stamps, and written area. Experiments were run using two images databases – one typical and another synthesized - that have many kinds of layouts and backgrounds. A quantitative methodology for evaluating the segmentation based on ground truth (reference) images, was introduced. This methodology works comparing pixel by pixel the aimed image and the final image. Promising results were reached in more than 97% of address block rate with less than 1% of noise. Finally, conclusions and future works are indicated in this dissertation.

**Key-words:** 1. Segmentation. 2. Postal Envelopes. 3. Feature Selection. 4. *Wavelets*. 5. Address Block.

# Capítulo 1

## 1. Introdução

A Automação Postal voltou a fazer parte da agenda de pesquisa das comunidades de Reconhecimento de Padrões e Visão por Computador, pois a aquisição e armazenamento de imagens de objetos dos correios (correspondências, envelopes, etc.) se tornaram baratas e factíveis comparando-se há uma década atrás, ao mesmo tempo, o poder de processamento dos computadores também evoluiu viabilizando essa tarefa.

A primeira máquina de triagem postal conhecida – o TRANSORMA (Figura 1.1, extraída de [PHPM 2003]) – foi utilizada pela primeira vez em 1931 em Roterdã, na Holanda. As máquinas TRANSORMA também foram utilizadas na Bélgica, Canadá, Alemanha, Suíça, Reino Unido, Estados Unidos, Filipinas, Venezuela, Argentina, e no Brasil (em 1934).



Figura 1.1: O TRANSORMA, a primeira máquina de triagem (1931).

Hoje, os Correios Brasileiros (CB) enfrentam um volume diário de 34 milhões de objetos e correspondências. No ano de 2001, o total da carga postal foi de mais de 9,5 bilhões e em 2002, de 9,4 bilhões [ECT 2003]. Os CB têm máquinas de triagem capazes de processar até 36.000 cartas por hora. Entretanto, essas máquinas trabalham com

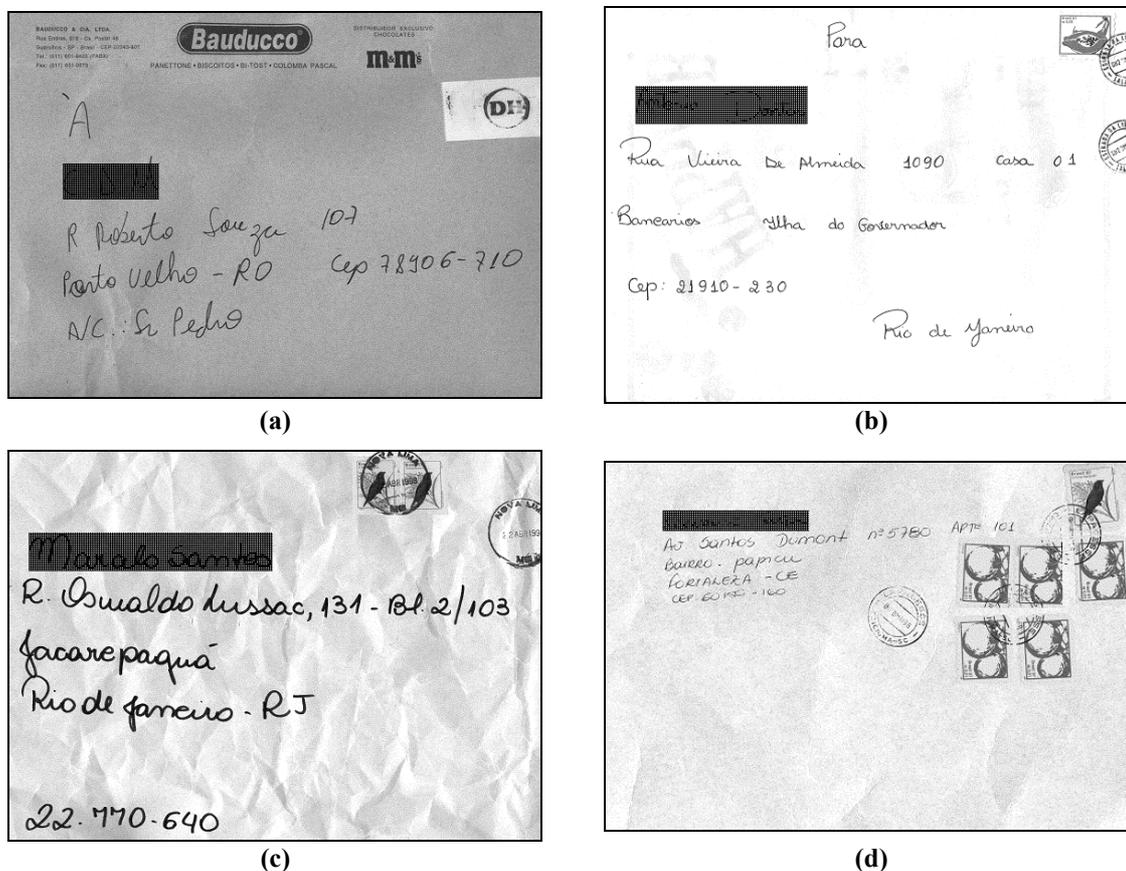
informações padronizadas, ou seja, as correspondências precisam obedecer a padrões rígidos de diagramação (*layout*), dimensão, tamanho e tipos de fontes, e fundo com bom contraste. Mesmo assim, esse serviço de triagem está disponível somente para grandes volumes de postagens, e deve ser previamente contratado nos CB. As correspondências que não obedecem a esses padrões, e até mesmo as correspondências dentro dos rígidos padrões estabelecidos pelos CB, mas que não foram contratados os serviços de triagem automática, são processadas manualmente. Essa tarefa de triagem manual é uma tarefa monótona e entediante para ser realizada por um ser humano. A necessidade de um sistema postal automático para os CB é de suma importância, visto que o trabalho de triagem da maioria das correspondências é feito manualmente, sendo essa uma motivação para a realização desse trabalho.

Dessa forma, faz-se necessário criar um sistema de automação postal que possa trabalhar com as correspondências que não obedecem a esses rígidos padrões, e com as quais os sistemas existentes não trabalham. O processo de automação postal pode consistir em um sistema de Visão que seja capaz de segmentar uma imagem de envelope postal, extraíndo as informações do remetente, e interpretando-as, possibilitando, dessa forma, o envio direto da correspondência.

O tema central desta dissertação consiste na tarefa de segmentação, que visa separar nos envelopes o fundo, o selo, o carimbo e o bloco endereço, sendo estes três últimos considerados objetos de segmentação. Apesar de haver vários trabalhos que tratam da segmentação de envelopes postais [WOLF *et al* 1997], [WOLF, NIEMANN 1997], [JAIN, BHATTACHARJEE 1992a], [JAIN, BHATTACHARJEE 1992b], [YU *et al* 1997], tal tarefa ainda é um problema desafiador, considerando a grande variedade de selos, fundos, e tipos e tamanhos de manuscritos/pré-impressos que compõem o bloco endereço, conforme pode ser observado em alguns exemplos da Figura 1.2.

Existe um padrão pré-determinado pelos CB na distribuição dos elementos existentes nos envelopes postais para facilitar o processo de triagem. Mas esse padrão não é obedecido pela população que utiliza o sistema de postagem brasileiro. O conteúdo existente nas imagens de envelopes postais que serão utilizadas neste trabalho pode ser resumido em alguns componentes, como o bloco endereço, os selos, os carimbos, e

informações pré-impressas (logotipos e textos ilustrativos); além do fundo do próprio envelope.



**Figura 1.2: Exemplos de imagens de envelopes utilizadas no trabalho, apresentando grande variações de diagramação, tipos de fundos, tamanho, espessura e tipos de fontes manuscritas, quantidade, tamanho e posição de selos e carimbos.**

O tamanho dos envelopes varia muito, desde cartas pessoais de tamanho padrão até grandes envelopes contendo cartas comerciais. As faces dos envelopes geralmente contêm textos e gráficos pré-impressos, além de selos, carimbos e, o mais importante, os endereços de remetente e destinatário. Não só a diagramação das imagens dos envelopes postais varia, mas também o contraste que os vários tipos de fundos proporcionam aos objetos presentes nos envelopes. Além disso, os fundos dos envelopes postais podem conter informações escritas.

O bloco endereço é composto por: nome do destinatário ou remetente, a rua ou avenida com número e completo, o bairro, a cidade, o estado ou unidade federativa, e também o CEP (Código de Endereçamento Postal). Não existe um padrão na distribuição geométrica desses dados que compõem o bloco endereço no envelope postal, mesmo para

as etiquetas de correspondências pré-impressas plastificadas (entrega de revistas e jornais por correio). Isto é, não existe uma ordem para a composição do bloco endereço postal especificada pelos CB. Salienta-se desta forma, a dificuldade de interpretação do bloco endereço, mesmo não sendo esse o objetivo deste trabalho.

Os objetos contidos nos envelopes, como o texto manuscrito pertencente ao bloco endereço, apresentam variações de tamanho, espessura e tipo. Além disso, não há uma posição fixa para o endereço do destinatário no envelope, apesar da recomendação dos CB.

As regiões relacionadas ao bloco endereço e ao carimbo dos CB têm propriedades físicas semelhantes, ou seja, intensidade parecida (em tons de cinza), distribuição geométrica semelhante, e a região de transição entre elas e o fundo do envelope é direta. O selo, todavia, tem uma distribuição espacial e geométrica diferente do bloco endereço e do carimbo. O selo pode ser sub-dividido em objetos e fundo do selo. A transição entre o fundo do envelope e o selo passa pelo fundo deste. A distinção entre o fundo do envelope e o fundo do selo faz-se pelo último ser mais claro (branco) que o primeiro. Os objetos do selo comportam-se como os objetos de segmentação que serão alvo do método de segmentação proposto aqui, entretanto, o fundo do selo não.

Geralmente, um carimbo é sobreposto sobre os selos contidos nos envelopes e, um ou mais carimbos aparecem em qualquer posição no fundo da correspondência. Às vezes, esses carimbos estão sobrepostos às informações do bloco endereço o que pode inviabilizar a identificação do bloco endereço.

As informações pré-impressas abrangem os logotipos e/ou textos pré-impressos no próprio fundo do envelope. Estas informações podem ter características semelhantes ao fundo do selo, ao bloco endereço, e ao carimbo. Portanto, estas informações podem ser segmentadas como objetos ou como fundo do envelope postal.

Todas essas variações existentes nos envelopes postais evidenciam ainda mais a necessidade de um algoritmo de segmentação que trate esses problemas. Logo, uma nova abordagem para segmentação de imagens de envelopes postais é apresentada neste trabalho. Esta abordagem consiste em um método de segmentação genérico e robusto que não é dependente de uma diagramação específica. O método proposto é dividido em cinco

passos, onde características salientes são localizadas no espaço *Wavelet* através de evidências e, usando estatísticas locais extraídas das imagens, hipóteses sobre as classes são construídas e testadas para alcançar a segmentação final.

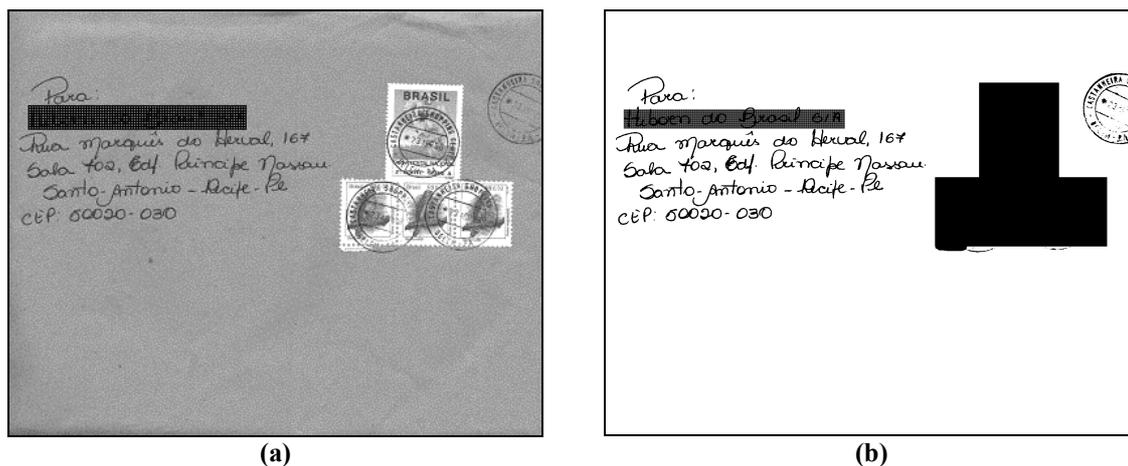
O método de segmentação proposto é fundamentado sobre a seguinte hipótese: “É possível utilizar pontos salientes como marcadores e geradores de evidências para segmentação dos elementos nas imagens”. Os experimentos realizados confirmam a hipótese criada.

Todas as imagens de envelopes postais apresentadas na dissertação estão em 256 tons de cinza (8 *bits* por *pixel*), foram obtidas através de um convênio entre a PUCPR e os CB (ou ECT - Empresa Brasileira de Correios e Telégrafos) e estão descritas em [ECT 2001]. Tarjas pretas tiveram que ser inseridas sobre os nomes dos destinatários existentes nos envelopes postais, obedecendo a uma cláusula de sigilosidade existente no convênio estabelecido entre a PUCPR e os CB.

O trabalho desenvolvido apresenta resultados promissores. Algumas análises e parte dos resultados aqui apresentados foram publicados por Menoti *et al* em [MENOTI *et al* 2003a] e [MENOTI *et al* 2003b]. Na média, mais de 97% dos *pixels* do bloco endereço foram segmentados com menos de 1% de ruído. Cada imagem levou, em média, seis segundos de processamento para ser segmentada. O sistema foi totalmente implementado em linguagem C++, executado em um ambiente dedicado em um Intel Pentium-III dual de 1.2 GHZ, com 3GB de memória RAM sobre a plataforma RedHat Linux.

## 1.1 Objetivos

Este trabalho visa definir uma metodologia para segmentação de envelopes postais para localização do bloco endereço utilizando um espaço de características *Wavelet*. Esta tarefa pode ser evidenciada na Figura 1.3. A Figura 1.3-b apresenta o resultado da segmentação que deve ser obtido pelo método de segmentação referente a imagem de envelope da Figura 1.3-a, separando, dessa forma, os objetos relacionados com o fundo do envelope, o selo, o carimbo e o bloco endereço.



**Figura 1.3: Objetivo da Segmentação. (a) Imagem típica; (b) Imagem-resultado que deve ser obtida pelo método proposto tendo como entrada a imagem (a).**

Para atingir tal objetivo devem ser cumpridos os seguintes objetivos específicos:

- Avaliar método de segmentação para os vários tipos de envelopes existentes;
- Uma forma de avaliação quantitativa "não subjetiva" (Grau da Precisão) deve ser empregada para avaliar o resultado da segmentação.

## 1.2 Justificativas

Nesta subseção, apresentam-se algumas justificativas em relação à viabilidade e ao uso das técnicas que compõem a abordagem proposta neste projeto.

A transformada *Wavelet* é uma ferramenta matemática poderosa, muito utilizada recentemente em problemas relacionados à área de Visão por Computador. Essa transformada possui várias propriedades que superam as dificuldades apresentadas pela transformada de Fourier e suas derivações. Alguns trabalhos que empregaram essas transformadas são apresentados em [JAIN, BHATTACHARJEE 1992b], [WU *et al* 1997]. A diferença básica entre a decomposição por Fourier e a decomposição por *Wavelet* é que as funções *Wavelet* possuem localização no tempo e na escala (frequência), enquanto que as funções senoidais (seno e cosseno), utilizadas pela transformada de Fourier, não possuem.

A transformada *Wavelet* permite identificar descontinuidades/singularidades na imagem, que são as evidências para os objetos de segmentação desse trabalho.

A transformada *Wavelet*, realizada através do algoritmo de decomposição de Mallat, permite reduzir significativamente o espaço de características (através de subamostragens), e ao mesmo tempo possibilita a reconstrução perfeita da imagem original. A representação conjunta de tempo-escala que a transformada *Wavelet* proporciona é uma propriedade muito útil para o método de segmentação aqui proposto. Essa representação permite a localização das informações no espaço original através do espaço *Wavelet*.

### **1.3 Contribuições**

Nesta subseção apresentam-se as contribuições que este trabalho de pesquisa produziu:

- Um método de segmentação, como um todo, de imagens de envelopes postais foi proposto para a localização do bloco endereço;
- Os resultados da segmentação foram avaliados de forma quantitativa.

### **1.4 Organização da Dissertação**

Esta dissertação está organizada em cinco capítulos. No segundo capítulo é apresentada uma revisão sobre os trabalhos relacionados. No Capítulo 3, o método proposto para a segmentação de envelopes postais é descrito. Os experimentos realizados neste trabalho no sentido de validar o método proposto são apresentados no Capítulo 4, juntamente com uma análise da sensibilidade dos parâmetros em relação aos resultados nas bases de imagens. Neste capítulo também é apresentada uma forma de avaliação quantitativa/objetiva (*pixel a pixel*) da segmentação. No quinto e último capítulo são apresentadas as conclusões e propostas de trabalhos futuros.

## Capítulo 2

### 2. Trabalhos Relacionados

Neste capítulo serão apresentados trabalhos que se relacionam com o tema central da dissertação, a segmentação de envelopes postais. Embora o relacionamento de alguns trabalhos com o tema proposto não seja direto, eles são focados em diferentes aspectos de segmentação de imagens que envolvem regiões de texto em imagens.

Uma abordagem para localizar as áreas de texto em fundos complexos baseada na decomposição *Wavelet* é apresentada por Jin e Tang em [JIN; TANG 2001]. Essa abordagem utiliza uma pseudomovimentação da imagem, que é uma translação que pode ser vista como uma mudança na posição de visão do observador em relação à imagem. A translação da imagem faz com que os coeficientes da transformada *Wavelet* (usando a função de base Haar) oscilem e esta oscilação permite a identificação das regiões de texto. Um algoritmo baseado na decomposição *Wavelet* com pseudomovimento é desenvolvido para modelar essas oscilações e identificar as áreas de texto em imagens de documento. Segundo os autores, esse método pode trabalhar com diferentes tipos de texto: tanto de grandes caracteres quanto de pequenos; caracteres pretos e brancos; caracteres em posições horizontais, verticais e inclinadas; com fundos simples e complexos. Resultados experimentais são apresentados para ilustrar o desempenho do algoritmo na detecção de áreas de textos nas imagens.

Etemad *et al* em [Etemad *et al* 1994] apresenta um algoritmo de segmentação de documentos independente de diagramação baseado em *Wavelet packets*, que trata regiões de textos, de imagens e de gráficos, presentes no documento, como três classes de textura diferentes. Decisões locais suaves/difusas sobre pequenos blocos – janelas quadradas – são realizadas usando dez canais *Wavelet* com mais energia (esses canais são oriundos da transformada *Wavelet packets*) baseado em vetores de características. A segmentação é executada pela propagação e integração das decisões locais suaves/difusas entre os blocos vizinhos, dentro da mesma escala e através das escalas. Estas propagações e integrações

das decisões locais são baseadas em ponderações na vizinhança dentro e através das escalas. As decisões locais suaves/difusas são modeladas por um classificador que é implementado através de uma rede neuronal de múltiplas camadas em avanço, sendo composta por vinte neurônios de entrada, oito neurônios na camada escondida e três unidades de saída. Cada uma dessas três unidades representa as decisões locais difusas (valores de 0 a 1) de um dado bloco pertencer a uma classe. Os dados de entrada consistem de várias imagens em tons de cinza adquiridas a 200 *dpi*. Essa rede foi treinada com 400 amostras de 16x16 *pixels* para cada uma das três classes. Segundo os autores, esse algoritmo pode ser aplicado a outras tarefas de segmentação de imagens. Não são apresentados, nem comentados, resultados sobre a avaliação da segmentação sobre nenhuma base de imagens, somente um exemplo ilustrativo.

Um sistema baseado em aglomeração (*clustering*) detectando e extraíndo/reconhecendo texto em imagens, tais como revistas, jornais, e correspondências, é apresentado em [WU *et al* 1997], e é dividido em quatro passos principais: 1) Um esquema de segmentação de textura é usado para localizar regiões onde o texto pode ocorrer, através de um banco de filtros gaussianos derivativos (com três escalas e três orientações, gerando nove filtros) seguidos por uma transformação não-linear ( $\tanh(\alpha)$ ) que produz um vetor de características. O processamento multi-escala é usado para considerar as variações do tamanho de fonte do texto, utilizando apenas três escalas. É realizado um agrupamento (*clustering*) no vetor de características usando o algoritmo *k-means* [TOU, GONZALEZ 1974] com  $k = 3$ . Uma das classes é definida como classe texto automaticamente e um fechamento morfológico é executado nesta classe; 2) Segmentos são extraídos das regiões de textos, e então são processados para formar caixas retangulares ao redor das cadeias de texto; 3) O texto é extraído através da remoção do fundo e pela binarização das cadeias de texto detectadas; 4) Caixas retangulares melhoradas são geradas baseadas nas cadeias de texto binarizadas e, estas novas caixas são novamente binarizadas e submetidas a um sistema de reconhecimento (*OCR*) comercial. Segundo os autores este sistema é estável, robusto, e funciona bem em imagens com e sem diagramação estruturada. O sistema foi testado sobre 48 imagens de várias origens: *frames* de vídeo digitalizados, fotografias, jornais, anúncios de revistas e cheques pessoais. Somente, 35 das 48 imagens foram submetidas a um OCR, onde cerca de 84% dos caracteres presentes nestas imagens foram

reconhecidos e 77% das palavras compostas por estes caracteres foram corretamente reconhecidas.

Uma abordagem para segmentar imagens de documentos em regiões tais como texto, fundo, cabeçalho e imagens (fotos) é apresentado por Cheng *et al* em [CHENG *et al* 1997, CHENG; BOUMAN 1998, CHENG; BOUMAN 2001], que explora tanto características de textura local quanto a estrutura da imagem. O método é baseado em uma metodologia bayesiana que permite modelar precisamente tanto as características da imagem quanto a estrutura contextual de cada região. As características de textura local são extraídas em cada resolução através da decomposição *Wavelet*. Os parâmetros que descrevem as características das imagens típicas são extraídos de uma base de imagens de treinamento que são produzidas por imagens típicas de documentos segmentados manualmente em componentes (ou regiões) desejados. Uma grande vantagem desse algoritmo de segmentação é que ele pode ser treinado para aplicações de segmentação específicas, simplesmente fornecendo exemplos de imagens com suas correspondentes imagens de referência (*ground truth*). Apesar de serem usadas imagens *ground truth* para o treinamento desta metodologia, os autores não relatam resultados quantitativos (*pixel a pixel*) sobre os testes realizados da segmentação, entretanto, as imagens utilizadas para ilustrar o desempenho do algoritmo são convincentes no sentido que o método apresenta visualmente uma segmentação satisfatória. Esta conclusão é feita através da observação das imagens, uma forma subjetiva de avaliação.

Choi e Baraniuk apresentam em [CHOI, BARANIUK 2001] um algoritmo de segmentação de textura baseado na transformada *Wavelet* e em um Modelo de Markov Escondido Hierárquico. Este modelo é um grafo probabilístico estruturado em árvore que captura as propriedades estatísticas dos coeficientes da transformada *Wavelet*. Esta árvore é construída segundo uma abordagem multi-escala bayesiana, agrupando as segmentações em várias escalas para obter uma segmentação final confiável. O algoritmo é testado com imagens sintetizadas, de fotos aéreas, e com documentos, mostrando resultados promissores. O algoritmo apresentado trabalha diretamente sobre a transformada *Wavelet* da imagem, isto possibilita que o algoritmo segmente imagens compactadas pela transformada *Wavelet* sem a necessidade de descompactá-las (ao domínio do espaço). Da mesma forma que em [CHENG *et al* 1997, CHENG; BOUMAN 1998, CHENG;

BOUMAN 2001], resultados quantitativos não são apresentados em relação aos testes realizados, embora os poucos exemplos apresentados no artigo apresentam visualmente uma boa segmentação.

Wolf e Niemann apresentam em [WOLF, NIEMANN 1997] uma abordagem para localização automática do bloco endereço em capas de revistas, cartas comerciais e jornais, baseada no reconhecimento da estrutura de um endereço de correspondências alemãs. A imagem é decomposta, através de uma técnica de segmentação básica (binarização), em pequenos blocos. Operações morfológicas de erosão e dilatação são utilizadas na imagem para que após este processamento, a imagem contenha poucos blocos não conectados. Esses blocos são classificados em blocos de caracteres normais e blocos de caracteres grandes, utilizando propriedades estruturais (largura e altura). Os blocos mais próximos no sentido vertical ou horizontal (dependendo da orientação) são fundidos e ditos blocos adjacentes. Esses blocos adjacentes, por sua vez, são mesclados para gerar candidatos ao bloco endereço do destinatário. Regras heurísticas e características típicas são usadas para identificar o bloco endereço do destinatário correto entre estes candidatos. Experimentos foram realizados em 213 imagens, sendo que em 195 delas o bloco endereço do destinatário foi recuperado com sucesso (91,5%). Esta abordagem é restrita a propriedades geométricas, visto que ela utiliza muitas heurísticas e só leva em consideração a variação de todo o bloco endereço e não a de seus componentes em separado, entretanto é voltada para as necessidades de desempenho do processo de automação postal dos correios alemães.

Wolf *et al* propuseram em [WOLF *et al* 1997] uma metodologia de localização do bloco endereço de imagens de envelopes postais em tempo real, baseada na detecção de blocos não homogêneos. A imagem do envelope é dividida em pequenos blocos retangulares, através dos quais medidas de homogeneidade são avaliadas. Depois de determinada uma binarização adaptativa, blocos não homogêneos são detectados em um processo de busca hierárquico. Os blocos não homogêneos adjacentes são mesclados em áreas de formas arbitrárias. Estas áreas são representadas pelo menor retângulo que as contém. Várias características são calculadas para cada área e um teste de plausibilidade é realizado. A área que contém o melhor escore é assumida como a área que contém o bloco endereço. Menos heurísticas são utilizadas neste método em relação à abordagem proposta

por Wolf e Niemann em [WOLF, NIEMANN 1997]. Esta abordagem cumpre a restrição de tempo de processamento (menos de 200 milissegundos), codificado em *Assembly*, em um computador baseado na arquitetura *i-860*, exigida para automação postal. Experimentos foram realizados em uma base de 2000 imagens de envelopes, dentre os quais cerca de 1600 deles eram pré-impressos (etiquetas), e cerca de 400 eram manuscritos de diferentes autores, variando o fundo e grandes quantidades de informações “estranhas” (logotipos e informações pré-impressas concernentes ao fundo) presentes nos envelopes. O sistema encontrou o bloco endereço em cerca de 98% das imagens dos envelopes, utilizando as três melhores áreas candidatas a bloco endereço.

Um método para localizar o bloco endereço em imagens onde a diagramação do texto impresso é conhecida *a priori* é apresentado por Yu *et al* em [YU *et al* 1997]. Diferente de simples envelopes que têm um alto grau de estrutura espacial global junto com um número restrito de entidades, muitas correspondências comerciais (cartas, revistas e jornais), enviadas pelo correio, têm o bloco endereço impresso em uma etiqueta que pode ser colada em uma posição e orientação arbitrárias junto com o texto, gráficos e imagens na capa da correspondência. O bloco endereço pré-impresso tem algumas propriedades: (i) ele satisfaz a regra do alinhamento à esquerda; (ii) ele é impresso em uma etiqueta branca ou amarelada, que pode ser colada sobre a correspondência com alguma inclinação; (iii) o contraste e o espaçamento do texto variam de uma impressora *laser* para uma impressora matricial, e de acordo com os tipos de estilo de fonte; (iv) o endereço freqüentemente deixa transparecer o fundo da correspondência. É nesta diagramação que o trabalho apresentado por Yu *et al* [1997] se baseia. Primeiramente, as correspondências são segmentadas segundo o algoritmo global de Otsu com uma modificação na função critério, para diminuir a influência da distribuição dos tons de cinza. O cálculo da função critério é baseado no canal vermelho – o de maior contraste – das imagens coloridas. Depois de binarizada, a imagem é submetida a uma extração de componentes conectados, baseada em grafo de blocos adjacentes (*BAG*). Muitas heurísticas, entretanto, são utilizadas nesta abordagem. Os componentes conectados devem estar sobre uma etiqueta que tem fundo branco ou amarelado, e estes são preservados caso seus *pixels* vizinhos sejam claros (segundo um limiar fixo). Os componentes conectados são modelados como blocos de linhas, que serão modelados como blocos de texto, e que finalmente serão modelados como bloco endereço. Essa tarefa de modelagem é realizada usando o conhecimento *a*

*priori* sobre o bloco endereço. Depois de modelados os blocos endereços são caracterizados, ou seja, algumas características são extraídas destes blocos. Estas características são utilizadas para determinar um grau de confiança. Os blocos de endereço são ordenados segundo o grau de confiança. A lista final ordenada (de 0 a 5 blocos endereços candidatos) é avaliada por um reconhecedor. Os experimentos foram realizados em 109 imagens RGB a 40 *dpi* a uma velocidade de 10 correspondências por segundo. Das 109 imagens utilizadas, 53 foram cedidas pela IBM, sendo que o sistema reconheceu corretamente o bloco endereço em cerca de 38 imagens (71,70%), as outras 56 imagens foram adquiridas pelos próprios autores, sendo que em 52 delas o sistema funcionou corretamente (92,86%).

Um outro trabalho que mostra resultados em envelopes postais é apresentado por Jain e Bhattacharjee em [JAIN, BHATTACHARJEE 1992b], que propõe um algoritmo de segmentação de textura baseado nos filtros de Gabor para localizar o bloco endereço do destinatário. Um vetor de características é gerado para cada *pixel* da imagem através de um banco de filtros (quatro orientações em cada uma das duas frequências/escalas) de Gabor mais as posições do *pixel* dentro da imagem. Uma função não linear é aplicada ao banco de filtros de Gabor, seguido de uma filtragem gaussiana por janela, que é realizada para suavizar os limites entre as regiões de textura. Nesse vetor de características é efetuado um algoritmo de agregamento (*clustering*) gerando-se três regiões/agregados. Uma das regiões segmentadas corresponde ao texto contido na imagem de envelope. Essa região é isolada e uma análise em componentes conectados é usada para identificar blocos individuais de texto. O bloco endereço do destinatário (BED) é identificado através destes blocos. Experimentos foram realizados sobre 174 imagens de envelopes postais onde: 81% dos envelopes tiveram o BED corretamente identificado e o processo de automação postal poderia ter sido completado; em 4% dos envelopes o BED, apesar de identificado corretamente, não possuía informação suficiente para a postagem automática; em 10% dos envelopes o BED foi identificado erroneamente; e nos 5% restantes de envelopes do experimento, nenhum BED foi identificado.

Já em [JAIN, BHATTACHARJEE 1992a], Jain e Bhattacharjee apresentam um método supervisionado utilizando os filtros de Gabor para classificar as regiões do envelope postal em texto e não-texto. Esta tarefa de classificação supervisionada é

realizada por uma rede neuronal de uma única camada com somente quatro entradas representando as quatro características de textura – sendo filtros de Gabor com duas orientações para cada uma das duas frequências/escalas – com uma única saída. O bloco endereço é identificado da mesma forma que em [JAIN, BHATTACHARJEE 1992b]. Experimentos detalhados como em [JAIN, BHATTACHARJEE 1992b] não são apresentados.

Um sistema integrado de interpretação de bloco endereço manuscrito nos Correios dos Estados Unidos através de leitura por computador remoto é apresentado por Srihari e Kuebert em [SRIHARI, KUEBERT 1997]. Este sistema é um gerenciador de imagens que atribui códigos de barras às correspondências que não foram totalmente processadas pelos *OCR's* postais. Nesse artigo são apresentados vários aspectos tecnológicos utilizados para a implantação desse sistema.

Uma análise sobre o universo de imagens de documentos é apresentada em [HARALICK; 1994], que abrange documentos como cartas comerciais, formulários e artigos científicos e técnicos como aqueles encontrados em periódicos e conferências técnicas. Este estudo preocupa-se com os aspectos geométricos da diagramação e com a estrutura lógica dos documentos.

Todos os trabalhos apresentados relacionam-se com a segmentação de regiões textuais, alguns estão relacionados à tarefa de segmentação do bloco endereço e outros abrangem o reconhecimento do bloco endereço. Entretanto, nenhum deles apresenta avaliações quantitativas (objetivas) sobre os objetos de segmentação extraídos. Também não mostram quantitativamente o ruído presente na segmentação. Esses trabalhos restringem-se a avaliações da segmentação de formas qualitativas ou visuais. Nesta dissertação será apresentada uma metodologia de avaliação (*pixel a pixel*), que permite determinar de forma quantitativa os resultados da segmentação.

É importante lembrar também, que a grande maioria dos trabalhos relacionados apresentados neste capítulo apresentam algumas restrições quanto a diagramação, tamanho do envelope, tipo de fundo, contraste do mesmo com os objetos, tipos e tamanhos de fontes, etc. O trabalho proposto aqui visa apresentar um método de segmentação que seja

genérico e robusto no sentido de que deseja tratar de todas essas variações existentes nos envelopes postais.

No capítulo seguinte é proposto um método para a segmentação de envelopes postais utilizando os canais da transformada *Wavelet* para gerar evidências para a segmentação. Os cinco passos principais que compõem o algoritmo de segmentação proposto aqui também serão descritos no próximo capítulo.

## Capítulo 3

### 3. Método Proposto

O método proposto consiste na segmentação de imagens de envelopes postais através da seleção de características no espaço *Wavelet*. O objetivo deste método é realizar a separação, de forma automática, nos envelopes postais, das regiões relacionadas ao fundo do envelope, ao selo, ao carimbo e ao bloco endereço.

A abordagem proposta pode ser dividida em cinco passos principais, a saber:

- 1) Inicialmente, a imagem original ( $I$ ) do envelope postal é decomposta no espaço *Wavelet*, utilizando-se o algoritmo de decomposição de Mallat [MALLAT 1989] com a base Haar. Esse algoritmo executa uma transformação não-redundante e ortogonal que resulta em quatro canais de saída de características, sendo estes:  $BB$  (banda de aproximação, ou baixa frequência),  $BA$  (banda de detalhe horizontal),  $AB$  (banda de detalhe vertical) e  $AA$  (banda de detalhe diagonal), gerando, assim, a saída do primeiro passo ( $I_w$ );
- 2) A identificação dos pontos salientes é baseada na intersecção dos coeficientes mais significativos dos canais de alta frequência  $BA$  (horizontal) e  $AB$  (vertical), produzindo desta forma a saída do segundo passo ( $I_s$ );
- 3) A rotulação controlada das janelas (de tamanho  $k$ , e  $k^2$  elementos) de pontos salientes, para eliminar o ruído e a separação do fundo do envelope, é baseada em um teste de hipótese estatístico para um conjunto de janelas, resultando a saída  $I_{RC}$ ;

- 4) A projeção reversa dos pontos salientes das janelas selecionados na imagem de níveis de cinza original, que produz  $I_{PRPS}$ ;
- 5) A perseguição de contorno através dos *pixels* referenciados pelos pontos salientes é baseada em teste de hipótese estatístico gerando o resultado final  $I_{FINAL}$ .

A Figura 3.1 mostra um diagrama de blocos da abordagem proposta.

A palavra “informação” referenciada à imagem, descrito nos passos deste capítulo, está intimamente ligada à dimensão das imagens. Quando a dimensão da imagem for  $n \times m$ , entenda a palavra informação como uma referência aos *pixels*. Quando a dimensão da imagem for  $n/2 \times m/2$ , a palavra “informação” deve fazer referência aos coeficientes ou pontos salientes, pois se trata do espaço *Wavelet*. E por último, o conceito de imagem deve ser entendido como uma matriz de janelas quando a dimensão da imagem for  $n/2k \times m/2k$ .

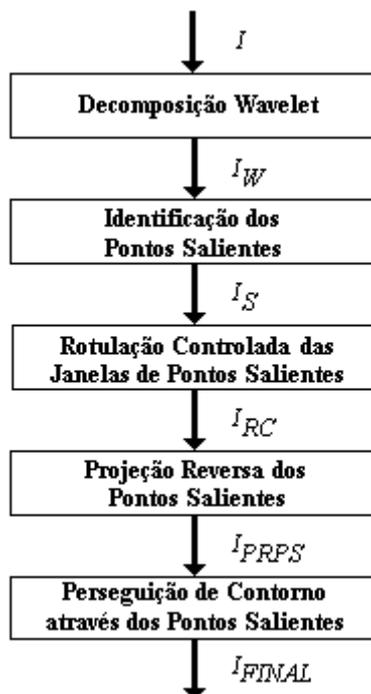


Figura 3.1: Diagrama de blocos do método de segmentação proposto, indicando os 5 passos principais e suas respectivas saídas.  $I$  tem dimensão  $n \times m$ ;  $I_W$  tem dimensão  $n/2 \times m/2$ ;  $I_S$  tem dimensão  $n/2 \times m/2$ ;  $I_{RC}$  tem dimensão  $n/2k \times m/2k$ ;  $I_{PRPS}$  tem dimensão  $n \times m$ ;  $I_{FINAL}$  tem dimensão  $n \times m$ , onde  $k$  é o tamanho de janela de  $k^2$  elementos.

As primeiras cinco subseções deste capítulo apresentam os cinco passos principais do algoritmo de segmentação proposto. Para ilustrar esses passos utilizar-se-á a Figura 3.2 como exemplo de imagem de entrada no algoritmo de segmentação, e outras imagens, derivadas da Figura 3.2, também serão utilizadas para ilustrar as entradas e saídas dos outros passos. Por último, na Seção 3.6, comentários finais são feitos sobre o método proposto.

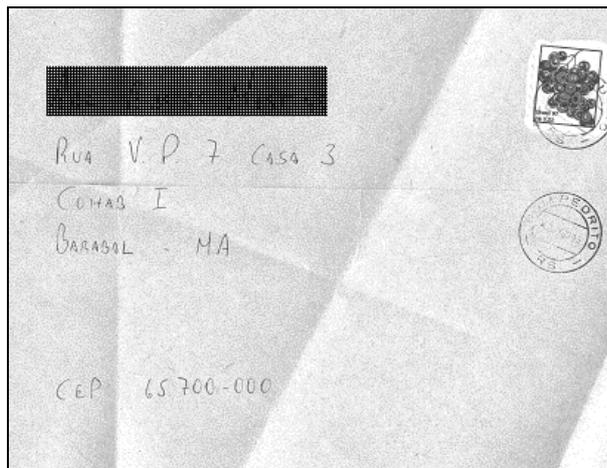


Figura 3.2: Exemplo ilustrativo de imagem de entrada para o algoritmo de segmentação proposto.

### 3.1 Decomposição *Wavelet*

A transformada *Wavelet* decompõe dados em blocos fundamentais de construção. A diferença básica entre a transformada de *Fourier* e a transformada *Wavelet* é que a transformada *Wavelet* utiliza funções *Wavelet*, que tem uma boa localização tanto no tempo quanto no espaço, enquanto que a transformada de *Fourier* que usa funções senoidais (seno / cosseno) não as têm. Visto que é possível desenvolver decomposições *Wavelet* com uma grande variedade de funções básicas, e também enfatizar a redundância ou eliminá-la através dos níveis de decomposição, existe uma farta literatura de diferentes técnicas úteis para isto [STRANG, NGUYEN 1996].

Entretanto, a tarefa de segmentação a ser resolvida aqui exige uma decomposição que ajude a localizar descontinuidades na imagem mais prováveis de serem as bordas do bloco endereço, selo e carimbo.

A decomposição de Mallat [MALLAT 1989] é um esquema de decimação que produz como saída quatro conjuntos relacionados à imagem original. O cálculo da transformada *Wavelet* de uma imagem pelo algoritmo de Mallat pode ser realizado por um algoritmo piramidal baseado em convoluções com filtros *QMF's* (*Quadrature Mirror Filters*) e sub-amostragens (*downsampling*). O algoritmo em questão é ilustrado pelo diagrama de blocos da Figura 3.3.

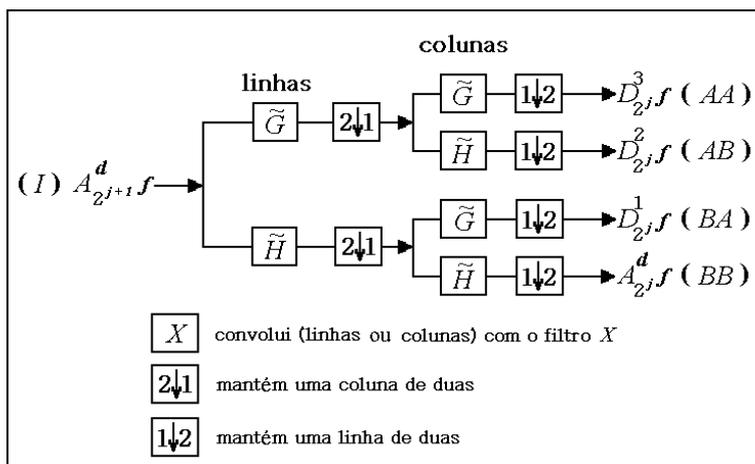


Figura 3.3: Diagrama de blocos ilustrando o algoritmo de decomposição *Wavelet* de Mallat.

A notação utilizada neste diagrama de blocos é a mesma utilizada em [MALLAT 1989], onde  $A_{2^j}^d f$  é chamado de aproximação (*A*) discreta (*d*) de uma função  $f(x, y)$  na resolução  $2^j$  e  $D_{2^j}^n f$  é chamado de detalhe (*D*) discreto de uma função  $f(x, y)$  na resolução  $2^j$  da *l*-ésima imagem; *l* indica uma das três imagens de detalhe. Este diagrama da Figura 3.3 já é a representação estendida para imagens (sinais bidimensionais).

A imagem original  $A_{2^{j+1}}^d f$  é decomposta em quatro imagens:  $A_{2^j}^d f$ ,  $D_{2^j}^1 f$ ,  $D_{2^j}^2 f$  e  $D_{2^j}^3 f$ , onde as três últimas representam as bandas de alta frequência ou coeficientes de detalhe e  $A_{2^j}^d f$  é a banda de baixa frequência ou coeficientes de aproximação. Este algoritmo é baseado em convoluções unidimensionais entre as linhas ou colunas da imagem  $A_{2^{j+1}}^d f$  e um dos filtros unidimensionais *QMF*  $\tilde{H}$  e  $\tilde{G}$ .

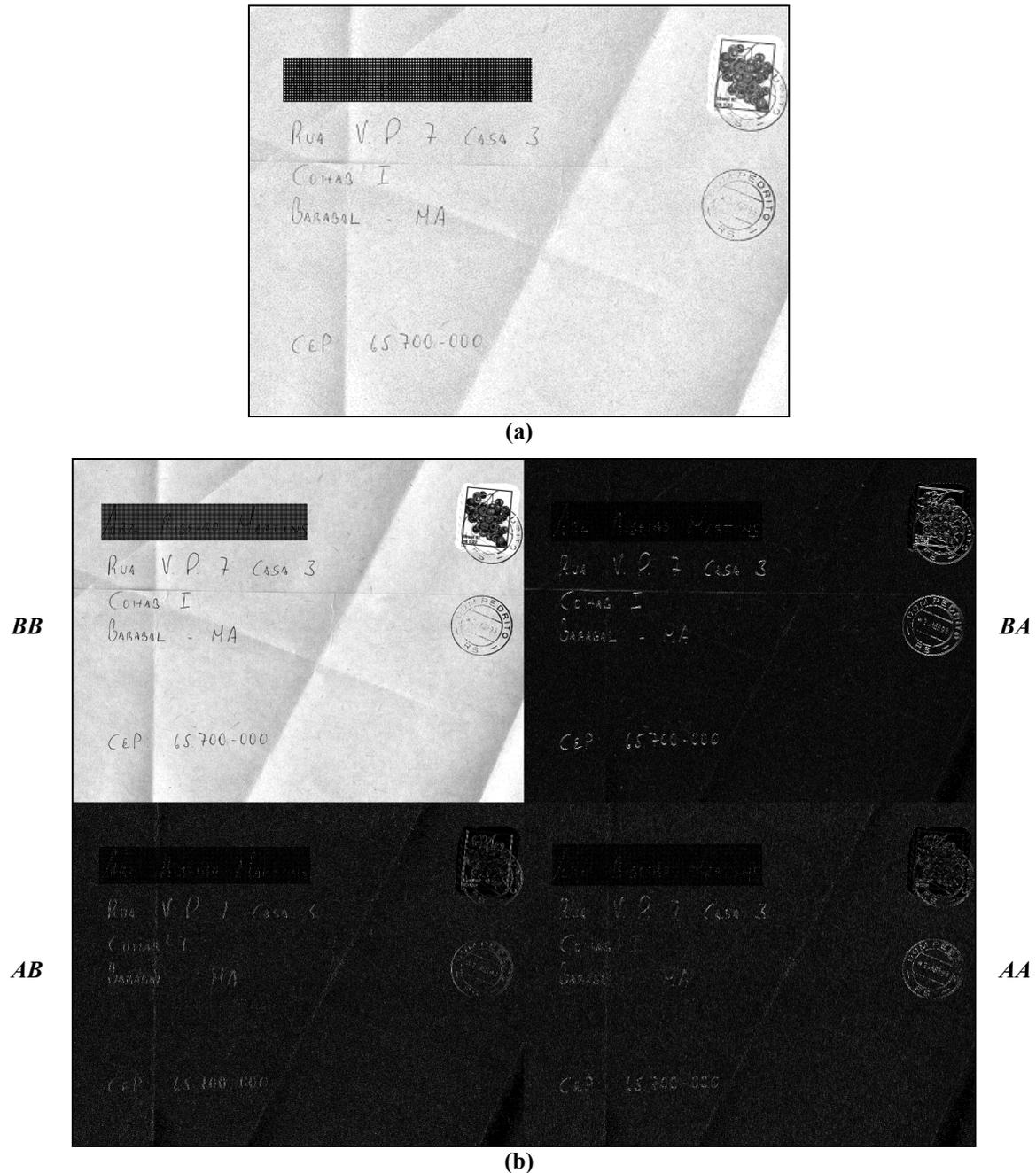
Os dois filtros *QMF* usados aqui,  $\tilde{G}$  (de alta frequência) e  $\tilde{H}$  (de baixa frequência), para a decomposição *Wavelet* no algoritmo de Mallat representando a função de base Haar são representados por vetores de comprimento dois, isto é:

$$\tilde{H} = \frac{\sqrt{2}}{2} [+1; +1] \quad (3.1)$$

$$\tilde{G} = \frac{\sqrt{2}}{2} [-1; +1] \quad (3.2)$$

O comprimento dos filtros *QMF* que representam estas funções de base é diretamente proporcional à complexidade  $O(n)$  do algoritmo de decomposição. Os filtros *QMF* que representam a função de base Haar têm o menor comprimento conhecido e, portanto, esta função de base apresenta uma menor complexidade algorítmica. A função de base Haar possibilita a localização de descontinuidades em imagens. Esses são dois fortes motivos para a escolha desta função de base para esta aplicação.

Os quatro conjuntos de saída da decomposição *Wavelet* são: um conjunto para dados suavizados ou de baixa frequência (*BB*); e mais três conjuntos de detalhes, sendo conjuntos direcionais de alta frequência: horizontal (*BA*), vertical (*AB*) e diagonal (*AA*). Todos estes conjuntos são números reais resultantes das convoluções com os filtros *QMF*. A Figura 3.4-b apresenta os quatros canais de saída do algoritmo de decomposição *Wavelet* de Mallat: *BB*, *BA*, *AB* e *AA*; tendo como entrada a imagem da Figura 3.4-a. O conjunto *BB* apresentado na Figura 3.4-b é uma aproximação sub-amostrada da imagem original. Já os conjuntos *BA*, *AB* e *AA* também são sub-amostragens da imagem original, porém esses contêm os detalhes que, somados a *BB*, reconstruem a imagem original. As imagens, dos coeficientes de alta frequência, *BA* (horizontal), *AB* (vertical), e *AA* (diagonal), apresentadas na Figura 3.4-b são representadas da seguinte forma: os coeficientes mais escuros representam as regiões de menor resposta aos filtros enquanto que os coeficientes mais claros representam as regiões de alta resposta do filtro. Para que esta visualização fosse possível, estas imagens foram normalizadas segundo oito vezes o valor do desvio padrão. Finalmente estas imagens foram tomadas em valor absoluto e distribuídas em 256 tons de cinza. Faz-se necessário ressaltar que este procedimento foi adotado somente para facilitar a visualização dos coeficientes de detalhe.



**Figura 3.4: Primeiro Passo. (a) Imagem original ( $n \times m$ ); (b) A transformada *Wavelet* da imagem original pelo algoritmo de Mallat utilizando a função de base Haar ( $n/2 \times m/2$ ), gerando  $I_w(BB)$ ,  $I_w(BA)$ ,  $I_w(AB)$  e  $I_w(AA)$ .**

A primeira etapa do método proposto consiste em transformar  $I$  (imagem original) em  $I_w$  (saída da primeira etapa) como um passo de preparação para a segmentação, usando somente o primeiro nível de decomposição. O algoritmo de decomposição *Wavelet* apresenta complexidade computacional  $O(n)$  e permite reduzir o espaço de características.

Para o algoritmo de decomposição em questão, utilizando a base Haar, a dimensão dos canais de saída é a “metade” da dimensão da imagem original, ou seja, se a imagem original  $I$  possui dimensão  $n \times m$  (linhas e colunas), então  $I_w$  possuirá dimensão  $n/2 \times m/2$  (linhas e colunas). Esta redução de dimensionalidade é realizada através da sub-amostragem da decomposição *Wavelet*.

### 3.2 Identificação dos Pontos Salientes

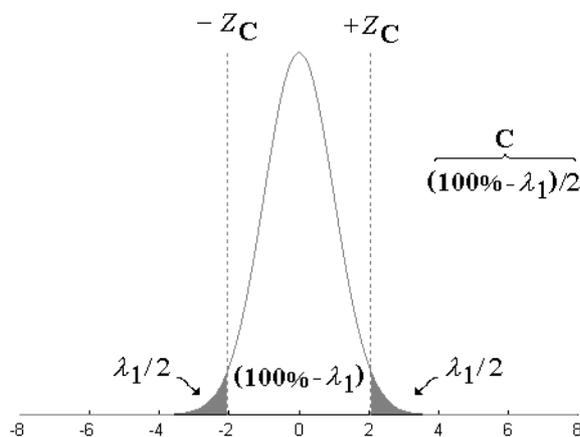
Para que os pontos salientes sejam identificados, é necessário encontrar, primeiramente, evidências nas imagens do envelope. Então, supõe-se que o valor absoluto dos coeficientes *Wavelet* de alta frequência,  $BA$ ,  $AB$  e  $AA$  seja diretamente proporcional à intensidade da borda em um dado sentido: horizontal, vertical e diagonal, respectivamente. A hipótese criada gera evidências através das bordas para os objetos de segmentação (bloco endereço, carimbo e selo), e também para ruídos presentes nos envelopes. Arbitra-se que a distribuição dos coeficientes *Wavelet* de alta frequência é uma distribuição normal (ou gaussiana)  $N(\mu, \sigma)$  com média zero (propriedade dos coeficientes *Wavelet* de detalhe ou alta frequência). Neste contexto, considera-se que os  $\lambda_1\%$  maiores coeficientes de detalhe contêm estas evidências, conforme ilustrado pela Figura 3.5. Então, mantendo-se somente  $\lambda_1\%$  dos coeficientes mais significativos no sentido da distribuição normal, conforme a Equação (3.3) [PRESS *et al* 1992], espera-se que estas evidências estejam presentes nesta amostra.

$$I_{wCS}(XX) = \left\{ \frac{I_w(XX) - I_{w\mu}(XX)}{I_{w\sigma}(XX)} < -Z_{(100\% - \lambda_1)/2} \right\} \cap \left\{ \frac{I_w(XX) - I_{w\mu}(XX)}{I_{w\sigma}(XX)} > +Z_{(100\% - \lambda_1)/2} \right\} \quad (3.3)$$

onde  $I_w(XX)$  e  $I_{wCS}(XX)$  são os conjuntos de todos os coeficientes e dos coeficientes mais significativos, respectivamente, em relação à banda de alta frequência  $XX$  ( $BA$ ,  $AB$  ou  $AA$ );  $I_{w\mu}(XX)$ ,  $I_{w\sigma}(XX)$  são a média e o desvio padrão referente a banda de alta

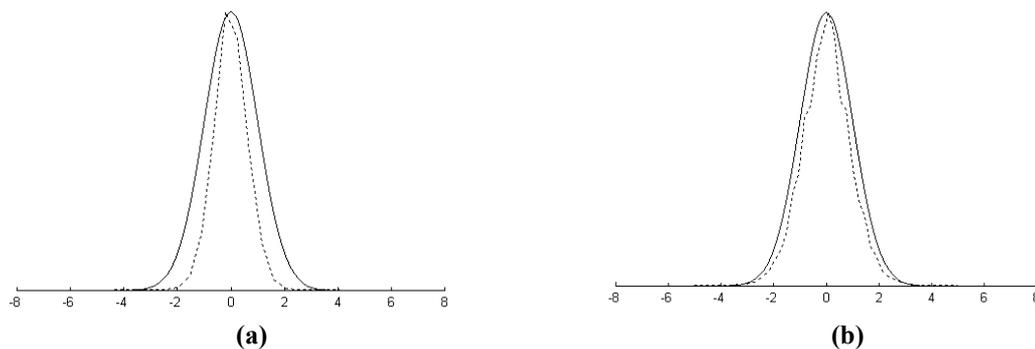
freqüência  $XX$ ;  $Z_{(100\%-\lambda_1)/2}$  é o ponto de probabilidade normalizada de nível de significância  $(100\% - \lambda_1)/2$ .

A Figura 3.5 ilustra uma função gaussiana, como uma função da distribuição de uma das bandas dos coeficientes *Wavelet* de alta freqüência, onde  $\lambda_1 = 4\%$  ( $Z_{(100\%-4\%)/2} = Z_{48\%} = 2,05$ ) é a área mais externa delimitada pela gaussiana e pelas duas retas pontilhadas, ou seja, a soma das duas áreas hachuradas na Figura 3.5 corresponde a 4% da distribuição, e estes são os 4% de coeficientes que armazenam a energia mais significativa desta distribuição no sentido da hipótese do valor absoluto dos coeficientes.



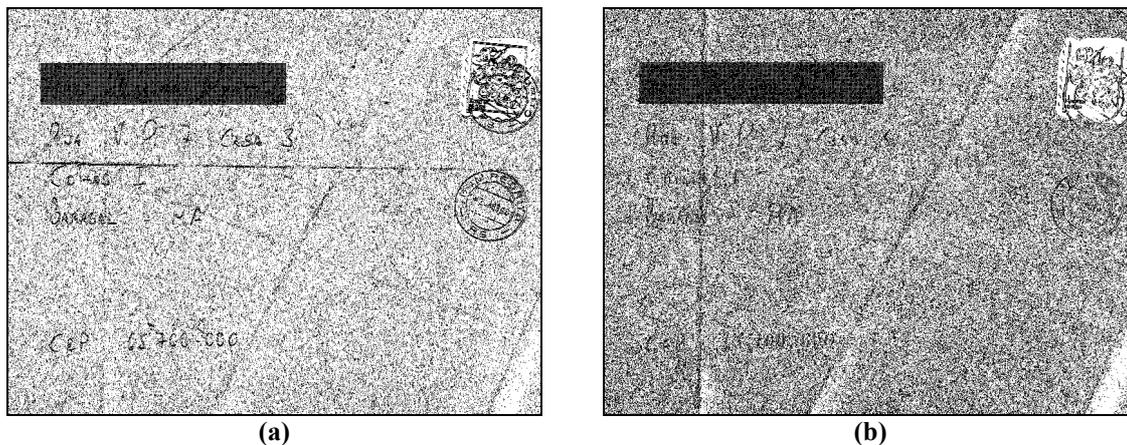
**Figura 3.5:** Gráfico típico de uma distribuição normal “Z” (ou gaussiana) com média zero e desvio padrão unitário; as retas indicam o ponto de corte  $(100 - \lambda_1)/2$ ; as duas áreas hachuradas correspondem aos  $\lambda_1\%$  dos coeficientes mais significativos.

Na Figura 3.6 pode-se observar que a suposição de distribuição normal (pontos contínuos) é semelhante à distribuição real dos coeficientes *Wavelet* (pontilhada). A distribuição normal mostrada na Figura 3.6 tem desvio padrão unitário e média zero, e a distribuição dos coeficientes *Wavelet* das Figura 3.6-a e b são normalizadas pelo desvio padrão unitário da própria distribuição e também são centrados em zero.



**Figura 3.6:** Gráfico das distribuições normal “Z” (pontilhadas) e dos coeficientes *Wavelet* (contínuas) de detalhe de uma imagem típica de envelope postal. (a) Distribuição dos coeficientes horizontais (*BA*); (b) Distribuição dos coeficientes verticais (*AB*).

Vale ressaltar que a decomposição *Wavelet* permite a redução do espaço dimensional, ou seja, do tamanho do espaço de dados. E levando-se em conta, também, que somente os  $\lambda_1\%$  dos coeficientes mais significativos no sentido da distribuição normal são mantidos, este passo permite uma redução significativa de informações a serem processadas pelo algoritmo de segmentação.

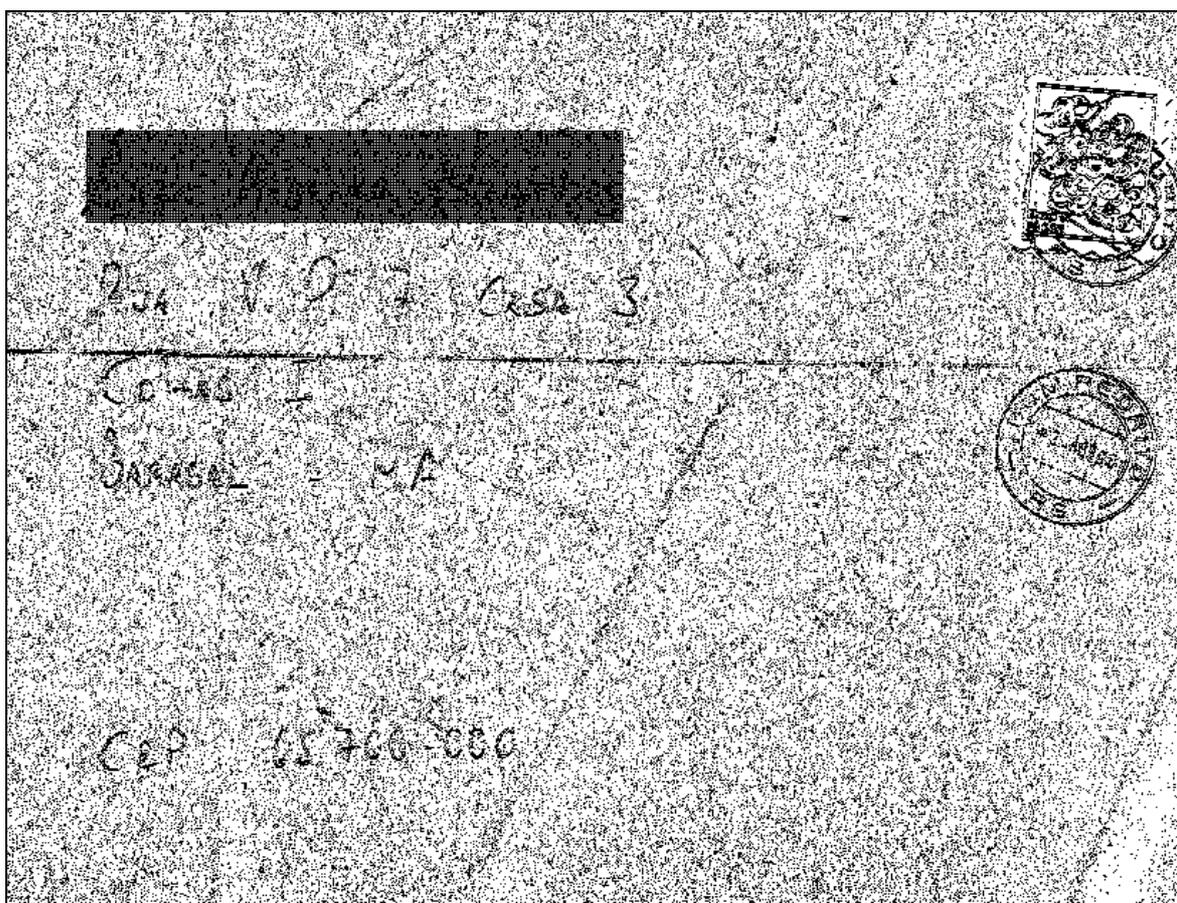


**Figura 3.7:** Coeficientes mais significativos (Imagem  $I_{WCS}$ ), com  $\lambda_1 = 43\%$ ; (a) Imagem  $I_{WCS(BA)}$ ; (b) Imagem  $I_{WCS(AB)}$ .

O resultado da seleção de coeficientes mais significativos,  $I_{WCS}$ , é uma imagem (ou matriz) binária, conforme ilustra a Figura 3.7, com dimensão  $n/2 \times m/2$  (linhas x colunas). Os pontos pretos desta imagem são aqueles  $\lambda_1\%$  dos coeficientes mais significativos no sentido da distribuição normal.

O leitor pode-se questionar o resultado apresentado na Figura 3.7, argumentando-se que os  $\lambda_1\%$  coeficientes esperados não foram realmente observados. Entretanto, nada foi

feito no sentido de ajustar a distribuição normal e, em distribuições pequenas o valor da distribuição normal é um superior seguro. Como não é garantido que exista uma distribuição conhecida única para todos os dados, optou-se por uma distribuição de comportamento previsível (a distribuição normal). Os experimentos, apresentados no próximo capítulo, mostram que a distribuição normal é adequada. Trabalhos futuros poderiam modelar uma distribuição típica do problema, contudo uma mudança no tipo da imagem (fundo, contexto da informação, etc.) poderia tornar restrita essa nova distribuição.



**Figura 3.8:** Resultado do segundo passo (imagem  $I_5$ ), intersecção entre as imagens da Figura 3.7.

Uma imagem típica de envelope contém vários componentes que são de interesse deste algoritmo de segmentação. Todos estes componentes – bloco endereço, selos, carimbos – estão envolvidos pelo fundo do envelope, ou seja, as regiões do fundo do envelope e dos objetos de segmentação fazem fronteira. Estas regiões de transição são fortes evidências da existência de objetos de segmentação. A transformada *Wavelet*, através de suas bandas direcionais obtidas pela decomposição de Mallat, pode representar

estas zonas de transição. As zonas de transição com forte representação serão denominadas, doravante, de pontos salientes.

Os conjuntos  $I_{WCS}$  são utilizados neste processo, pois tem um poder de representação de zonas de transição. O objetivo neste passo é identificar evidências através de bordas de regiões mais consistentes que são parecidas com o fundo do envelope ou texto (bloco endereço e carimbo) ou selo. Então, define-se o conjunto de pontos salientes  $I_S$  como sendo o conjunto de pontos que têm forte evidência de ser um detalhe. Mesmo que estes pontos possam ser somente simples manchas, neste instante do processo, todavia haverá outros passos, mais adiante, para verificar a consistência (eliminando ruídos) e incluir mais evidências. Esta identificação é realizada através da interseção posicional de dois canais de detalhes de  $I_{WCS}$ : o canal vertical ( $AB$ ) e o horizontal ( $BA$ ); isto é, onde quer que seja encontrada a presença (ponto a ponto) de detalhes horizontal e vertical (de  $BA$  e  $AB$ ), que foram mantidos segundo os  $\lambda_1\%$  dos coeficientes mais significativos, devem ser ditos pontos salientes. Os pontos diagonais (do canal diagonal  $AA$ ) são abandonados uma vez que são muito ruidosos e não contribuem muito com este procedimento. O procedimento de identificação de pontos salientes pode ser resumido segundo a Equação (3.4), e o resultado (imagem  $I_S$ ) é ilustrado pela Figura 3.8.

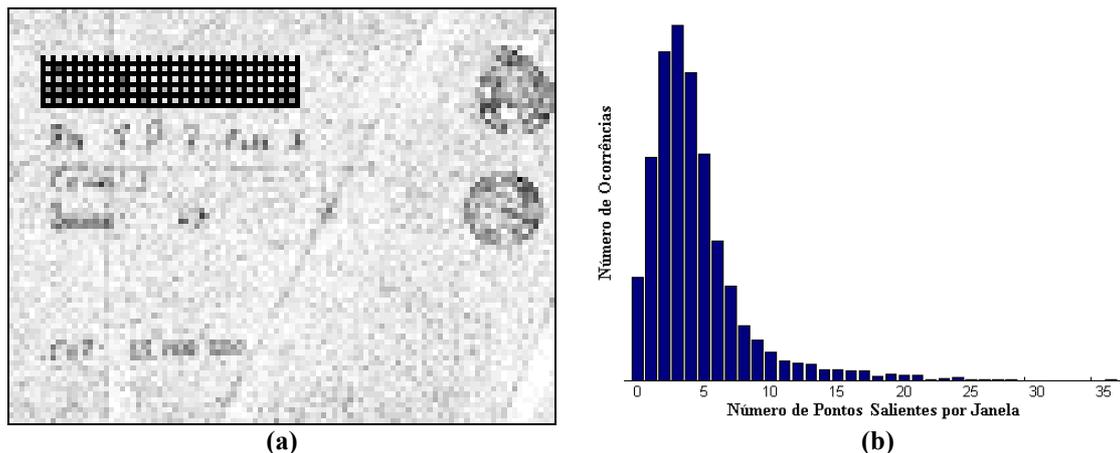
$$I_S \leftarrow I_{WCS(LH)} \cap I_{WCS(HL)} \quad (3.4)$$

A imagem de saída deste passo tem a mesma dimensão ( $n/2 \times m/2$ ) e também é binária como a imagem do passo anterior ( $I_{WCS}$ ). Estes dois passos iniciais reduzem muito a quantidade de informação a ser processada pelo algoritmo de segmentação, conforme pode ser observado pela quantidade de pontos salientes apresentados na imagem da Figura 3.8.

### 3.3 Rotulação Controlada das Janelas de Pontos Salientes

O conjunto de pontos salientes  $I_S$  é evidência de textura esparsa de componentes conectados, e isto é o objeto de segmentação deste trabalho. Alguns pontos salientes devem aparecer em uma determinada região com uma certa distribuição diferente da distribuição de outras regiões, isto é, basicamente deseja-se considerar dois tipos de regiões locais (janelas) de pontos salientes: uma região com alta densidade de presença de pontos salientes (medidas em uma janela quadrada de tamanho  $k$ ), e uma outra região com baixa densidade.

A Figura 3.9-a mostra uma imagem representando a densidade de pontos salientes em janelas da imagem  $I_S$  da Figura 3.8. As janelas (que são representadas por *pixels*) mais escuras da Figura 3.9-a são regiões de alta densidade enquanto que as janelas claras são de baixa densidade. Esta imagem de densidade/concentração foi gerada segundo uma janela de  $8 \times 8$  (pontos salientes). É possível estudar o efeito da granularidade, isto é, a função de densidade de probabilidade em relação ao tamanho da janela escolhida; mediante os experimentos a janela de  $8 \times 8$  mostrou-se ser uma boa escolha. Cada janela é representada pelo número de pontos salientes existentes dentro da mesma segundo a imagem  $I_S$ . O histograma de pontos salientes em janelas da imagem da Figura 3.9-a é apresentado na Figura 3.9-b, isto é, cada elemento  $i$  deste histograma, representa a ocorrência de janelas com  $i$  pontos salientes.



**Figura 3.9: Janela de pontos salientes. (a) Imagem representativa das janelas de pontos salientes ( $n/2k \times m/2k$ ), agrupados por uma janela de  $k \times k$  ( $8 \times 8$ ); (b) Histograma das janelas de pontos salientes de (a).**

As regiões com alta densidade de pontos salientes são mais prováveis de serem regiões conectadas e objetos de segmentação, enquanto que as outras regiões, de baixa concentração, passam a ser consideradas como ruído. A decisão de estabelecer um limite entre estas regiões de alta densidade e de baixa densidade, não pode ser feita de modo fixo, por um limiar, visto que para cada imagem de envelope os *pixels* da imagem e suas distribuições mudam muito, por diversos fatores como: iluminação a qual o envelope está exposto, o tipo de fundo de envelope, a espessura da caligrafia, o tipo de selo, etc. Para resolver este problema foi desenvolvido um algoritmo iterativo de controle especial para realizar testes de significância estatística [PRESS *et al* 1992] sobre as janelas de pontos salientes.

O algoritmo de controle desenvolvido executa os testes de hipótese de forma global, isto é, o algoritmo não leva em consideração a posição das janelas dentro da imagem. Por isto, este algoritmo pode ser visto como um problema global de binarização do histograma das janelas de pontos salientes, onde se deseja identificar duas regiões distintas.

Estas duas regiões são determinadas de forma isolada, isto é, uma região não interfere na construção da outra. O teste de hipótese é utilizado para verificar quais janelas pertencem a uma dada classe/região (de alta ou baixa densidade) em cada iteração.

Se, em um nível de significância  $\lambda_2\%$ , duas distribuições de janelas – uma da região/classe e outra do conjunto local – mostrarem uma diferença não significativa entre

elas, então o conjunto local deve ser incorporado à região/classe, e a Equação (3.5) deve ser mantida, isto é,

$$\frac{|\mu_R - \mu_L|}{\sqrt{\frac{\sigma_R^2}{n_R} + \frac{\sigma_L^2}{n_L}}} \leq Z_{\lambda_2/2} \quad (3.5)$$

onde  $\mu_L$  é a média e  $\sigma_L^2$  é a variância local dos pontos salientes de cada janela, e são calculados de acordo com as Equações (3.6) e (3.7) respectivamente,  $n_L$  são os números de amostras possíveis dentro de uma janela ( $k^2$ );  $\mu_R$  é a média e  $\sigma_R^2$  é a variância das regiões rotuladas e são calculadas pelas Equações (3.8) e (3.9) respectivamente,  $n_R$  é o número de amostras contido na região;  $Z_{\lambda_2/2}$  é o ponto de probabilidade normalizada para um nível de significância  $\lambda_2\%$ .

As equações abaixo são utilizadas no cálculo das medidas locais e das distribuições das regiões, ou seja, a média local dos pontos salientes de cada janela:

$$\mu_L(n_S) = \frac{n_S}{k^2}, \quad (3.6)$$

o desvio padrão local dos pontos salientes de cada janela:

$$\sigma_L(n_S) = \sqrt{\frac{\sum_{i=1}^{k^2} [x_i - \mu_L(n_S)]^2}{k^2}} = \frac{n_S}{k^2} - \frac{n_S^2}{k^4}, \quad (3.7)$$

onde:  $n_S$  é o número de pontos salientes existentes dentro da janela;  $k$  é o tamanho da janela de  $k^2$  elementos;  $x_i$  é o  $i$ -ésimo elemento da janela que pode (1) ou não (0) ter sido selecionado como ponto saliente. A média das regiões rotuladas:

$$\mu_R = \left( \sum_{i=R_{\min}}^{R_{\max}} i \cdot hist[i] \right) / \left( k^2 \cdot \sum_{i=R_{\min}}^{R_{\max}} hist[i] \right), \quad (3.8)$$

e o desvio padrão das regiões rotuladas:

$$\sigma_R^2 = \left( \sum_{i=R_{\min}}^{R_{\max}} (i - \mu_R)^2 \cdot hist[i] \right) / \left( k^2 \cdot \sum_{i=R_{\min}}^{R_{\max}} hist[i] \right). \quad (3.9)$$

na qual  $i$  é um par de histograma das janelas;  $hist[i]$  é o número de ocorrências de janelas de tamanho  $k$ , e  $k^2$  elementos, com  $i$  pontos salientes contidos nela;  $R_{\min}$  e  $R_{\max}$  são os limites inferiores e superiores, respectivamente, da região rotulada.

A rotulação em duas classes de regiões de pontos salientes é inicializada pelos extremos do histograma. A classe de maior densidade é inicializada pela extrema direita do histograma, ou seja, pela janela de maior média local  $\mu_A$  (isto é, a janela que possui o maior número de pontos salientes), e a classe menor densidade é inicializada pela extrema esquerda do histograma, ou seja, pela janela de menor média local  $\mu_B$  (isto é, a janela que possui o menor número de pontos salientes) acima de zero.

O algoritmo começa a execução em ambos os níveis, da base, pelas regiões/janelas de baixa densidade de pontos salientes, e do topo, pela região/janela de maior densidade de pontos salientes, ao mesmo tempo em que vai decidindo os rótulos das janelas/regiões gerando assim o novo conjunto  $I_{RCW}$  com as janelas rotuladas como de alta densidade e descartando as janelas/regiões rotuladas como sendo de baixa densidade. Caso alguma janela tenha ficado sem rótulo após a execução do algoritmo por toda a imagem (todo histograma), os valores das médias,  $\mu_A$  e  $\mu_B$ , das variâncias,  $\sigma_A^2$  e  $\sigma_B^2$ , e dos número de amostras,  $n_A$  e  $n_B$ , são atualizados segundo o conjunto de janelas já rotuladas, e, portanto, novas médias e variâncias são calculadas, bem como o número de amostras dentro de cada região. O algoritmo é executado até que todas as regiões/janelas estejam rotuladas como sendo regiões de alta e de baixa densidade. Caso exista interseção entre as regiões/janelas, isto é, caso existam janelas que estejam na região de confusão das duas regiões rotuláveis, toma-se a decisão de que estas janelas serão rotuladas como sendo janelas de baixa densidade de pontos salientes e por conseqüência são consideradas como ruído. Depois deste passo a imagem  $I_{RCW}$  deve conter somente as janelas de alta densidade de pontos salientes. Um exemplo ilustrando este conjunto pode ser visto na Figura 3.11-a.

```

// Ini é o primeiro elemento do histograma (acima de zero)
// Max é o último elemento do histograma com ocorrência
// Inicialização
 $\mu_B \leftarrow \mu_L(\text{Ini}), \text{var}_B \leftarrow \text{var}_L(\text{Ini}), n_B \leftarrow k^2, \text{RMin} \leftarrow 0;$ 
 $\mu_A \leftarrow \mu_L(\text{Max}), \text{var}_A \leftarrow \text{var}_L(\text{Max}), n_A \leftarrow k^2, \text{RMax} \leftarrow \text{Max};$ 
// Algoritmo Iterativo
faça
// Teste de hipótese para região de baixa densidade
para  $i$  de RMin até Max passo 1 faça:
    se TestHipo(L,  $i$ ) >  $Z\lambda^2/2$  então
        RMax  $\leftarrow i-1;$ 
        sai loop;
    fim-se
fim-para
// Teste de hipótese para região de alta densidade
para  $i$  de RMax até 1 passo -1 faça:
    se TestHipo(H,  $i$ ) >  $Z\lambda^2/2$  então
        RMin  $\leftarrow i+1;$ 
        sai loop;
    fim-se
fim-para
// médias, variâncias e número de amostras
atualize dados por região segundo RB's e RA's
enquanto (RMax < RMin)
// Intersecção de regiões é tomada por área baixa
RMin  $\leftarrow$  RMax + 1;

```

Figura 3.10: Pseudo-código do algoritmo de rotulação controlada de janelas de pontos salientes.

Um resumo para melhor entendimento deste passo é apresentado na Figura 3.10 na forma de um pseudo-código do algoritmo.

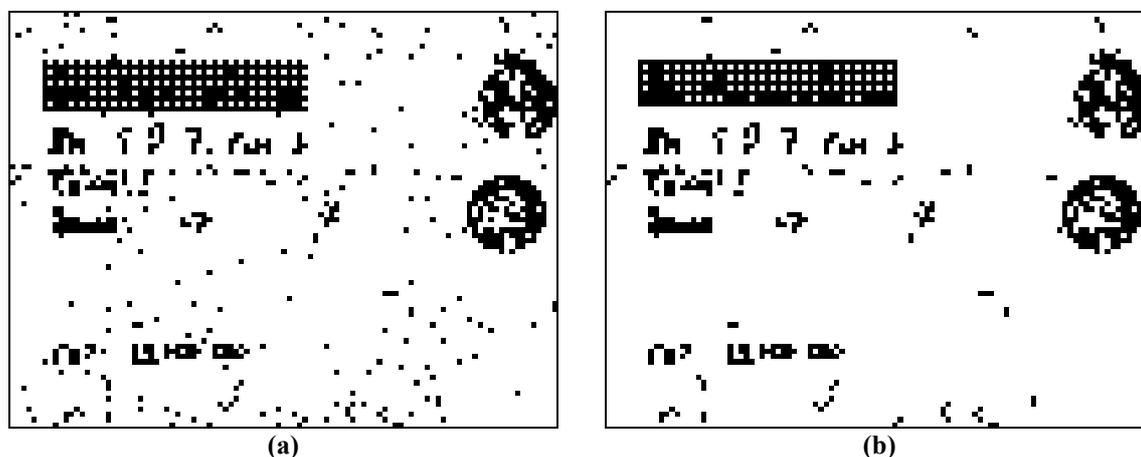


Figura 3.11: Saída do terceiro passo ( $I_{RC}$  - dimensão  $n/2k \times m/2k$ ). (a) Conjunto  $I_{RCW}$ , imagem das janelas de alta densidade de pontos salientes; (b) Conjunto  $I_{RC}$ , removidas as janelas sem conexão com nenhuma outra janela vizinha.

As janelas que não estão conectadas a nenhuma outra janela em sua vizinhança são removidas do conjunto  $I_{RCW}$ , eliminando supostos ruídos, gerando assim o conjunto  $I_{RC}$  (Figura 3.11-b) de saída do terceiro passo. Este último conjunto deve estar quase isento de

evidências que apontem para o fundo ou ruído. Os ruídos que ainda persistirem serão barrados/bloqueados nos próximos passos.

### 3.4 Projeção Reversa dos Pontos Salientes Selecionados

O conjunto  $I_{RC}$  é uma imagem de  $n/2k \times m/2k$  representando janelas que contém os pontos salientes que pertencem a uma região dita de alta densidade de pontos salientes. Cada ponto saliente, por sua vez, tem quatro filhos relacionados à imagem original. A Figura 3.12 ilustra essa projeção reversa.

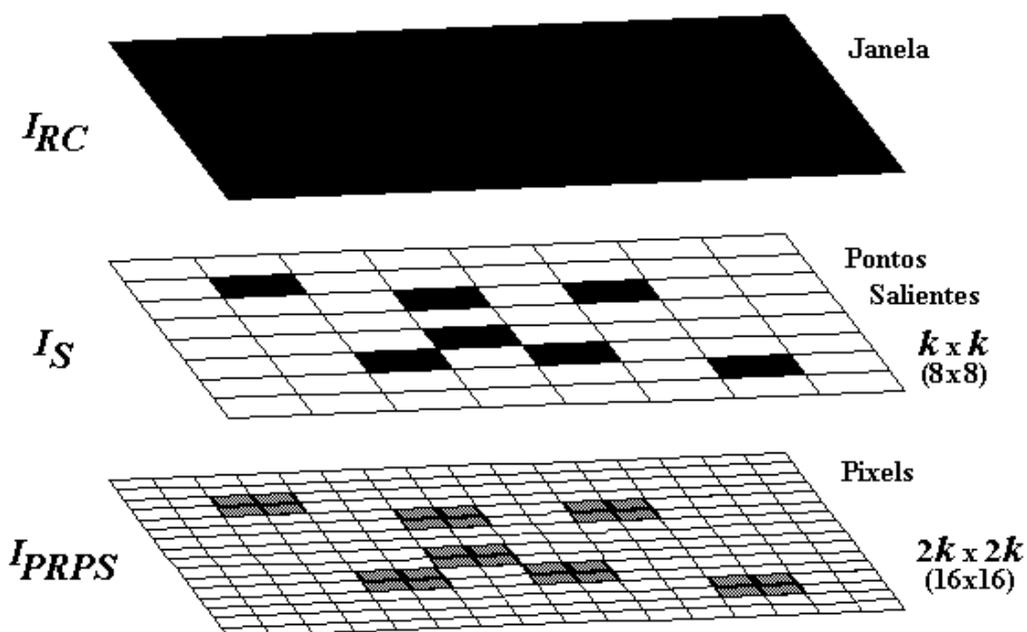
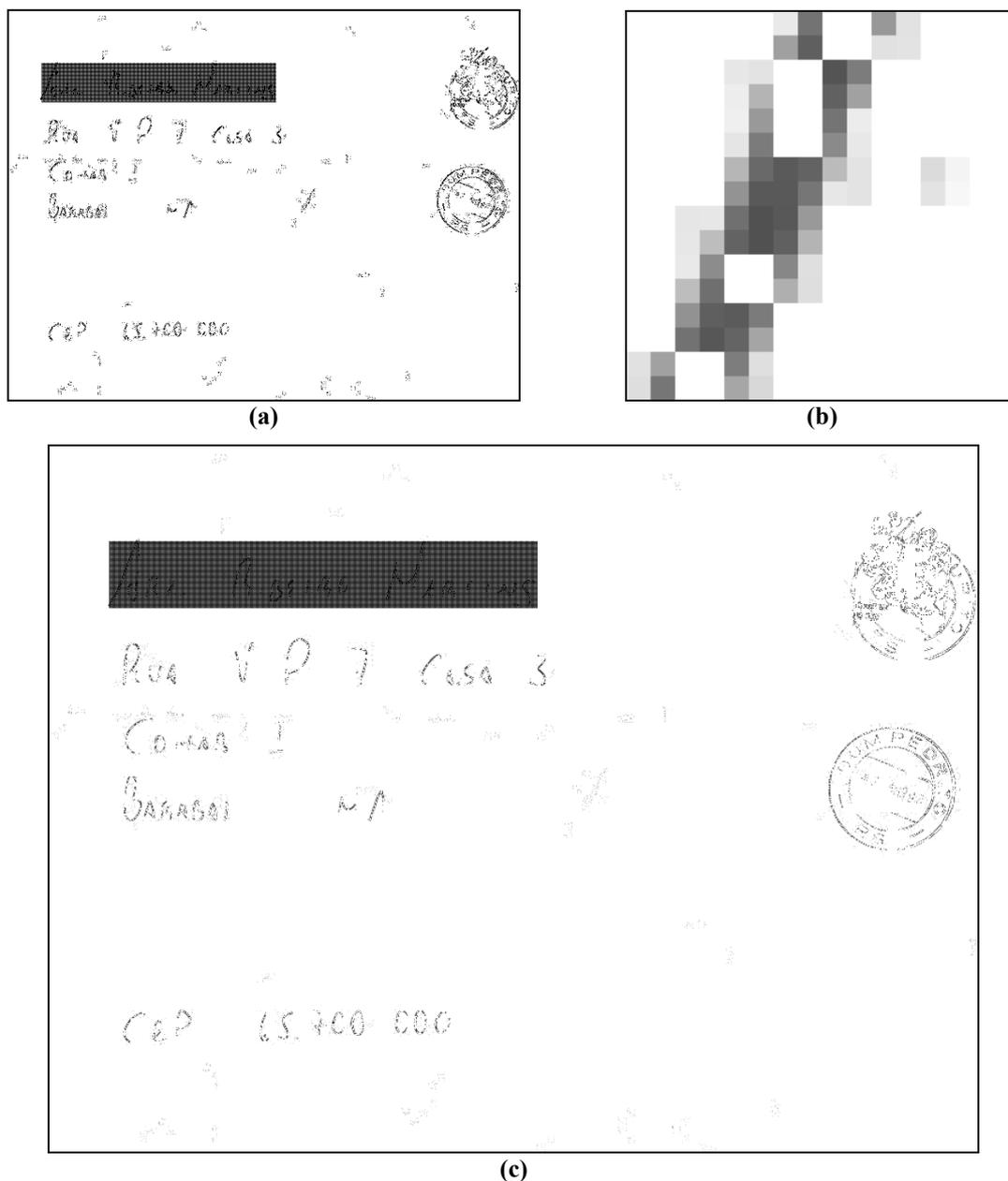


Figura 3.12: Ilustração da projeção reversa dos pontos salientes de janelas de alta densidade na imagem original em tons de cinza.

Pode-se observar através desta figura que uma janela de tamanho  $k$  possui  $k^2$  elementos relacionados a imagem de intersecção  $I_S$ . Cada um destes elementos está relacionado com quatro filhos da imagem original em tons de cinza. Todavia, somente os pontos salientes identificados no conjunto  $I_S$  e rotulados pelo conjunto  $I_{RC}$  serão utilizados para a projeção reversa. Estes pontos salientes são ilustrados pela Figura 3.13-a, e representam os pontos salientes rotulados pelo conjunto  $I_{RC}$ , e esta imagem

representativa possui dimensão  $n/2 \times m/2$ . A diferença entre o conjunto representado pela imagem  $I_S$  da Figura 3.8 e o conjunto de pontos representados na Figura 3.13-a são os ruídos eliminados no passo anterior.



**Figura 3.13: Projeção reversa dos pontos salientes. (a) Conjunto de pontos salientes correspondentes as janelas de alta densidade ( $n/2 \times m/2$ ); (b) Exemplo de uma janela de alta densidade de pontos salientes projetados reversamente na imagem original ( $2k \times 2k$ ); (c) Conjunto  $I_{PRPS}$  ( $n \times m$ ).**

A projeção reversa destes quatro filhos de cada ponto saliente, exemplificados pelo conjunto da Figura 3.13-a, origina o conjunto  $I_{PRPS}$  (Figura 3.13-c) com dimensão igual à imagem original ( $n \times m$ ). Este conjunto é composto pelos *pixels* com o nível de cinza da

imagem original. Os *pixels* que não são originários de pontos salientes rotulados como de alta densidade são inicializados como não pertencentes ao conjunto  $I_{PRPS}$ . Um exemplo de uma janela de  $2k \times 2k$  elementos de alta densidade de pontos salientes, projetada reversamente, é apresentada na Figura 3.13-b.

### 3.5 Perseguição de Contorno através dos Pontos Salientes

Nesta etapa a imagem  $I_{PRPS}$  deve conter evidências selecionadas para cada objeto de segmentação, isto é, os *pixels* provavelmente pertençam a uma das classes: bloco endereço, carimbo, ou selos. Visto que a grande maioria do fundo já está selecionada pelo complemento de  $I_{PRPS}$ , tendo como universo a imagem original. Desta forma, esta evidência deve ser usada apropriadamente para encontrar no resto da imagem somente os *pixels* em tons de cinza que são coerentes à imagem  $I_{PRPS}$ , bem como a idéia de que regiões locais devem compartilhar atributos similares.

Uma decisão, sobre quais pontos devem ser finalmente selecionados, deve ser estabelecida. Esta decisão é feita levando-se em conta duas hipóteses:

1ª Hipótese: uma informação de contexto muito forte; considera-se que os objetos de segmentação (bloco endereço, carimbos e selos) têm uma intensidade mais intensa que o fundo do envelope postal, ou seja, os tons de cinza dos objetos de segmentação estão mais próximos do preto que o fundo. Esta é uma hipótese que leva em conta a distribuição dos tons de cinza de uma forma global. Para isto, faz-se necessário supor uma distribuição para a imagem do envelope postal. Esta suposição pode ser realizada baseando-se na média  $I_\mu$  e no desvio padrão  $I_\sigma$  da imagem original, como sendo uma distribuição gaussiana  $N(I_\mu, I_\sigma)$ . Um exemplo de distribuição (ou histograma) dos tons de cinza de uma imagem original e de sua imagem suposta podem ser observadas na Figura 3.14-a e b, respectivamente. Dados estes dois histogramas, pode-se dizer que esta hipótese é aceitável, baseando-se na semelhança existente entre os dois. Entretanto, este fato só será confirmado

com os resultados que serão vistos no próximo capítulo. A argumentação para isto é a mesma apresentada na Seção 3.2;

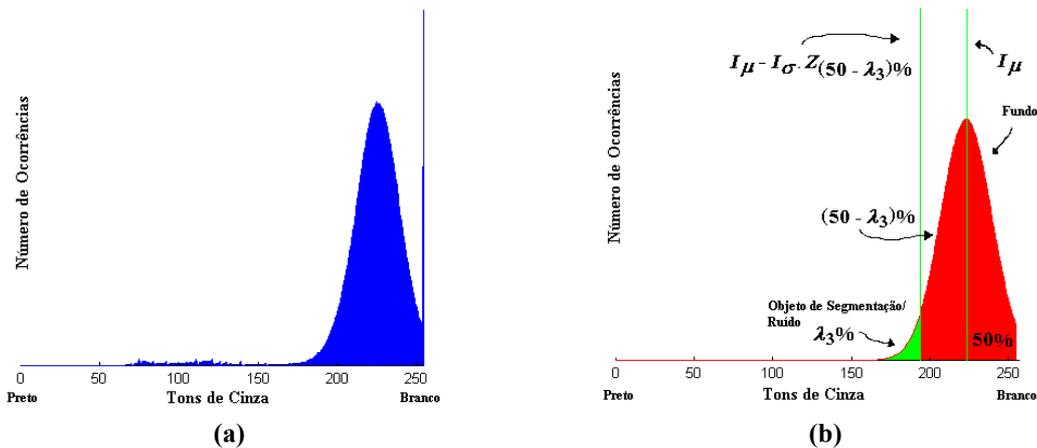


Figura 3.14: Informação de Contexto. (a) Histograma de um envelope típico; (b) Histograma suposto  $(N(I_\mu, I_\sigma))$  de um envelope, com objetos de segmentação em  $\lambda_3\%$ .

2ª Hipótese: Supõe-se que os pontos salientes representem regiões (de  $2 \times 2$  pixels) de transição, isto é, possuem um alto grau de energia (valor absoluto) no sentido dos coeficientes *Wavelet* de alta frequência ( $AB$  e  $BA$ ), e são a transição entre as regiões de objeto de segmentação e o fundo conforme ilustra a Figura 3.15.

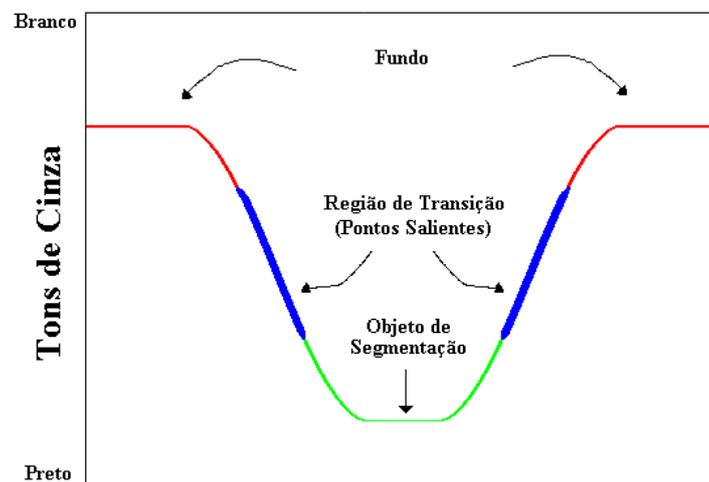
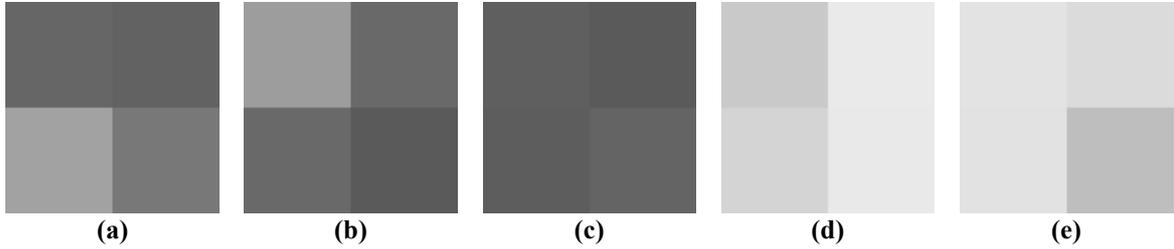


Figura 3.15: Transição entre o objeto de segmentação e o fundo; os pontos salientes. Um corte transversal ilustrativo em uma imagem típica de envelope.

Estes pontos salientes são evidência para os objetos de segmentação (Figura 3.16-a,b,c). Entretanto, alguns destes pontos salientes podem ainda ter filhos ( $2 \times 2$  pixels) que

apontam para o fundo do envelope (Figura 3.16-d,e) e estes últimos devem pertencer a região de fundo de envelope, segundo a imagem estimada  $N(I_\mu, I_\sigma)$ .



**Figura 3.16: Os filhos (2x2 pixels) de pontos salientes de janelas de alta densidade. (a), (b) e (c) Evidência para objeto de segmentação; (d) e (e) Evidência para o fundo do envelope.**

Baseando-se nestas duas hipóteses, foi desenvolvido o passo final de segmentação. Este tem início com a imagem  $I_{PRPS}$ , de cada um dos  $2 \times 2$  pixels ( $I_{SET}(i, j)$ ,  $I_{SET}(i+1, j)$ ,  $I_{SET}(i, j+1)$ ,  $I_{SET}(i+1, j+1)$ ) referenciados pelos pontos salientes. O pixel com o menor tom de cinza – mais próximo ao objeto de segmentação conforme ilustra a Figura 3.15 – deve ser selecionado, como pertencente ao contorno do objeto de segmentação, gerando assim a imagem  $I_{CONTORNO}$ , caso um valor limite local fique dentro dos  $\lambda_3\%$  - região dos objetos de segmentação - da distribuição da imagem estimada  $N(I_\mu, I_\sigma)$ , isto é,

$$\begin{aligned} \text{se } LimLocal(\alpha) < I_\mu - Z_{50\%-\lambda_3} \cdot I_\sigma \\ \text{então } I_{CONTORNO} \leftarrow I_{SET}(\text{menor}) \end{aligned} \quad (3.10)$$

onde  $I_\mu$  é a média global da imagem;  $I_\sigma$  é o desvio padrão global da imagem; e  $Z_{50\%-\lambda_3}$  é o ponto de probabilidade normalizada para  $(50\% - \lambda_3)$  da distribuição;  $LimLocal(\alpha)$  é um valor local baseado nos 4 pixels filhos do conjunto  $I_{SET}$  referenciados pelo ponto saliente  $\alpha$ . A Figura 3.17-a ilustra um pequeno recorte de uma imagem típica com o conjunto,  $I_{CONTORNO}$ , de pixels inicializados.

O valor que limita a inicialização dos pixels,  $LimLocal(\alpha)$ , referentes aos pontos salientes é baseado em arranjos entre os  $2 \times 2$  pixels do conjunto  $I_{SET}$ . Este valor é usado para determinar se o ponto saliente dos quais esses pixels são originados pertencem ao

fundo ou não. Os arranjos propostos aqui são: (1) o menor dos quatro filhos (Equação (3.11)); (2) o segundo menor dos quatro filhos (Equação(3.12)); (3) a média entre os três menores filhos (Equação(3.13)); (4) a média entre os quatro filhos (Equação (3.14)); (5) o maior dos quatro filhos (Equação (3.15)). Esses arranjos acima são apresentados de forma ordenada ascendente.

$$LimLocal(\alpha) = I_{SET}(menor) \quad (3.11)$$

$$LimLocal(\alpha) = I_{SET}(2^\circ menor) \quad (3.12)$$

$$LimLocal(\alpha) = (I_{SET}(menor) + I_{SET}(2^\circ menor) + I_{SET}(3^\circ menor))/2 \quad (3.13)$$

$$LimLocal(\alpha) = (I_{SET}(i, j) + I_{SET}(i, j + 1) + I_{SET}(i + 1, j) + I_{SET}(i + 1, j + 1))/2 \quad (3.14)$$

$$LimLocal(\alpha) = I_{SET}(maior) \quad (3.15)$$

Os resultados, apresentados no próximo capítulo, mostram qual é a escolha de  $LimLocal(\alpha)$  mais adequada. Ressalte-se que os *pixels* filhos, do conjunto  $I_{SET}$ , são ordenados segundo o seu tom de cinza, e que o tom de cinza é mais intenso quanto mais claro (ou branco) ele for. E, também é importante observar que somente um dos quatro *pixels*, inicialmente, será selecionado como pertencentes ao contorno (ou como objeto de segmentação), ou seja, cada ponto saliente da janela de alta densidade referencia somente uma única evidência em  $I_{CONTORNO}$  em tons de cinza. Cada um destes *pixels* inicializados carrega consigo o valor de  $LimLocal(\alpha)$  ( $\alpha$  é o ponto saliente que referencia os 2 x 2 *pixels* filhos) para ser utilizado como um limitador local do algoritmo de perseguição do contorno.

Depois de inicializado o algoritmo de perseguição de contorno, o passo seguinte deste algoritmo consiste em localizar todos os outros *pixels* que pertencem aos objetos de segmentação. Esta decisão é feita entre os *pixels* pertencentes ao contorno (inicialmente os pontos inicializados ( $I_{SET}(menor)$ ) e seus *pixels* vizinhos. Os vizinhos dos *pixels* do contorno pertencerão ao objeto de segmentação (ou contorno) caso a Equação (3.18) seja mantida, então seja a razão:

$$\beta = \frac{|PIXEL - VIZ|}{\max(PIXEL, VIZ)}, \quad (3.16)$$

e a razão:

$$\varepsilon = \frac{|LimLocal(\alpha) - VIZ|}{\max(LimLocal(\alpha), VIZ)}. \quad (3.17)$$

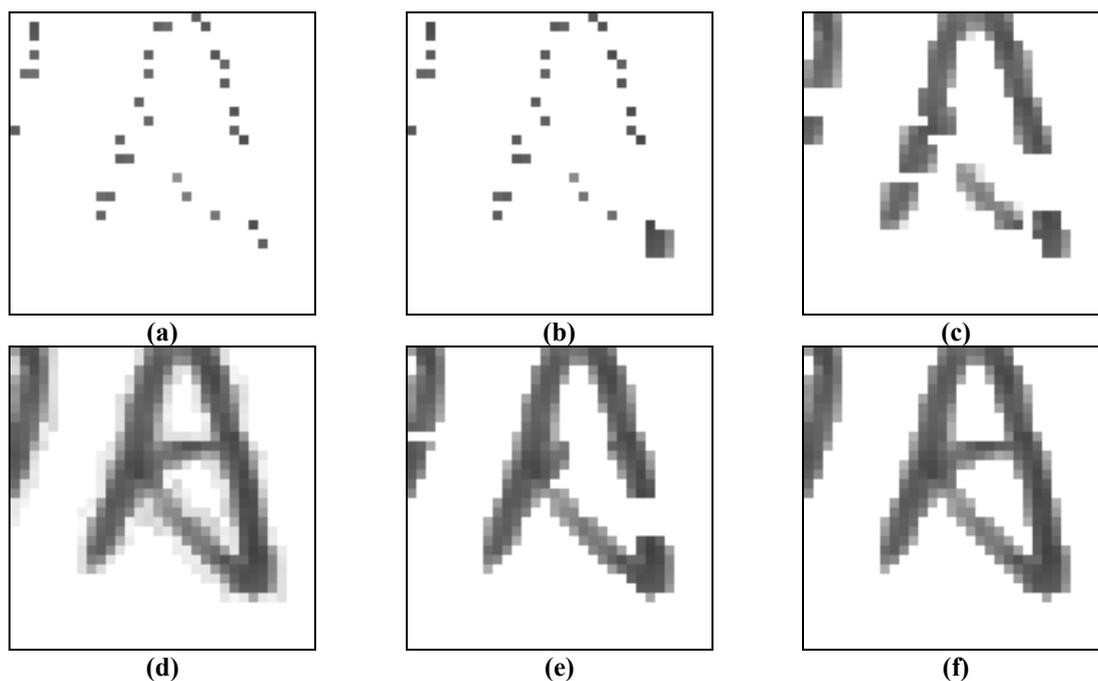
que são utilizadas para o cálculo da Equação (3.18).

$$\begin{aligned} \text{se } \beta \leq \varepsilon \text{ então } I_{VIZ} \text{ é selecionado} \\ \text{caso contrário } I_{VIZ} \text{ é descartado} \end{aligned} \quad (3.18)$$

onde *PIXEL* é um *pixel* pertencente ao conjunto  $I_{CONTORR}$ , *VIZ* é um *pixel* vizinho de *PIXEL*, em uma vizinhança 3x3 (com 8 elementos vizinhos); *LimLocal*( $\alpha$ ) é o limitador local armazenado pelo *pixel* referenciado pelo ponto saliente  $\alpha$  inicial do contorno local; a função  $\max(x, y)$  indica o maior valor entre  $x$  e  $y$ .

As razões  $\beta$  (Equação (3.16)) e  $\varepsilon$  (Equação (3.17)) são escolhidas de forma que os vizinhos, *VIZ*, do *pixel* pertencentes a um contorno local (micro região), respeitem um limite local, *LimLocal*( $\alpha$ ), determinado de forma automática para cada região (2 x 2 *pixels*) da imagem. Desta forma o algoritmo pode ser executado em diferentes imagens e diferentes fundos sem nenhuma intervenção manual. Caso a Equação (3.18) não seja verificada, o *pixel* é desprezado e, o algoritmo continua a execução analisando o próximo *pixel* da fila. Este algoritmo é finalizado quando não existirem mais *pixels* na fila a serem processados, ou de outro ponto de vista, que todos os *pixels* que devam pertencer ao contorno sejam incorporados.

Os *pixels* vizinhos que mantêm a Equação (3.18) são ditos pertencentes ao contorno e são armazenados em uma fila para que sejam processados posteriormente. A perseguição do contorno, então, é realizada segundo uma pesquisa em largura, ou seja, cada novo *pixel* inserido no contorno só será processado depois que todos os outros *pixels* que estavam na fila sejam processados. Este processamento de pesquisa em largura é necessário para evitar que limites locais de uma micro região não influenciem em outras micro regiões. Os novos *pixels* vizinhos que farão parte do contorno herdarão o limitador local, *LimLocal*( $\alpha$ ), do *pixel* inicial do contorno e serão colocados no final da fila.



**Figura 3.17:** Um pequeno recorte ilustrando o algoritmo de perseguição de contorno, utilizando  $LimLocal(\alpha)$  como sendo a média dos quatro filhos. (a) Inicialização do contorno; (b) Incorporação dos vizinhos; (c)-(d) Evolução e finalização do contorno sem restrição global; (e)-(f) Evolução e finalização do contorno com restrição global.

A primeira hipótese utilizada pelo algoritmo de perseguição de contorno, apresentada no começo desta seção, pode ser estendida para a evolução do contorno como sendo uma restrição global ao crescimento do contorno. Esta restrição pode ser feita adaptando-se a Equação (3.10) para utilizar o *pixel* vizinho em questão,  $I_{VIZ}$ , em vez de usar o limitador local  $LimLocal(\alpha)$  gerando assim a Equação (3.19). Esta restrição impede que o contorno agregue elementos pertencentes ao fundo, conforme pode ser observado em Figura 3.17-f, o que não ocorre quando esta restrição não é imposta (Figura 3.17-d).

$$\begin{aligned} \text{se } I_{VIZ} < I_{\mu} - Z_{(50\%-\lambda_3)} \cdot I_{\sigma} \text{ então } I_{VIZ} \text{ é selecionado} \\ \text{caso contrário } I_{VIZ} \text{ é descartado} \end{aligned} \quad (3.19)$$

A Figura 3.17 ilustra a inicialização e evolução do algoritmo de perseguição de contorno em um pequeno recorte feito em uma imagem de envelope. Primeiramente os pontos salientes estão referenciando quatro *pixels* filhos na imagem original em tons de cinza. Estas referências feitas pelos pontos salientes aos *pixels* só serão efetivadas – iniciarão o contorno – caso a Equação (3.10) seja verificada. Uma ilustração dos *pixels* inicializados está na Figura 3.17-a. Cada um destes *pixels* inicializados são colocados em

uma fila juntamente com o seu limitador local,  $LimLocal(\alpha)$ , que é utilizado para restringir o crescimento de cada micro região. Estas micro regiões são determinadas por cada *pixel* inicializado que foi referenciado por um ponto saliente, isto é, cada ponto saliente determina uma micro região de crescimento do contorno. O conjunto destas micro regiões – muitas delas conectadas – são o contorno, que representam os objetos de segmentação. Os *pixels* são retiradas da fila na ordem em que foram inseridos e, agora, os vizinhos de cada *pixel* são avaliados segundo a condição da Equação (3.18). Comparando a Figura 3.17-a e a Figura 3.17-b, pode-se notar como um dos *pixels* inicializados incorpora os seus *pixels* vizinhos ao contorno. A evolução deste procedimento pode ser visualizada na Figura 3.17-c (sem restrição global) e na Figura 3.17-e (com restrição global).

```

// inicialização do contorno
para cada 2x2 pixels em IPRPS faça:
    se  $LimLocal(\alpha) < I\mu - Z(50\%-\lambda_3) \cdot I\sigma$  então
        ICANTORNO  $\leftarrow$  ISET(MENOR);
        // a fila armazena a posição do pixel
        // e o valor de  $LimLocal(\alpha)$ 
        FILACANTORNO  $\leftarrow$  ISET(MENOR);
    fim-se
fim-para
// algoritmo de perseguição de contorno (evolução)
enquanto (FILACANTORNO não esta vazia) faça:
    PIXEL  $\leftarrow$  Remove(FILACANTORNO);
    para cada vizinho (VIZ) dos 8 vizinhos de PIXEL faça:
        beta  $\leftarrow$   $|PIXEL - VIZ| / \max(PIXEL, VIZ)$ ;
        epsilon  $\leftarrow$   $|LimLocal(\alpha) - VIZ| / \max(LimLocal(\alpha), VIZ)$ ;
        se beta  $\leq$  epsilon e
             $LimLocal(\alpha) < I\mu - Z(50\%-\lambda_3) \cdot I\sigma$  então
                ICANTORNO  $\leftarrow$  VIZ;
                FILACANTORNO  $\leftarrow$  VIZ;
        fim-se
    fim-para
fim-enquanto

```

Figura 3.18: Pseudo-código do algoritmo de perseguição de contorno.

Novamente, pode-se observar pela comparação entre as Figura 3.17-d (contorno final sem restrição global) e Figura 3.17-f (contorno final com restrição global) o efeito da imposição global que foi estendida à evolução do contorno, ou seja, alguns *pixels* do fundo do envelope são incorporados ao contorno quando esta imposição não é feita. Então, a evolução do contorno fica dependente da condição da Equação (3.18) e da restrição global imposta pela condição da Equação (3.19), gerando a Equação (3.20), que rege a evolução do algoritmo de perseguição de contorno.

se  $(\beta \leq \epsilon)$  e  $(I_{VIZ} < I_{\mu} - Z_{(50\%-\lambda_3)} \cdot I_{\sigma})$  então  $I_{VIZ}$  é selecionado  
 caso contrário  $I_{VIZ}$  é descartado (3.20)

A Figura 3.18 mostra um pseudo-código que ilustra o funcionamento do algoritmo de perseguição de contorno apresentado nesta seção.

Após executado o algoritmo de perseguição de contorno, o conjunto  $I_{CONTORNO}$ , contém os tons de cinza de todos os *pixels* que foram segmentados (Figura 3.19-a). Entretanto, esta não é a representação recomendada para uma segmentação. Então, todos os *pixels* que foram rotulados/classificados/segmentados como bloco endereço, selo, ou carimbo são transformados em *pixels* pretos de uma imagem binária. Gerando, desta forma, a imagem de saída do algoritmo de segmentação proposto  $I_{FINAL}$  (Figura 3.19-b).

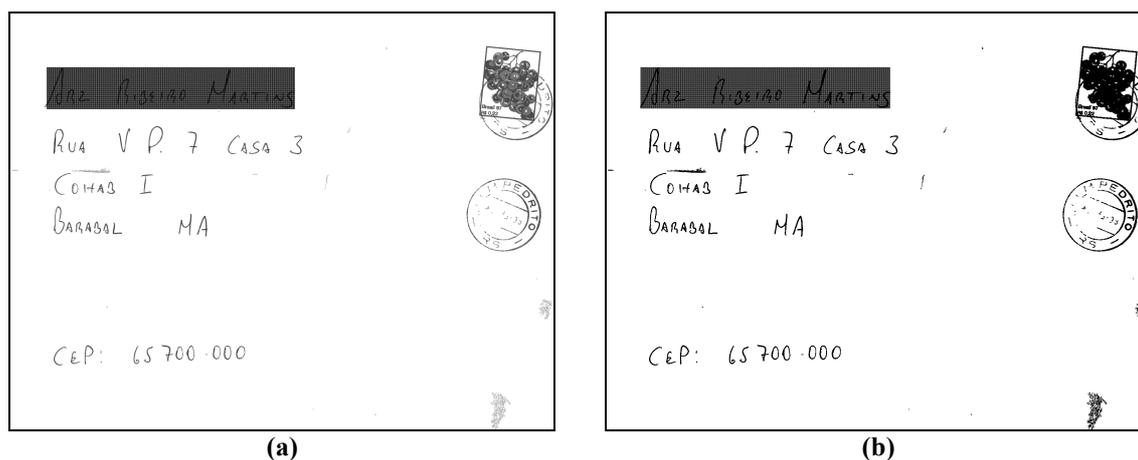


Figura 3.19: Resultado do algoritmo de segmentação proposto. (a)  $I_{CONTORNO}$  em tons de cinza; (b) Imagem  $I_{FINAL}$  binária.

### 3.6 Comentários Finais

Neste capítulo foi apresentada uma metodologia aplicada ao problema de segmentação de imagens de envelopes postais, salientando os pontos essenciais da abordagem como: a decomposição *Wavelet* utilizada; a forma como foram identificados os pontos salientes; a separação de regiões feita por um teste de hipóteses com base nas janelas de pontos salientes; a projeção reversa dos pontos salientes; e o algoritmo de perseguição de contorno. No próximo capítulo serão mostrados os experimentos realizados para validar o

método proposto, bem como os resultados obtidos e a análise dos mesmos. Estes experimentos visam validar as hipóteses levantadas aqui e justificar algumas escolhas.

## Capítulo 4

### 4. Experimentos

Neste capítulo, são apresentados a estratégia de avaliação dos resultados obtidos pela abordagem proposta e os experimentos realizados para averiguar a sua eficiência no caso de envelopes postais. A abordagem *ground truth* é apresentada na Seção 4.1. A Seção 4.2 apresenta as bases de imagens utilizadas nos experimentos. A estratégia de avaliação é descrita na Seção 4.3. Os resultados são analisados na Seção 4.4. Um estudo sobre a sensibilidade dos parâmetros do algoritmo de segmentação é apresentado na Seção 4.5. Na Seção 4.6, considerações sobre a complexidade computacional do algoritmo são feitas. E por último, comentários finais são feitos na Seção 4.7 acerca dos experimentos.

#### 4.1 Abordagem "*ground truth*"

Segundo Facon em [FACON 2001], pelo fato de não existir critérios de avaliação padrão, a avaliação de técnicas de segmentação é uma tarefa difícil. O que explica que a comunidade de processamento de imagens em geral e a área de segmentação de imagens mais especificamente carece do desenvolvimento de estratégias sólidas/indiscutíveis de avaliação de algoritmos. No caso de envelopes postais, não poderia ser diferente. De fato, existem poucos trabalhos que apresentam metodologias de avaliação da segmentação. Jain e Bhattacharjee em [JAIN, BHATTACHARJEE 1992a] avaliam os objetos de segmentação através da análise de componentes conexos e caixas retangulares (*bounding boxes*). Estas caixas retangulares são avaliadas visualmente de forma subjetiva. Já Wu *et al* em [WU *et al* 1997] submete o resultado da segmentação a um sistema de reconhecimento, que é avaliado pelo resultado do reconhecimento e não da segmentação. O que exige que se tenha disponível um reconhecedor de palavras manuscritas eficiente, o que está longe de ser trivial.

Portanto, decidiu-se empregar o conceito de abordagem "*ground truth*" em inglês que pode ser traduzida como abordagem de referência ou ainda abordagem verdadeira em português. Uma imagem "*ground truth*" é uma imagem apresentando os resultados ideais que se espera obter de qualquer algoritmo de segmentação. A Figura 4.1-a ilustra um exemplo de imagens "*ground truth*" (de referência) no caso da segmentação de um envelope postal Brasileiro. A Imagem original em tons de cinza que foi utilizada para gerar a imagem de referência da Figura 4.1-a pode ser vista na Figura 4.3-a. Visando extrair e separar o bloco endereço, o(s) selo(s), e o(s) carimbo(s), a Figura 4.1-b, a Figura 4.1-c e a Figura 4.1-d ilustram cada classe segmentada separadamente de forma ideal. Por conter grandes variações de níveis de cinza, de textura e de espessura da parte escrita, a segmentação dos selos é de fato bastante complexa. Como pode ser observado na Figura 4.1-c, decidiu-se classificar o fundo do selo como pertencente ao selo. Esse procedimento de inclusão do fundo do selo à classe selo foi realizado somente para facilitar o processo de construção das imagens de referência (*ground truth*). O fundo do selo não possui propriedades relacionadas aos objetos a serem segmentados pelo método proposto. Sabe-se, então, que vários *pixels* pertencentes ao selo, os *pixels* do fundo do selo, não serão segmentados pelo algoritmo proposto, apesar de terem sido classificados como pertencentes ao selo, na segmentação *ground truth* (ideal/desejada). Mas, isso não atrapalha a avaliação que será proposta na Seção 4.3, pois o objetivo principal da metodologia de avaliação é avaliar o resultado da segmentação do bloco endereço.

A geração das imagens segmentadas *ground truth* é de suma importância para poder avaliar com rigor (de forma quantitativa) a metodologia de segmentação proposta aqui. Estas imagens são utilizadas para realizar uma avaliação *pixel a pixel* do resultado da segmentação por classe. Vale lembrar que a segmentação manual desse tipo de imagem demanda muito tempo, além de ser uma tarefa entediante.

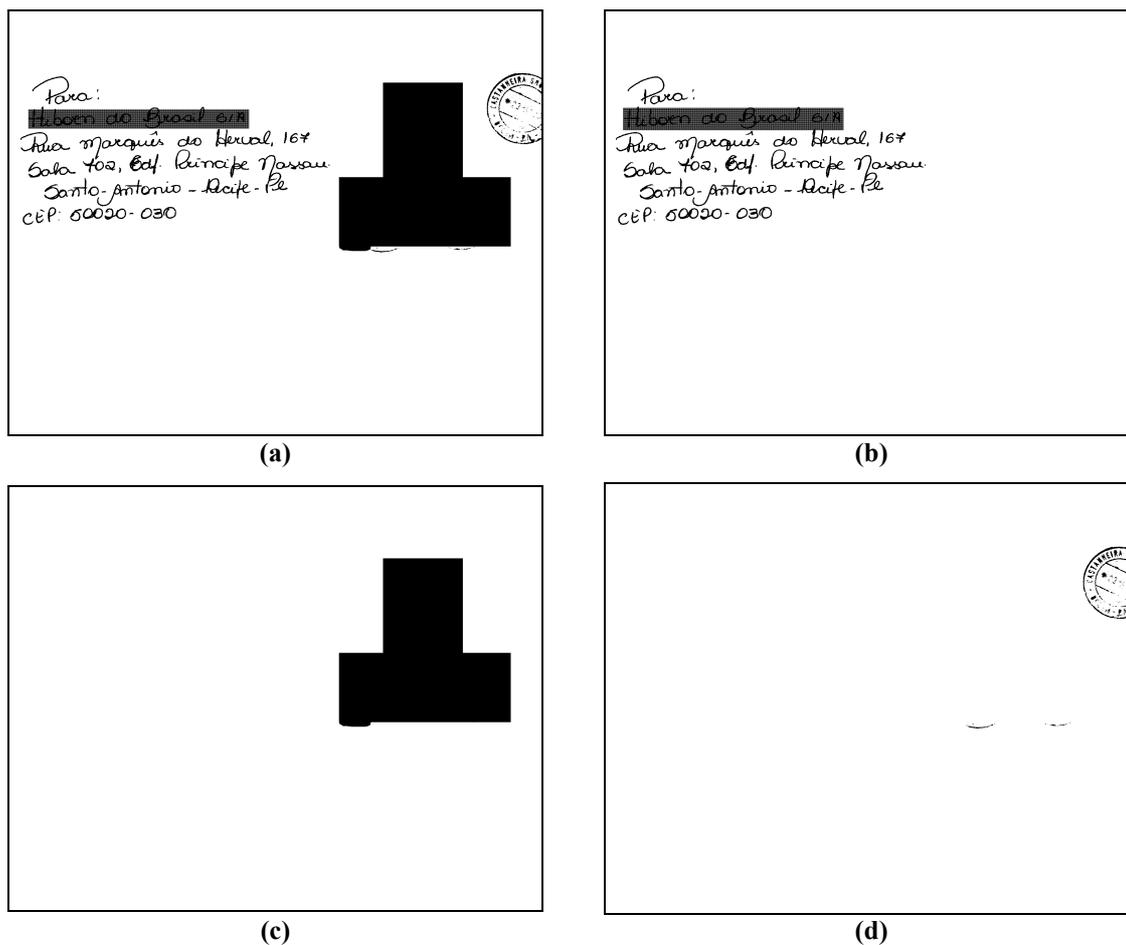


Figura 4.1: Imagens de referência (*ground truth*). (a) Resultado da segmentação desejada; Imagens (b), (c) e (d) com os *pixels* das classes bloco endereço, selo e carimbo, respectivamente.

## 4.2 As Bases de Imagens

Foram usadas duas bases de imagens de envelopes postais, uma primeira base composta de 200 imagens reais de envelopes postais (em 256 tons de cinza) retiradas de uma base de imagens de 55.000 correspondências postais obtida através de um convênio entre a PUCPR e os CB (ou ECT – Empresa Brasileira de Correios e Telégrafos), descrita em [ECT 2001]. Foram utilizadas somente 200 imagens desta base de 55.000 envelopes, pois ela contém menos de 500 envelopes postais com o bloco endereço manuscrito. O restante da base é composto de correspondências com o bloco endereço impresso em etiquetas. Uma segunda base composta de 2.000 imagens sintetizadas de envelopes postais geradas a partir do conteúdo das 200 imagens reais de envelopes postais e de alguns (10) tipos de envelopes postais com fundo complexo.

### 4.2.1 Base de imagens reais de envelopes postais brasileiros

Desta base dos CB (ECT) foram coletadas duzentas (200) imagens de envelopes postais como as apresentadas na Figura 4.2 que foram adquiridas a 200 *dpi*. Essa base é composta basicamente de dois tipos de envelopes: noventa e nove (99) imagens tem o fundo branco (como a Figura 4.2-a) e as outras cento e uma (101) imagens tem o fundo complexo (Figura 4.2-b).

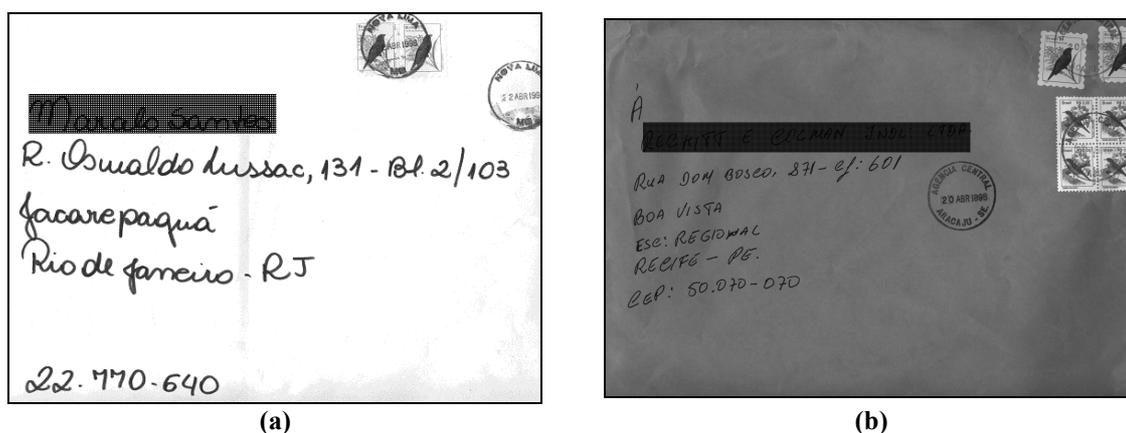


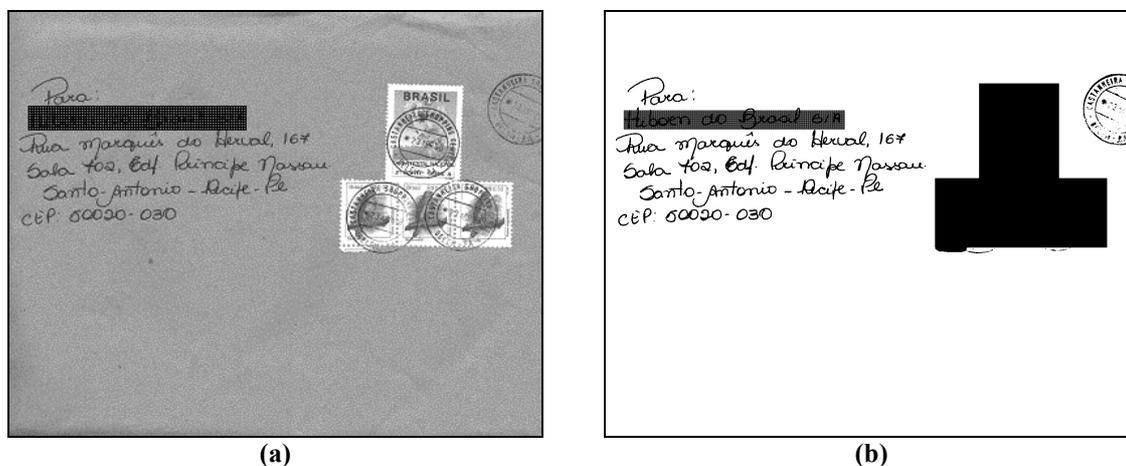
Figura 4.2: Imagens Reais. (a) Fundo branco; (b) Fundo complexo.

De cada uma dessas duzentas imagens reais foram segmentados manualmente o bloco endereço, os selos e os carimbos, gerando assim as soluções teóricas esperadas da segmentação (imagens *ground truth*). A Figura 4.1-a ilustra um exemplo de como esta segmentação em classes foi realizada. As classes estão ilustradas em imagens da Figura 4.1, onde o bloco endereço está na Figura 4.1-b, os selos estão na Figura 4.1-c e os carimbos estão na Figura 4.1-d.

### 4.2.2 Peculiaridades das imagens de envelopes postais

Propõe-se segmentar as imagens dos envelopes postais Brasileiros visando extrair e separar do fundo do envelope o bloco endereço, o(s) selo(s), e o(s) carimbo(s). A Figura 4.3-b ilustra tal propósito para uma imagem típica de um envelope postal da Figura 4.3-a. Vale a pena destacar que os componentes objetos de segmentação pelo método proposto, têm uma natureza sintática bem parecida (distribuição espacial, intensidade), porém uma natureza semântica diferente (bloco endereço: informação do destinatário; carimbo:

autenticação/informação de postagem; selo: autorização da postagem/contextualização histórica).



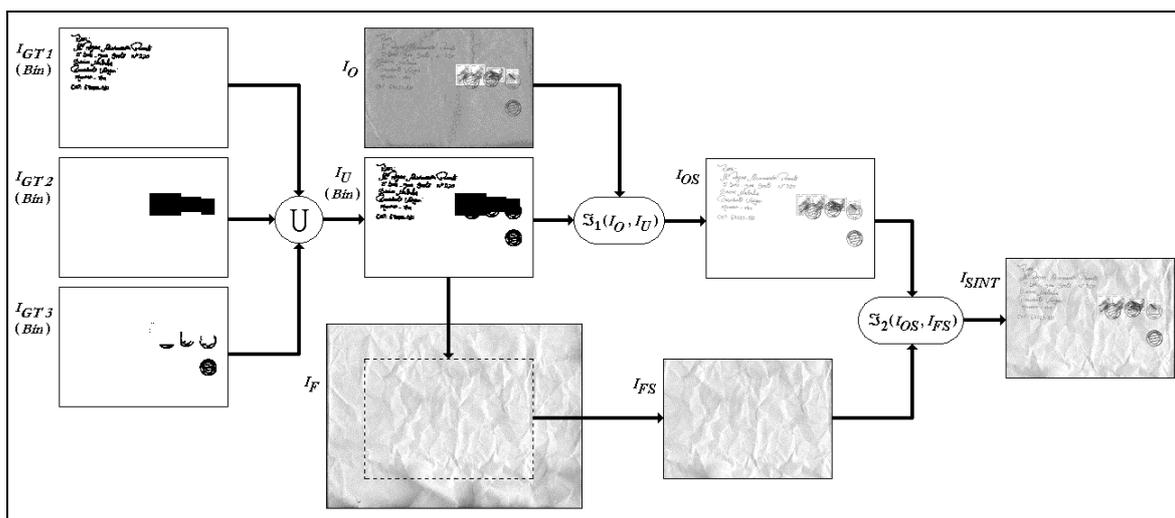
**Figura 4.3: Envelope postal. (a) Imagem típica; (b) Imagem segmentada desejada de (a).**

Vários fatores podem afetar a localização e extração dos componentes acima citados. Entre eles, podem ser citados, fundos variados de envelopes provenientes do próprio tipo de envelopes, de variações de iluminação durante a aquisição, de manipulação inadequada (gerando envelopes amassados) durante a postagem. Também o não respeito às regras de preenchimento de envelopes podem prejudicar o processo de segmentação, como por exemplo, a adição de texto manuscrito que não concerne ao endereço, o posicionamento errado. O tamanho e a espessura das letras (em geral manuscritas) pertencentes ao bloco endereço e algumas informações pré-impresas variam muito, bem como a posição destes componentes no envelope. O selo também apresenta variações de tamanho, de posição e de tipos nas imagens dos envelopes. Um envelope postal tem geralmente dois carimbos espalhados: um sobre o selo; e outro sobre o fundo do envelope. Existe uma pequena variação tanto no tamanho quanto na espessura dos carimbos presentes nos envelopes. Todas estas variações de informações nas imagens dos envelopes devem ser tratadas pelo algoritmo de segmentação.

### **4.2.3 Base de imagens sintetizadas de envelopes postais brasileiros**

Para avaliar a abordagem de segmentação proposta aqui, considerou-se o número (200) de imagens disponíveis coletadas da base dos CB (ECT) como muito pequeno. Além disso, considerou-se que essas imagens não representavam um universo suficientemente

abrangente de variações de envelopes postais. Logo, decidiu-se gerar uma base de imagens mais representativa (2000 imagens). A partir das imagens de referência (*ground truth*), segmentadas manualmente, do bloco endereço, dos selos e dos carimbos de cada uma das duzentas imagens reais e dos tipos de envelopes postais com fundos complexos, foram geradas imagens sintetizadas [FACON 2001].



**Figura 4.4:** Esquema de construção de imagens sintetizadas através de imagens reais e suas respectivas imagens de referências e tipos de fundos complexos.  $I_O$  é a imagem real original;  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$  são as imagens de referência (*ground truth*) das classes bloco endereço, selo e carimbo, respectivamente;  $I_U$  é a união entre  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$ ;  $I_{OS}$  contém todos os objetos de segmentação da imagem  $I_O$ ;  $I_F$  é o fundo complexo de envelope postal virgem (sem componentes);  $I_{FS}$  é a região selecionada para ser o fundo do envelope postal da imagem a ser sintetizada;  $I_{SINT}$  é a imagem sintetizada resultante deste processo; (*Bin*) indica que as imagens são marcadores.

A Figura 4.4 ilustra o procedimento de construção dessa base de imagens sintetizadas. Primeiramente, as imagens de referência (*ground truth*) de cada classe: bloco endereço, selo e carimbo, ( $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$  respectivamente); relacionadas à uma imagem real (da base de imagens reais), são tomadas em união para gerar a imagem  $I_U$ , que contém todos os marcadores para os objetos desejados que um método de segmentação obtenha. Os *pixels* que representam os objetos de segmentação da imagem real  $I_O$  são selecionados pela imagem de marcadores  $I_U$  através da transformação  $\mathfrak{S}_1(I_O, I_U)$ , gerando dessa forma a imagem  $I_{OS}$ . Ao mesmo tempo, uma imagem de envelope com fundo complexo virgem (sem componentes)  $I_F$  é recortada, ao centro, com as mesmas dimensões que a imagem original  $I_O$ , gerando, desta forma, a imagem  $I_{FS}$ . A partir de  $I_{OS}$  e  $I_{FS}$ , a transformação  $\mathfrak{S}_2(I_{OS}, I_{FS})$  realiza a sobreposição dos objetos de

segmentação apontados por  $I_{OS}$  sobre a imagem recortada do fundo de envelopes sem componentes  $I_{FS}$ , gerando, assim, finalmente, a imagem sintetizada  $I_{SINT}$ . A imagem sintetizada possui dimensões geométricas (altura e largura) iguais a imagem original  $I_O$  e as mesmas imagens de referência ( $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$ ), porém com um fundo diferente, relacionado a imagem  $I_{FS}$ . A Figura 4.5 mostra dois exemplos de imagens sintetizadas resultantes desse processo.

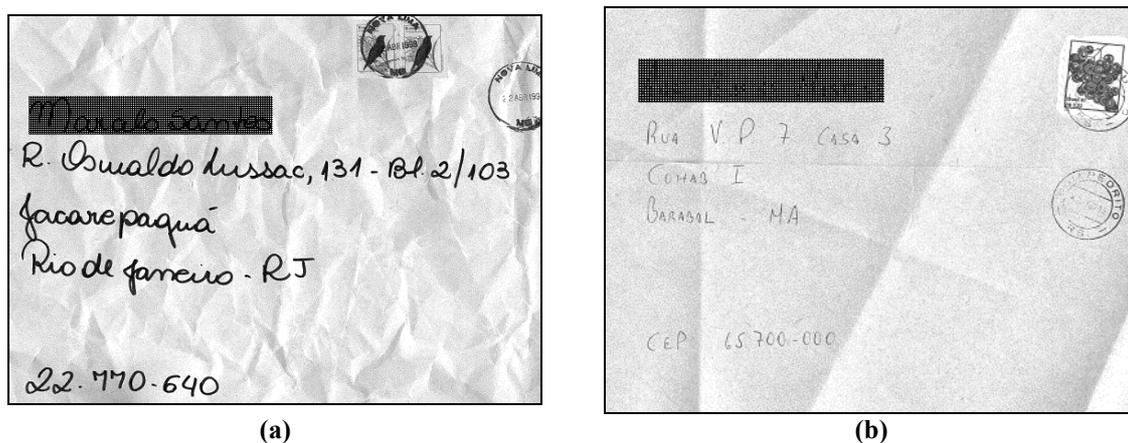


Figura 4.5: Exemplos de imagens sintetizadas.

Para cada conjunto de três imagens de referência correspondentes a uma imagem real da base de imagens reais são geradas dez imagens sintetizadas através de dez imagens de fundos complexos de envelopes diferentes. Logo, a base sintetizada é composta de 2.000 imagens (200 conjuntos de imagens de referência de cada imagem real x 10 imagens de fundos complexos de envelopes). Os tipos de fundos complexos de envelopes postais utilizados para a construção desta base estão apresentados no Apêndice A.

A respeito das 99 imagens  $I_{OS}$ , que contém os objetos de segmentação oriundos das imagens reais de fundo branco, um tratamento especial foi desenvolvido para que existisse um contraste visual entre o fundo do envelope das imagens sintetizadas e os objetos extraídos das imagens reais. Estas imagens,  $I_{OS}$ , tiveram suas intensidades corrigidas em 30 níveis de cinza, ou seja, os objetos contidos em  $I_{OS}$  tornaram-se mais escuros. Este procedimento de correção de intensidade é necessário para que se tenha um bom contraste entre o fundo do envelope e os objetos de segmentação. Esta ação é justificada pelo seguinte. Suponha que dois envelopes com fundos diferentes (um branco e outro complexo

acinzentado) foram escritos com a mesma caneta (mesma intensidade). Quando as imagens são adquiridas, os objetos de segmentação (principalmente, o manuscrito) pertencentes a elas têm uma intensidade diferente. Os objetos do envelope com fundo branco são mais claros, enquanto que os objetos do envelope com fundo complexo são mais escuros; os fundos de envelopes utilizados para gerar as imagens sintetizadas são todos complexos (acinzentados), como mostra a Figura A.1. Então, desta forma, justifica-se o uso de uma correção de intensidade nas 99 imagens  $I_{os}$  oriundas das imagens reais de fundo branco.

Nenhum levantamento estatístico, no entanto, foi realizado para tomar este número (30) como referência para a correção de intensidade. Simplesmente, observou-se visualmente que as imagens sintetizadas após esta correção tinham um contraste suficiente entre os objetos e o fundo.

#### 4.2.4 Números sobre as bases de imagens

Um levantamento realizado na base de imagens reais (composta de 200 imagens) sobre as dimensões geométricas dos envelopes postais mostrou que as imagens têm (média  $\pm$  desvio padrão): largura de  $1.857,8 \pm 20,5$  ( $\mu \pm \sigma$ ) *pixels*; altura de  $1.340,0 \pm 41,9$  ( $\mu \pm \sigma$ ) *pixels* e um número total (largura x altura) de *pixels* de  $2.549.180,2 \pm 56.822,6$  ( $\mu \pm \sigma$ ). Esse mesmo levantamento vale para a base de imagens sintetizadas (2000 imagens), pois a base de imagens sintetizadas tem proporcionalmente as mesmas dimensões geométricas (largura e altura) que aquela, porém em quantidades diferentes (10 vezes mais). Estes dados visam mostrar a dimensão do espaço de amostras (*pixels*) que cada imagem contém bem como a tarefa computacional envolvida neste processo.

**Tabela 4.1: Quantidade de *pixels* das classes nas bases de imagens de envelopes postais em relação ao número total de *pixels* da imagem.**

Classes		<i>Pixels</i> (%) ( $\mu \pm \sigma$ )
Objetos	Bloco Endereço	01,50 (01,41 $\pm$ 00,84)
	Selo	04,00 (03,97 $\pm$ 04,34)
	Carimbo	01,00 (00,93 $\pm$ 01,25)
Fundo	Fundo do Envelope	93,50 (93,75 $\pm$ 04,57)
Total		100%

Outro levantamento feito nas bases de imagens é apresentado na Tabela 4.1. Neste levantamento foram consideradas as quantidades de *pixels* que cada classe tem dentro do envelope em relação ao tamanho total da imagem. A grande maioria dos *pixels* do envelope pertencem ao fundo (aproximadamente 93,50% dos *pixels*). Os objetos perfazem um total de 6,50% dos *pixels*, sendo que: o bloco endereço fica com 1,50%; o selo com 4,00%; e o carimbo com 1,00% dos *pixels*. Estes valores são as médias de cada classe, havendo variações nestas classes que são refletidas pelo desvio padrão ( $\sigma$ ) apresentado na Tabela 4.1. Este levantamento foi feito aqui, mas é bom que se tenha em mente a proporção de *pixels* que pertencem ao fundo do envelope para ser usado na Seção 4.4, quando for feita a análise dos resultados em relação à proporção do ruído.

Depois que a abordagem *ground truth* e as bases de imagens utilizadas foram apresentadas, pode-se apresentar a metodologia quantitativa e objetiva de avaliação que é proposta aqui e foi utilizada para avaliar as imagens segmentadas dos envelopes postais obtidas neste trabalho.

### 4.3 Estratégia de avaliação da metodologia de segmentação

A grande vantagem de empregar o conceito de abordagem "*ground truth*" é que, tanto no caso das duzentas (200) imagens de envelopes postais coletadas da base dos CB (ECT) como das duas mil (2000) imagens sintetizadas, as soluções *ground truth* ideais do bloco endereço, dos selos e dos carimbos são conhecidas. Isto permite quantificar os resultados dos experimentos.

A metodologia de avaliação empregada é uma avaliação *pixel a pixel* comparando a segmentação obtida com a solução *ground truth* ideal. Conforme já visto anteriormente, para cada imagem a ser segmentada foi gerada, de forma manual, uma imagem para cada classe com os *pixels* que devem ser segmentados como objetos de segmentação (Figura 4.1-b, Figura 4.1-c e Figura 4.1-d, da seção anterior). Estas imagens foram geradas em separado por motivos de avaliação classe a classe do resultado da segmentação, entretanto o resultado final observado pelo algoritmo de segmentação e o desejado (Figura 4.1-a) apresentam somente uma única classe.

O esquema de avaliação *pixel a pixel* das imagens de envelopes postais é ilustrado pela Figura 4.6. Todas as imagens que são submetidas ao algoritmo de segmentação possuem sua respectiva imagem esperada, isto é, o resultado da segmentação desejado. Esta imagem esperada ou desejada é a imagem de referência (*ground truth*) previamente separada em três classes. A imagem resultante (imagem observada) do algoritmo de segmentação  $I_{FINAL}$  é comparada com as imagens de referência (imagens esperadas). Cada *pixel* de  $I_{FINAL}$ , que foi segmentado (*pixel* preto), é comparado com o respectivo *pixel* em  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$ . Caso esse *pixel* de  $I_{FINAL}$  seja encontrado em uma dessas três imagens (e só em uma delas), o *pixel* é classificado como sendo pertencente ao objeto de segmentação da respectiva classe. Caso contrário, o *pixel* é classificado como ruído e pertencente ao fundo do envelope. A medida que os *pixels* de  $I_{FINAL}$  são comparados com  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$  as imagens  $I_{Ob1}$ ,  $I_{Ob2}$ ,  $I_{Ob3}$  e  $I_{Ruído}$  são geradas. A razão entre o número de *pixels* existente nas imagens observadas ( $I_{Ob1}$ ,  $I_{Ob2}$  e  $I_{Ob3}$ ) e nas imagens de referência ( $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$ ) são as taxas de retorno das classes. O ruído do processo é medido como a razão entre o número de *pixels* que foram segmentados em  $I_{FINAL}$  e não pertencem a  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$ , e o número de *pixels* que pertencem ao fundo do envelope. Estas taxas de retorno e o ruído podem ser formalmente definidos como:

$$TR_{Classe\ 1} = \frac{\text{n}^\circ \text{ pixel } ( I_{FINAL} \cap I_{GT1} )}{\text{n}^\circ \text{ pixel } ( I_{GT1} )} = \frac{\text{n}^\circ \text{ pixel } ( I_{Ob1} )}{\text{n}^\circ \text{ pixel } ( I_{GT1} )} \quad (4.1)$$

$$TR_{Classe\ 2} = \frac{\text{n}^\circ \text{ pixel } ( I_{FINAL} \cap I_{GT2} )}{\text{n}^\circ \text{ pixel } ( I_{GT2} )} = \frac{\text{n}^\circ \text{ pixel } ( I_{Ob2} )}{\text{n}^\circ \text{ pixel } ( I_{GT2} )} \quad (4.2)$$

$$TRClasse\ 3 = \frac{\text{n}^\circ \text{ pixel } (I_{FINAL} \cap I_{GT3})}{\text{n}^\circ \text{ pixel } (I_{GT3})} = \frac{\text{n}^\circ \text{ pixel } (I_{Ob3})}{\text{n}^\circ \text{ pixel } (I_{GT3})} \quad (4.3)$$

$$Ruido = \frac{\text{n}^\circ \text{ pixel } (I_{FINAL} - (I_{GT1} \cup I_{GT2} \cup I_{GT3}))}{\text{n}^\circ \text{ pixel } ((I_{GT1} \cup I_{GT2} \cup I_{GT3}))} \quad (4.4)$$

onde  $TRClasse1$ ,  $TRClasse2$  e  $TRClasse3$  são as taxas de retorno das classes 1 (bloco endereço), 2 (selo) e 3 (carimbo), respectivamente;  $I_{FINAL} \cap I_{GT1}$ ,  $I_{FINAL} \cap I_{GT2}$  e  $I_{FINAL} \cap I_{GT3}$  são os conjuntos de intersecção entre a imagem segmentada  $I_{FINAL}$  e as imagens *ground truth* (de referência)  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$ , respectivamente, que representam os conjuntos  $I_{Ob1}$ ,  $I_{Ob2}$  e  $I_{Ob3}$ ;  $(I_{GT1} \cup I_{GT2} \cup I_{GT3})$  é o conjunto de *pixels* que representam todos os objetos de segmentação;  $\overline{(I_{GT1} \cup I_{GT2} \cup I_{GT3})}$  é o complemento do conjunto  $(I_{GT1} \cup I_{GT2} \cup I_{GT3})$ ; e  $\text{n}^\circ \text{ pixel } (X)$  é o número de *pixels* (pretos) existentes no conjunto  $X$ .

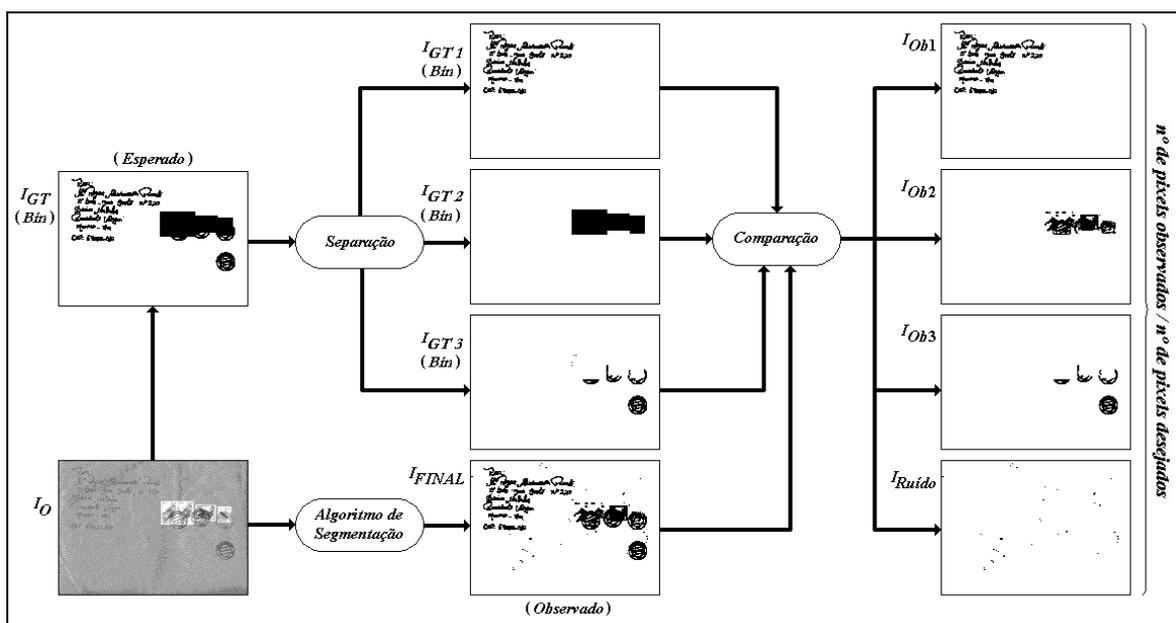


Figura 4.6: Esquema de avaliação das imagens de envelopes postais.  $I_O$  é a imagem original;  $I_{GT}$  é a segmentação esperada (ou desejada) para  $I_O$ ;  $I_{GT1}$ ,  $I_{GT2}$  e  $I_{GT3}$  são as imagens *ground-truth* (de referência) das classes bloco endereço, selo e carimbo, respectivamente, previamente separadas;  $I_{FINAL}$  é o resultado final do algoritmo de segmentação;  $I_{Ob1}$ ,  $I_{Ob2}$  e  $I_{Ob3}$  são as imagens que contém os *pixels* segmentados de cada uma das classes: bloco endereço, selo e carimbo, respectivamente;  $I_{Ruido}$  é a imagem que contém os *pixels* segmentados que não pertencem a nenhuma das três classes. (Bin) indica que as imagens são marcadores.

Para ilustrar melhor o método de avaliação, considere o seguinte exemplo hipotético: uma imagem que contém 10.000 *pixels*, distribuídos conforme mostra a Tabela 4.2. Os *pixels* presentes na imagem  $I_{FINAL}$  resultante do algoritmo de segmentação estão

distribuídos na coluna de *pixels* observados, enquanto que os *pixels* das imagens de referência estão na coluna de *pixels* desejados. A razão entre essas duas colunas (observados/desejados) representa a taxa de retorno de cada uma das classes. Os *pixels* da imagem segmentada  $I_{FINAL}$  que não pertencem a nenhuma das três classes alvo são considerados como ruídos e pertencentes ao fundo do envelope. Os resultados hipotéticos apresentados na Tabela 4.2 são proporcionais aos resultados verificados nos experimentos apresentados neste capítulo.

**Tabela 4.2: Exemplo hipotético para avaliação**

Classes	n° de <i>pixels</i> desejados	n° de <i>pixels</i> observados	n° observados / n° desejados (%)
Objetos	650	375	57,69
Bloco Endereço	150	145	96,57
Selo	400	150	37,50
Carimbo	100	80	80,00
Fundo do Envelope/ Ruído	9.300	93	01,00
Total	10.000	763	07,63

Para que os resultados dos experimentos sejam analisados segundo um critério objetivo foi necessário apresentar a metodologia de avaliação quantitativa, que por sua vez, precisou da apresentação da abordagem *ground truth* e da descrição das bases de imagens utilizadas. Então, agora os resultados dos experimentos são analisados.

#### 4.4 Análise de Resultados

Nesta subseção os resultados dos experimentos são apresentados e analisados. A metodologia de avaliação *pixel a pixel* usada para “medir” o resultado da segmentação, que foi apresentada na Seção 4.3 é utilizada. Algumas considerações sobre a metodologia de avaliação proposta são evidenciadas na Seção 4.4.1. Na Seção 4.4.2, comentários finas sobre a análise dos resultados são feitos.

Os resultados apresentados nesta seção em forma de Imagens, Tabelas e Gráficos, visam elucidar os experimentos realizados para validação do método de segmentação proposto. O algoritmo de segmentação deve levar em conta todas as possíveis variações de

informações existentes nas imagens dos envelopes. Existem variações apresentadas pelos fundos de envelope, bem como pelo tamanho e pela posição do bloco endereço. O selo também apresenta variações de posição, quantidade e distribuição, e também o carimbo apresenta variações de quantidade, de posição, e até mesmo sobreposição sobre o selo. De uma forma geral, o algoritmo deve lidar com variações de diagramação e ser independente do mesmo. Os experimentos foram realizados sobre as duas bases de imagens apresentadas na Seção 4.2. A base de imagens reais, que contém 200 imagens, e a base de imagens sintetizadas, que contém 2000 imagens, possuem muitas variações na diagramação, tipos de fundos, tamanho e tipo de áreas manuscritas e pré-impresas necessárias para validar a metodologia proposta. Todas as imagens geradas pelo método de segmentação proposto, apresentadas em figuras, para ilustrar os experimentos, tiveram os parâmetros do algoritmo de segmentação proposto configurados de acordo com a Tabela 4.3, para ambas as bases.

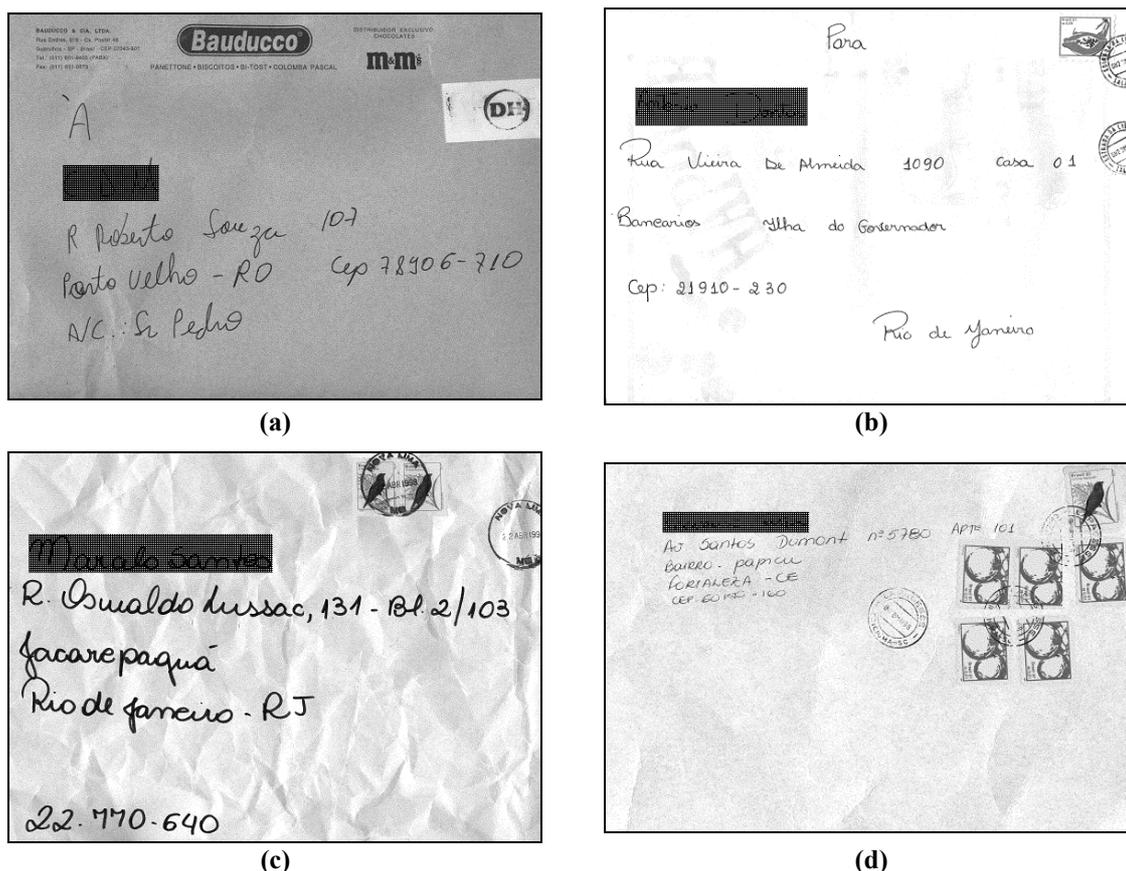


Figura 4.7: Exemplos de imagens utilizadas nos experimentos. (a) e (b) Imagens que pertencem à base de imagens reais; (c) e (d) Imagens que pertencem à base de imagens sintetizadas.

A Figura 4.7 apresenta 4 imagens das bases de imagens usadas nos experimentos. A Figura 4.8 apresenta as imagens  $I_{FINAL}$  resultantes do algoritmo de segmentação.

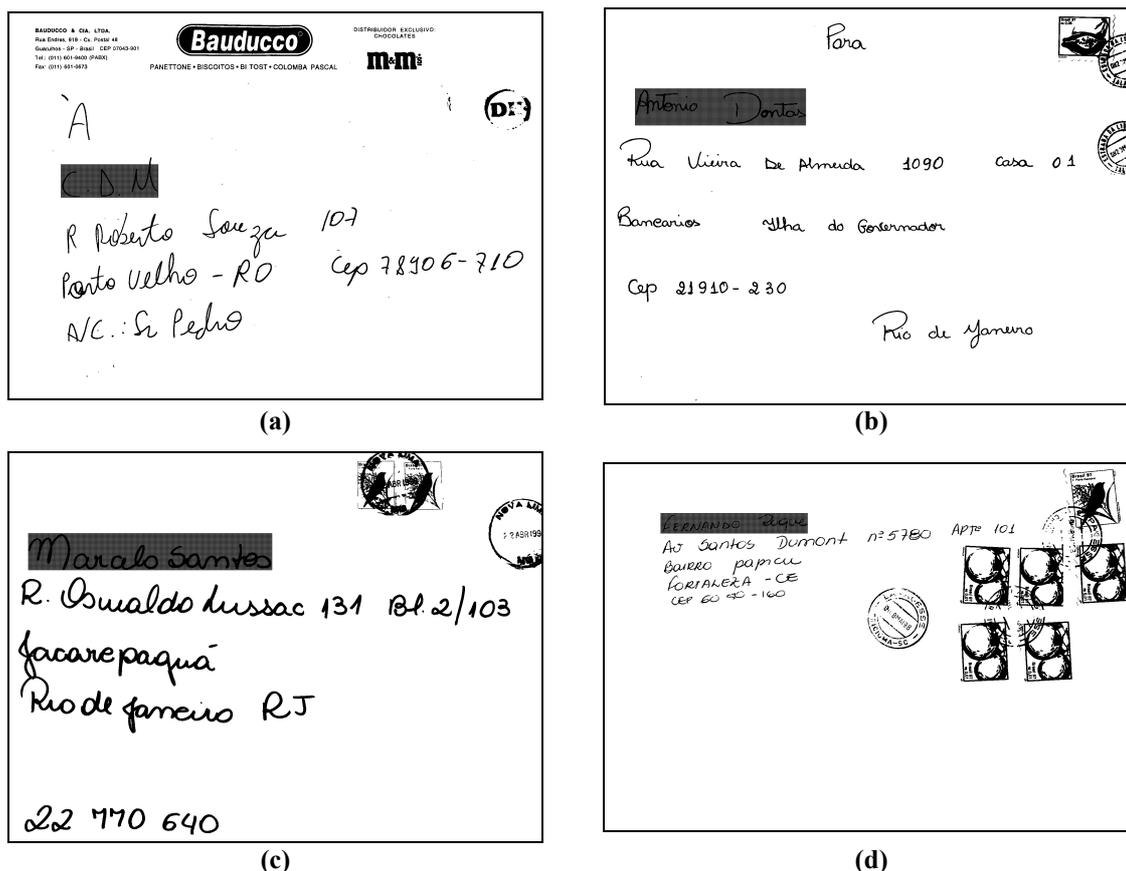
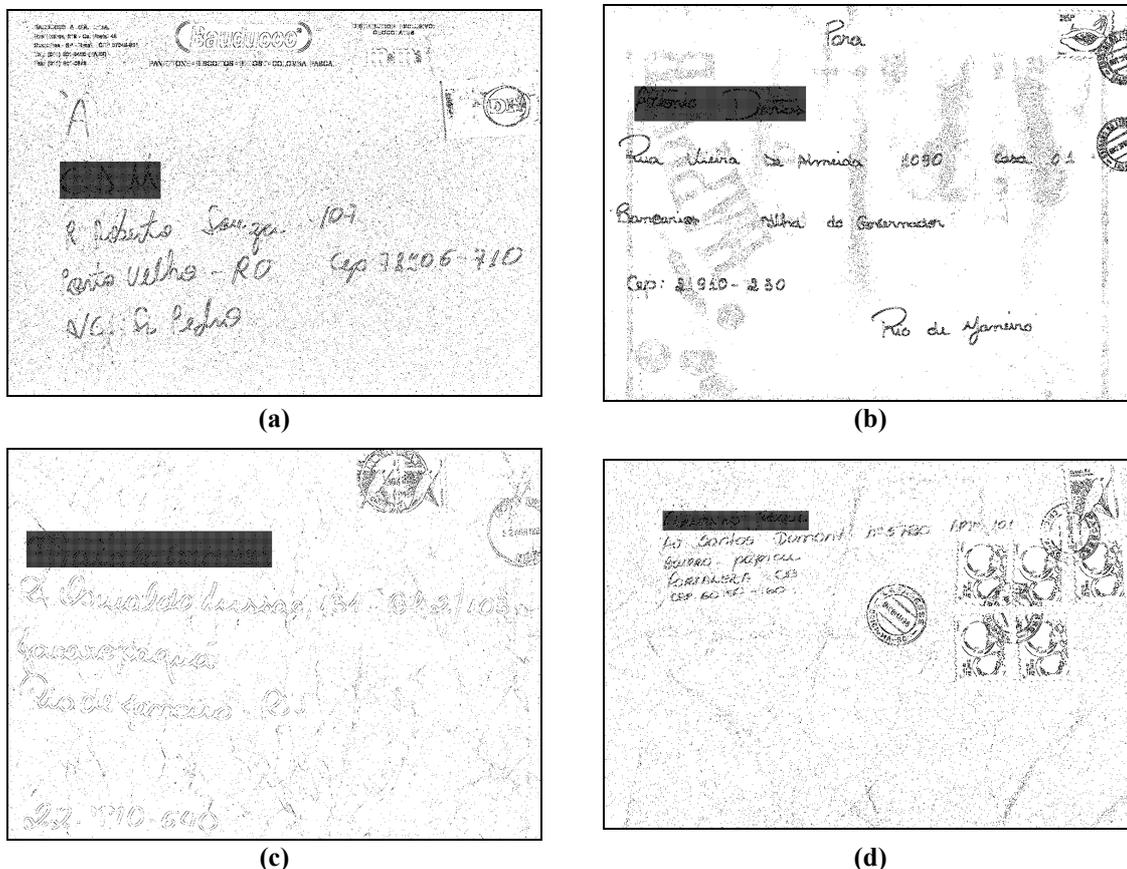


Figura 4.8: Exemplo de imagens  $I_{FINAL}$  obtidas nos experimentos pelo método proposto referente às imagens da Figura 4.7. (a) e (b) Resultados referentes a imagens que pertencem a bases de imagens reais; (c) e (d) Resultados referentes a imagens que pertencem à base de imagens sintetizadas.

Independente das diagramações, dos tipos de fundos e de todas as variações existentes nas imagens de entrada (Figura 4.7), pode-se observar através das imagens resultantes da segmentação (Figura 4.8), que todos os objetos de segmentação (blocos de endereço, selos e carimbos) foram recuperados com sucesso. As imagens apresentadas na Figura 4.7 são algumas imagens das bases de imagens reais (Figura 4.7-a e Figura 4.7-b) e sintetizadas (Figura 4.7-c e Figura 4.7-d), suas respectivas segmentações pelo método proposto são apresentadas na Figura 4.8. Vale lembrar que os parâmetros utilizados pelo algoritmo proposto são os mesmos, em ambas as bases. Isto mostra a robustez do algoritmo neste sentido.

A Figura 4.9, Figura 4.10, Figura 4.11 e a Figura 4.12 apresentam imagens resultantes dos passos intermediários do algoritmo de segmentação proposto relacionadas as imagens de entrada da Figura 4.7, isto é, as imagens correspondem respectivamente as imagens  $I_S$ ,  $I_{RC}$ ,  $I_{PRPS}$  e a imagem com os *pixels* que realmente iniciam o algoritmo de

perseguição de contorno. Estas imagens intermediárias são apresentadas para que o leitor possa melhorar o entendimento do funcionamento do algoritmo de segmentação proposto.



**Figura 4.9:** Exemplos de imagens intermediárias criadas nos experimentos. Imagens de pontos salientes  $I_s$  (dimensão  $n/2 \times m/2$ ), no espaço *Wavelet*, saída do segundo passo do algoritmo de segmentação proposto.

As imagens ( $I_s$ ) da Figura 4.9 apresentam os pontos salientes evidenciados pela transformada *Wavelet* (Seção 3.1), selecionados segundo o critério de valor absoluto dos coeficientes *Wavelet* de alta frequência apresentado na Seção 3.2. O valor do parâmetro  $\lambda_1$  para estas imagens é de 43% (ver Tabela 4.3). Pode-se notar visualmente através das imagens da Figura 4.9 que evidências para os objetos de segmentação foram extraídas, em conjunto com muitos ruídos. Entretanto, esta ainda é a imagem de saída do segundo passo e refinamentos foram/serão feitos nos passos posteriores.

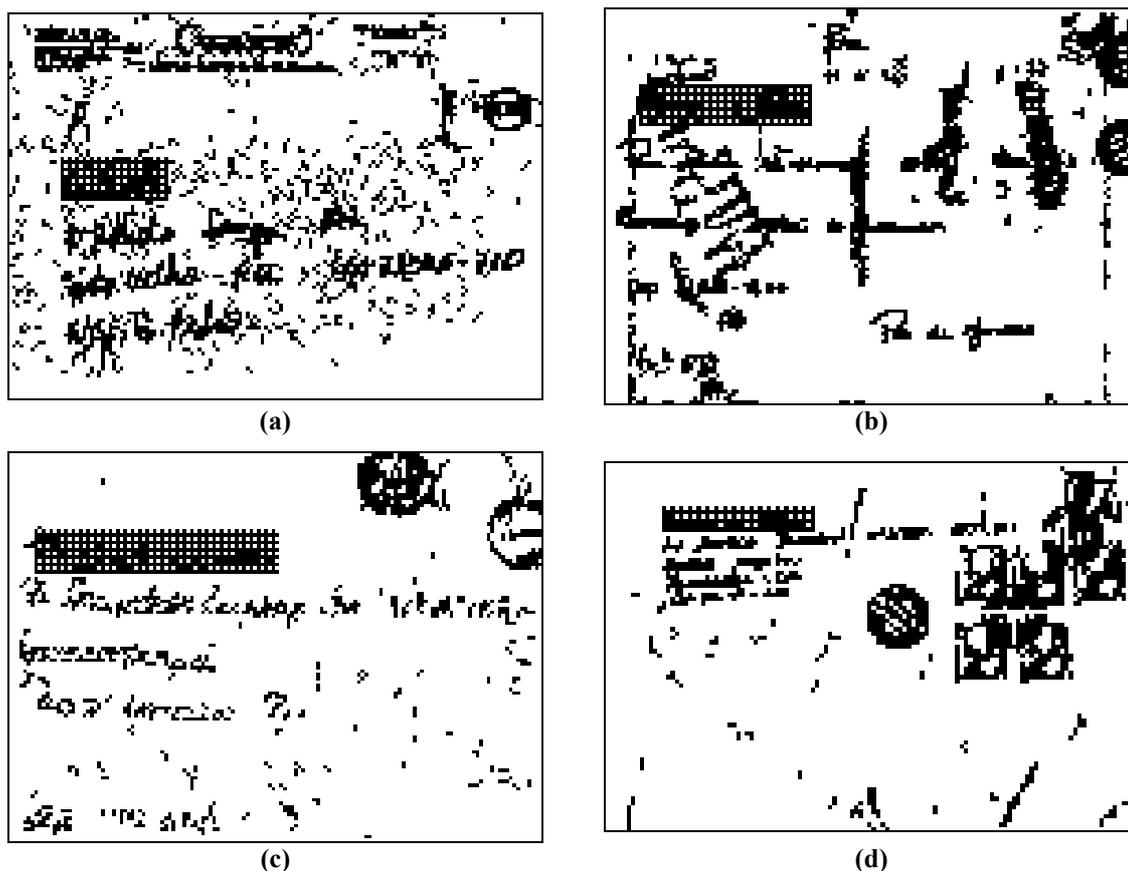
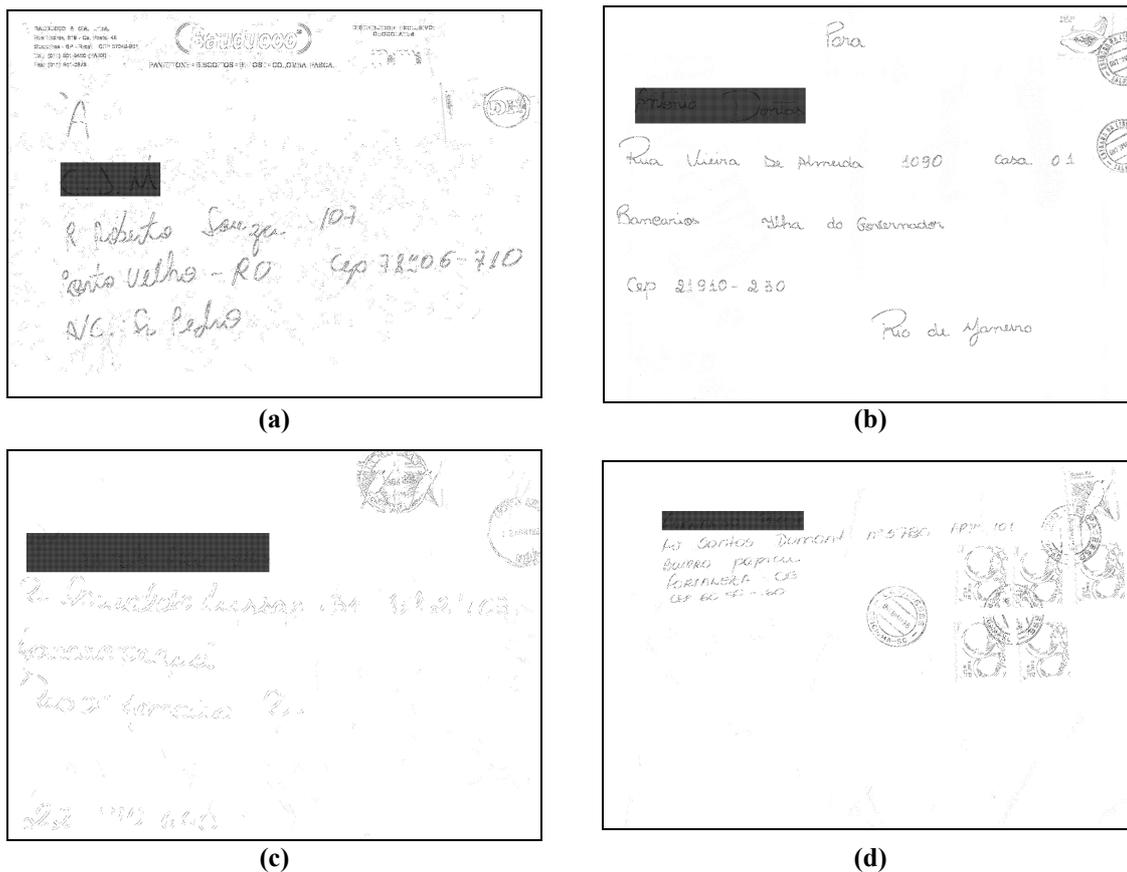


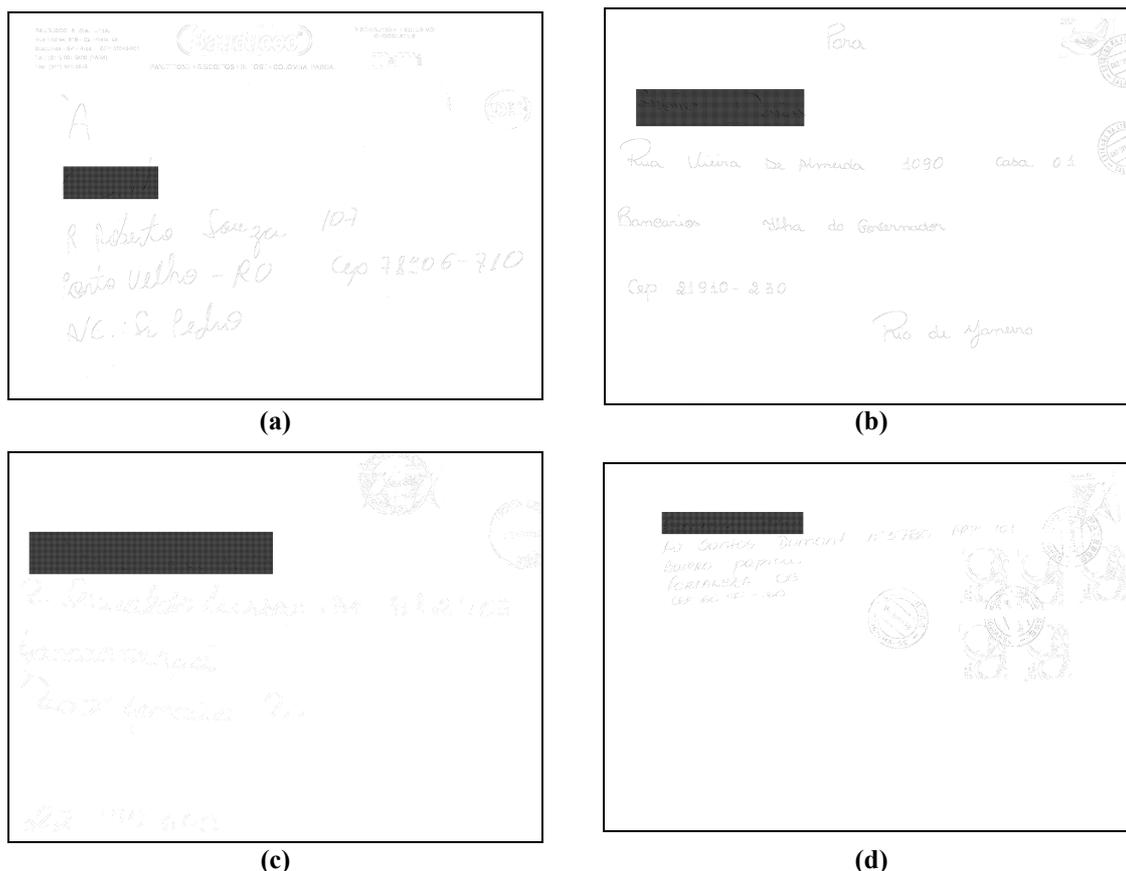
Figura 4.10: Exemplos de imagens intermediárias criadas nos experimentos. Imagens das janelas rotuladas/selecionadas  $I_{RC}$  (dimensão  $n/2k \times m/2k$ , onde  $k$  é o tamanho da janela de pontos salientes), saída do terceiro passo do algoritmo de segmentação proposto.

As imagens ( $I_{RC}$ ) da Figura 4.10 apresentam as janelas rotuladas como regiões de alta densidade de pontos salientes, conforme descrito na Seção 3.3. Neste algoritmo de rotulação de janelas de pontos salientes o tamanho  $k$  da janela de pontos salientes utilizado foi de 8, e o nível de significância  $\lambda_2$  foi de 80% de acordo com a Tabela 4.3. Observando as imagens da Figura 4.10 ainda existem muitos ruídos, isto é evidências que apontam para o fundo do envelope postal. Entretanto, estas janelas ruidosas, classificadas como de alta densidade de pontos salientes, devem possuir poucos pontos salientes em relação às janelas que realmente apontam para os objetos de segmentação, e foram/serão barradas pelas restrições globais impostas na Seção 3.5.



**Figura 4.11: Exemplos de imagens intermediárias criadas nos experimentos. Imagens da projeção reversa dos pontos salientes selecionados  $I_{PRPS}$  (dimensão  $n \times m$ ), saída do quarto passo do algoritmo de segmentação proposto.**

As imagens ( $I_{PRPS}$ ) da Figura 4.11 apresentam as projeções reversa dos pontos salientes selecionados, conforme descrito na Seção 3.4. Este passo de projeção reversa não envolve nenhum parâmetro. Esta projeção reversa simplesmente toma os pontos salientes selecionados no espaço *Wavelet* e refinados através do algoritmo de janela descrito na Seção 3.3 e reprojeta os seus quatro filhos sobre o espaço da imagem original. Nestas imagens ainda há muito ruído, apesar de não ser visível, pois estas imagens estão em tons de cinza, e os ruídos presentes geralmente são mais claros que os objetos de segmentação do bloco endereço, carimbo e selo.



**Figura 4.12:** Exemplos de imagens intermediárias criadas nos experimentos. Imagens com os *pixels* que iniciaram o contorno, tarefa inicial do quinto passo do algoritmo proposto.

As imagens da Figura 4.12 correspondem ao *pixels* que realmente iniciam o algoritmo de contorno descrito na Seção 3.5, isto é, dados os quatro *pixels* filhos referenciados pelo ponto saliente, somente um deles, será inicializado como pertencente ao contorno desde que obedecesse a Equação (3.10). Neste estágio do algoritmo, as imagens da Figura 4.12, que contém os *pixels* que inicializaram o contorno, devem estar isentas de ruídos, caso contrário esses ruídos remanescentes evoluirão podendo incorporar artefatos indesejados do fundo do envelope aos objetos de segmentação.

Dando continuidade a análise dos resultados. Em vez de avaliar a saída do algoritmo de segmentação com caixas retangulares (como em [JAIN; BHATTACHARJEE; 1992a]) ou submetendo-as a um sistema de reconhecimento (como em [WU *et al* 1997]), foi proposta na Seção 4.3 uma forma de avaliação quantitativa dos resultados da segmentação através de imagens de referência (*ground truth*) comparando *pixel a pixel* o resultado obtido com o desejado.

Embora, tenha sido desenvolvida uma metodologia para a análise de todas classes, a atenção do trabalho é voltada para a classe bloco endereço. Os componentes interessantes para a tarefa de automação postal são somente aqueles que fazem parte do bloco endereço, onde é necessário saber todas as informações do destinatário para automatizar o processo de postagem dos correios. As imagens segmentadas devem também possuir um baixo nível de ruído, para que esse ruído não atrapalhe o processo seguinte à segmentação, a identificação do bloco endereço. A tarefa da segmentação deve, então, aumentar a taxa de retorno da classe bloco endereço e ao mesmo tempo diminuir o ruído nas imagens segmentadas de envelopes postais.

**Tabela 4.3: Conjunto de parâmetros que apresentaram melhor resultado nos experimentos.**

Parâmetro	Valor
$\lambda_1$	43,00 %
$\lambda_2$	80,00 %
$\lambda_3$	10,00 %
$k$	8
$LimLocal(\alpha)$	média dos 4 <i>pixels</i> filhos

Os experimentos mostraram que o conjunto de parâmetros que apresentaram os melhores resultados é o apresentado na Tabela 4.3. O parâmetro  $\lambda_1$  determina a quantidade (%) de coeficientes *Wavelet* de alta frequência que devem ser selecionados no segundo passo do algoritmo de segmentação. O segundo parâmetro  $\lambda_2$  determina o nível de significância estatístico para o teste de hipótese realizado no terceiro passo do algoritmo. Já  $\lambda_3$  delimita globalmente a região onde os objetos de segmentação devem ocorrer. O tamanho  $k$  da janela de pontos salientes representa a granularidade que o algoritmo utiliza. O limitador local  $LimLocal(\alpha)$  limita o crescimento do contorno localmente. Estes parâmetros foram apurados empiricamente segundo as hipóteses que recaem sobre cada um deles, entretanto vale lembrar que existe uma forte correlação entre eles, devido ao encadeamento existente entre eles. Uma análise da sensibilidade destes parâmetros é apresentada na Seção 4.5.

Tabela 4.4: Resultados na base de imagens reais.

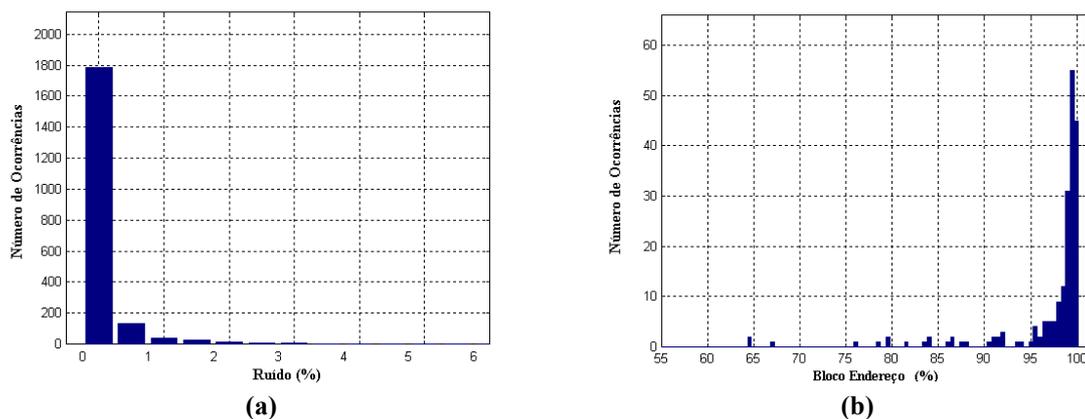
$(\mu \pm \sigma)$ (%)	Imagens Reais		
	Todos os tipos de Fundo (200 imagens)	Fundo branco (99 imagens)	Fundo complexo (101 imagens)
<b>Bloco Endereço</b>	97,13 ± 05,86	97,66 ± 05,64	96,61 ± 06,01
<b>Selo</b>	25,44 ± 19,29	31,83 ± 18,17	19,17 ± 18,26
<b>Carimbo</b>	88,12 ± 19,36	93,23 ± 14,71	83,11 ± 21,90
<b>Ruído</b>	00,58 ± 00,88	00,36 ± 00,26	00,80 ± 01,17

A Tabela 4.4 apresenta os resultados obtidos pelo método de segmentação proposto na base de imagens reais, enquanto que a Tabela 4.5 apresenta os resultados na base de imagens sintetizadas. Nos próximos parágrafos desta seção, comentários sobre os dados apresentados na Tabela 4.4 e na Tabela 4.5 serão feitos.

Tabela 4.5: Resultados na base de imagens sintetizadas.

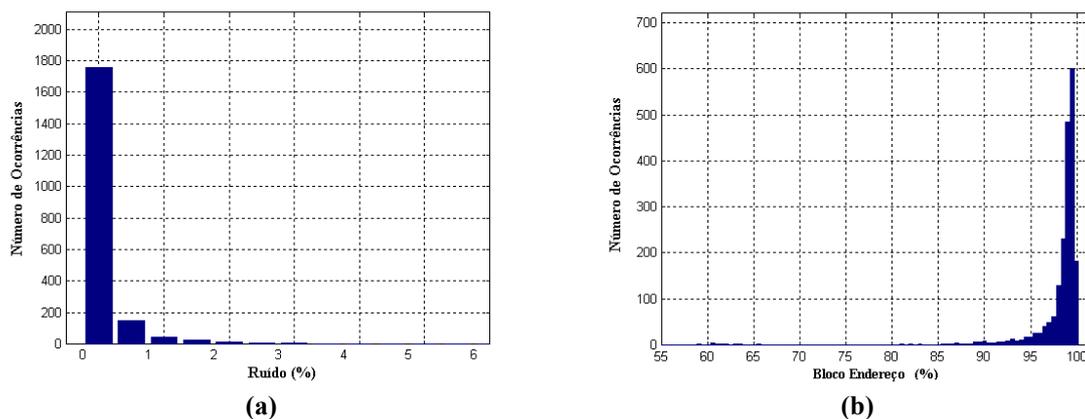
$(\mu \pm \sigma)$ (%)	Imagens Sintetizadas		
	Sintetizadas a partir de todos os tipos de fundo (2000 imagens)	Sintetizadas a partir de fundo branco (990 imagens)	Sintetizadas a partir de fundo complexo (1010 imagens)
<b>Bloco Endereço</b>	97,78 ± 04,76	96,95 ± 06,34	98,59 ± 02,02
<b>Selo</b>	33,03 ± 20,86	31,56 ± 18,06	34,47 ± 23,18
<b>Carimbo</b>	92,85 ± 15,67	91,98 ± 14,73	93,70 ± 16,49
<b>Ruído</b>	00,12 ± 00,35	00,11 ± 00,36	00,13 ± 00,35

Cada elemento numérico dessas tabelas (Tabela 4.4 e Tabela 4.5) é composto por dois valores. O primeiro representa a média e o segundo o desvio padrão das taxas de retorno das classes e do ruído. Esses dois valores permitem que uma suposição sobre a distribuição dos dados seja feita. Entretanto, os números (a média e o desvio padrão), das tabelas de resultados, sozinhos não permitem induzir de forma clara a distribuição dos resultados apresentados nos experimentos. Portanto, os resultados da Tabela 4.4 são apresentados graficamente na Figura 4.13 representando a distribuição do ruído (Figura 4.13-a) bem como a taxa de retorno da classe bloco endereço (Figura 4.13-b) ocorridos na base de imagens reais.



**Figura 4.13: Gráfico da distribuição dos resultados da base de imagens reais (200 imagens). (a) Distribuição do ruído; (b) Distribuição da taxa de retorno do bloco endereço.**

E, os resultados da base de imagens sintetizadas da Tabela 4.5 podem ser apresentados graficamente pela Figura 4.14, onde o gráfico da distribuição do ruído é apresentado na Figura 4.14-a e o da taxa de retorno do bloco endereço na Figura 4.14-b. Através dos gráficos das distribuições apresentados na Figura 4.13 e na Figura 4.14 é possível conhecer a distribuição da taxa de retorno do bloco endereço e do ruído em ambas as bases. Existe uma diferença relativamente grande entre os ruídos da base de imagens reais e da base de imagens sintetizadas. Esta diferença está nos *pixels* segmentados como objetos de segmentação nas imagens da base real pelo algoritmo de segmentação que estão ao redor dos objetos de segmentação desejados. Na base de imagens sintetizadas estes tipos de *pixels* quase não existem, pois quando as imagens sintetizadas são geradas somente os *pixels* que foram selecionados nas imagens de referências são inseridos sobre os fundos dos envelopes, e é justamente através destas imagens de referência que a segmentação é avaliada. Esta diferença é evidenciada de uma forma mais clara na Seção 4.4.1 (Figura 4.17).



**Figura 4.14: Gráfico da distribuição dos resultados da base de imagens sintetizadas (2000 imagens). (a) Distribuição do ruído; (b) Distribuição da taxa de retorno do bloco endereço.**

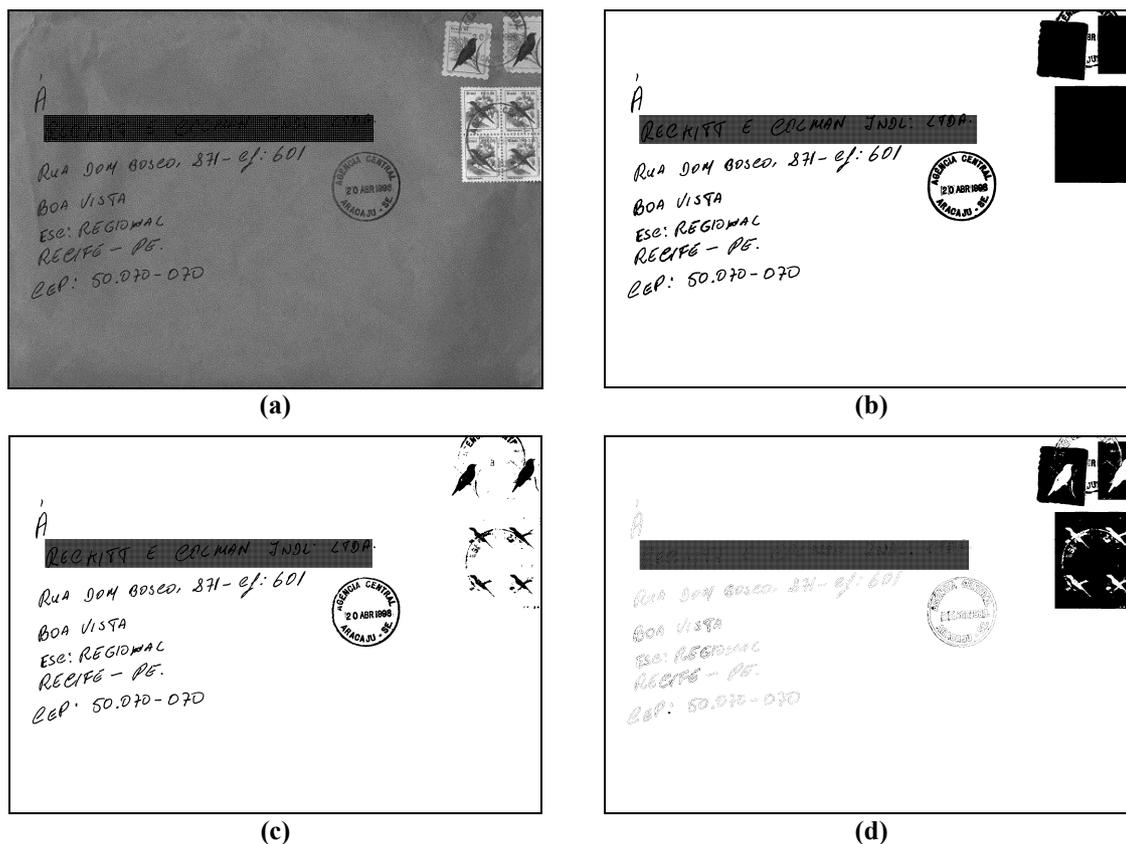
A primeira coluna numérica da Tabela 4.4 e Tabela 4.5 apresentam os resultados do método proposto em toda a base de imagens. A segunda e a terceira coluna da Tabela 4.4 apresentam os resultados do método proposto sobre partes da base de imagens reais; envelopes com fundo branco e envelopes com fundo complexo, respectivamente. Comparando-se os ruídos dessas duas colunas da Tabela 4.4 observa-se que existe mais ruído nas imagens de envelope com fundo complexo do que nas imagens de fundo branco. Isto pode ser justificado pelo fato de que o contraste existente entre os objetos de segmentação e o fundo branco é maior do que o contraste entre os objetos e o fundo complexo e, portanto, a separação dos objetos de segmentação do fundo do envelope branco é maior e mais nítida. Já os resultados da segunda e terceira coluna da Tabela 4.5 não apresentam variação significativa no ruído. Isto mostra que a correção de intensidade realizada nos objetos de segmentação das imagens de envelope de fundo branco não interferiu nos resultados.

Por conter grandes variações de níveis de cinza, de textura e de espessura da parte escrita, como já era esperada, a taxa de retorno da classe selo é baixa (menos de 40% em todos os experimentos). Esta expectativa foi confirmada pelo fato de que os *pixels* pertencentes ao fundo do selo eram classificados nas imagens de referências como pertencentes à classe selo. E, que estes *pixels* não seriam classificados como objetos de segmentação pelo método proposto, pois não possuíam características dos mesmos.

#### 4.4.1 Outras considerações a respeito da Metodologia de Avaliação Proposta

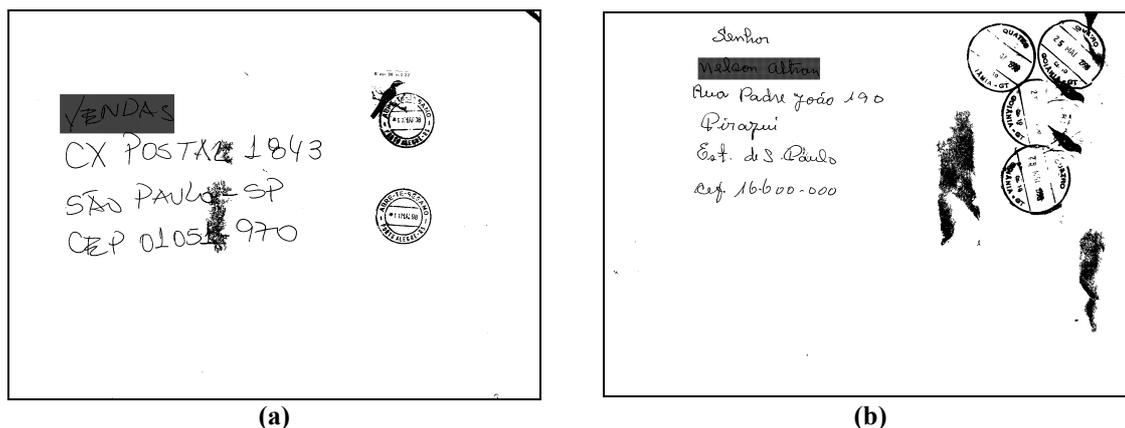
Nesta subseção são feitas algumas considerações a respeito da metodologia de avaliação quantitativa/objetiva da segmentação proposta neste trabalho. Esta metodologia de avaliação permite apresentar a quantidade de ruído presente nas imagens segmentadas, mas não a sua posição. Ela também permite apresentar a quantidade de separação ou a taxa de retorno das classes segmentadas. Mas nem sempre esse resultado quantitativo é suficiente para determinar a qualidade dos objetos segmentados. Estas dificuldades de avaliações foram observadas em poucas imagens das bases e alguns exemplos ilustrarão isso.

A Figura 4.15 ilustra um exemplo de uma imagem real segmentada pelo método proposto com taxa de retorno do bloco endereço de 79,44% e ruído de 0%. As imagens original, desejada e observada são apresentadas pela Figura 4.15-a, Figura 4.15-b e Figura 4.15-c respectivamente. Os *pixels* que deveriam ter sido segmentados como objetos de segmentação e não foram, são apresentados na Figura 4.15-d. Esses *pixels* são a parte mais externa dos objetos de segmentação. O resultado da taxa de retorno do bloco endereço da imagem da Figura 4.15-c é baixo (79,44%) quando comparado ao resultado médio da taxa de retorno do bloco endereço da base de imagens reais (97,13%). Entretanto, uma avaliação visual entre a imagem obtida Figura 4.15-c e a imagem desejada Figura 4.15-b permite afirmar que a imagem segmentada Figura 4.15-c é uma boa segmentação da imagem original da Figura 4.15-a. Esta boa segmentação deve ser entendida no sentido de que o bloco endereço segmentado (Figura 4.15-c) pode ser identificado por um sistema de reconhecimento, possibilitando, dessa forma, automatizar todo o sistema de postagem dos correios. Mostra-se, assim, uma dificuldade que a metodologia de avaliação proposta tem na identificação/separação dos objetos de segmentação.



**Figura 4.15:** Influência da taxa de retorno do bloco endereço na análise do resultado (taxa de retorno do bloco endereço de 79,44% e ruído de 0,00%). (a) Imagem original; (b) Imagem desejada; (c) Imagem segmentada (observada) pelo método proposto; (d) *Pixels* que não foram segmentados pelo algoritmo e que deveriam pertencer aos objetos de segmentação.

Através de uma inspeção visual sobre todas as imagens resultantes dos experimentos, verificou-se, de uma forma geral, que as imagens que continham ruído de 1% ou mais, tinham o bloco endereço danificado. Entretanto, através da metodologia de avaliação descrita até aqui, não é possível determinar a posição do ruído em relação aos objetos segmentados. Um exemplo desta dificuldade é apresentada na Figura 4.16. A imagem da Figura 4.16-a apresenta taxa de retorno de classificação de 98,70% e ruído de 0,79%, enquanto que a imagem da Figura 4.16-b apresenta taxa de retorno de classificação de 99,86 e ruído 2,28%. Apesar da imagem da Figura 4.16-b apresentar quase três vezes mais ruído que a imagem da Figura 4.16-a, a imagem da Figura 4.16-b tem os objetos de segmentação intactos (principalmente o bloco endereço). Assim, mostra-se outra dificuldade da metodologia de avaliação proposta em identificar a posição do ruído em relação aos objetos de segmentação.



**Figura 4.16: Localização do ruído na segmentação. (a) Bloco endereço danificado (ruído de 0,79%); (b) Imagem segmentada com alto ruído (2,28%), porém objetos de segmentação intactos.**

A Figura 4.17 ilustra um exemplo de uma imagem real segmentada pelo método proposto com taxa de retorno do bloco endereço de 99,82% e ruído de 0,45%. As imagens original, desejada e observada são apresentadas pela Figura 4.17-a, Figura 4.17-b e Figura 4.17-c respectivamente. A imagem que representa os 0,45% de *pixels* de ruído é apresentado na Figura 4.17-d. Nesta imagem, pode-se observar visualmente que estes *pixels* classificados como ruído fazem parte da transição entre os objetos de segmentação e o fundo do envelope e não prejudicam em nada o resultado da segmentação tendo em vista um posterior reconhecimento. Comparando-se visualmente as imagens desejada (Figura 4.17-b) e observada (Figura 4.17-c), conclui-se que este ruído (0,45%) não é significativo. Os *pixels* que existem a mais na Figura 4.17-c (representados na Figura 4.17-d) contribuem para uma melhor definição dos objetos do bloco endereço e não atrapalhariam um posterior reconhecimento.



**Figura 4.17:** Influência do ruído na análise do resultado (taxa de retorno do bloco endereço de 99,82% e ruído de 0,45%). (a) Imagem original; (b) Imagem desejada; (c) Imagem segmentada (observada) pelo método proposto; (d) *Pixels* que foram classificados como ruídos.

Voltando a questão, levantada na Seção 4.4, sobre a distribuição do ruído apresentado pela base de imagens reais e sintetizadas (Ver Tabela 4.4 e Tabela 4.5, e Figura 4.13 e Figura 4.14). As distribuições são diferentes; o ruído na base de imagens reais é maior que o ruído na base de imagens sintetizadas. Entretanto, não foi feito nenhum ajuste nos parâmetros para que o ruído na base de imagens fosse menor. Os parâmetros utilizados (ver Tabela 4.3) para avaliar a base de imagens sintetizadas são os mesmos que foram utilizados para avaliar a base de imagens reais. Esta diferença também pode ser justificada por ruídos provenientes de bordas de contorno como os apresentados na Figura 4.17-d que existem nos resultados da base de imagens reais e que não existem nos resultados da base de imagens sintetizadas. Pois, somente os *pixels* que foram rotulados como objetos de segmentação (imagens de referência) foram utilizados para a construção das imagens da base sintetizada.

#### 4.4.2 Comentários Finais sobre a Análise dos Resultados

Nesta subseção (Seção 4.4) os resultados do método proposto para segmentação de envelopes postais foram apresentados e analisados. Ambas tabelas dos experimentos com melhores resultados (Tabela 4.4 e Tabela 4.5) mostram taxa de retorno do bloco endereço superior a 97% e menos de 1% de ruído existente no fundo do envelope. Visto que nenhum tratamento adicional é aplicado as imagens resultantes  $I_{FINAL}$ , tais como filtragens ou operações morfológicas, estes resultados são significativos e promissores. Os resultados mostram que o método é robusto para diferentes diagramações, tipos e tamanhos de texto, e fundos de envelope.

As imagens segmentadas foram avaliadas segundo uma metodologia objetiva (*pixel a pixel*). A metodologia de avaliação permitiu avaliar a segmentação de uma forma quantitativa, sem a necessidade de avaliações visuais, que são subjetivas, apesar de terem sido apresentadas algumas dificuldades da mesma.

#### 4.5 Análise da Sensibilidade dos Parâmetros do Algoritmo de Segmentação

A análise realizada nesta seção tem como objetivo mostrar o comportamento da eficiência do algoritmo de acordo com a variação dos parâmetros, mostrando também que estes parâmetros são pontos de controle existentes no algoritmo. A análise é feita sobre o resultado da variação de um parâmetro por vez, mantendo-se todos os outros constantes. Esta pode não ser a melhor forma para estudar o comportamento da eficiência do algoritmo (utilizar um único parâmetro por vez). Mas, é, sem dúvida, uma maneira de estudar o efeito que cada um destes tem sobre o algoritmo de segmentação proposto. Na última sub-seção (Seção 4.5.6) um resumo sobre a análise realizada é apresentada .

Esta variação de parâmetros busca investigar os limites que cada parâmetro tem, além de mostrar como cada um funciona como um ponto de controle dentro do algoritmo. Esta investigação é feita através dos resultados dos experimentos nas bases de imagens

expressados por cada combinação de parâmetros. Estas limitações estão ligadas às hipóteses sustentadas por cada um dos parâmetros.

O conjunto de parâmetros com melhores resultados é o apresentado na Tabela 4.3. Em cada subseção, a avaliação do parâmetro é feita sobre cinco valores. Para gerar esses cinco valores, o valor mediano do espaço a ser gerado é o valor do referido parâmetro apresentado na Tabela 4.3. A partir desse valor são escolhidos mais quatro valores de uma forma *ad hoc*: dois valores maiores que o valor de referência e mais dois valores menores. Esta regra não é válida para  $LimLocal(X)$ , pois este já possui cinco arranjos propostos. O objetivo deste procedimento é buscar um pequeno espaço para poder avaliar o comportamento do resultado dos experimentos da segmentação do método proposto.

#### 4.5.1 Sensibilidade ao fator de seleção de coeficientes ( $\lambda_1$ )

A Tabela 4.3 mostra o conjunto de parâmetros que obtiveram os melhores resultados, e vale lembrar que o valor do  $\lambda_1$  nesta tabela é 43%. Para a análise da sensibilidade de  $\lambda_1$  foram escolhidos os valores apresentados na Tabela 4.6. Este conjunto de valores foi escolhido indiretamente através do valor de  $Z_{(100\%-\lambda_1)/2}$ . Os valores de  $Z_{(100\%-\lambda_1)/2}$  foram escolhidos de uma forma *ad hoc* segundo valores usuais da tabela de probabilidades normalizada.

**Tabela 4.6: Valores de  $\lambda_1$  usados na análise da sensibilidade.**

$\lambda_1$ (%)	$(100\% - \lambda_1) / 2$	$Z_{(100\% - \lambda_1) / 2}$
20,00 (20,06)	79,94	1,28
34,00 (33,70)	66,30	0,96
43,00 (42,38)	57,62	0,80
55,00 (54,86)	45,14	0,60
60,00 (60,30)	39,70	0,52

Os resultados dos experimentos para avaliar a sensibilidade do parâmetro  $\lambda_1$  na base de imagens reais é apresentado na Tabela 4.7. A Tabela 4.8 apresenta os resultados dos experimentos desta subseção na base de imagens sintetizadas.

**Tabela 4.7: Resultados obtidos com a variação de  $\lambda_1$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.**

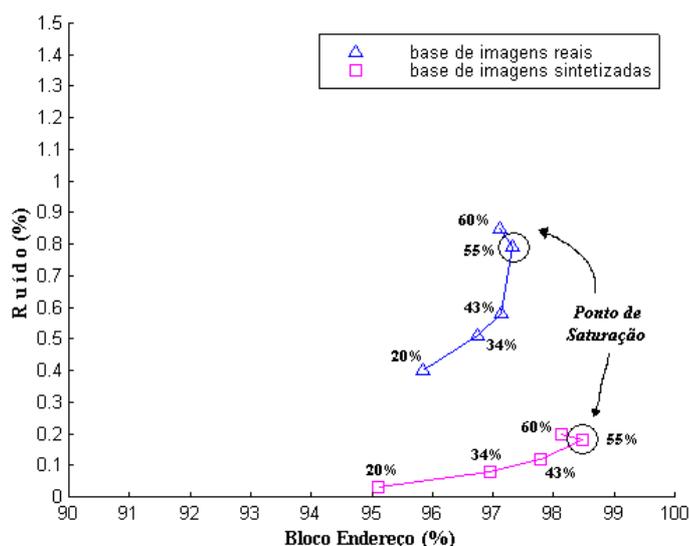
$(\mu \pm \sigma)$ (%)	$\lambda_1 / Z_{(100\%-\lambda_1)/2}$				
	60% / 0,52	55% / 0,60	43% / 0,80	34% / 0,96	20% / 1,28
<b>Bloco Endereço</b>	97,12 ± 06,25	97,33 ± 05,70	97,13 ± 05,86	96,75 ± 06,11	95,85 ± 06,71
<b>Selo</b>	25,31 ± 18,70	25,59 ± 19,38	25,44 ± 19,29	25,27 ± 19,16	24,44 ± 18,32
<b>Carimbo</b>	88,63 ± 18,81	88,64 ± 18,93	88,12 ± 19,36	87,51 ± 19,84	85,94 ± 20,90
<b>Ruído</b>	00,85 ± 01,16	00,79 ± 01,12	00,58 ± 00,88	00,51 ± 00,75	00,40 ± 00,47

**Tabela 4.8: Resultados obtidos com a variação de  $\lambda_1$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$\lambda_1 / Z_{(100\%-\lambda_1)/2}$				
	60% / 0,52	55% / 0,60	43% / 0,80	34% / 0,96	20% / 1,28
<b>Bloco Endereço</b>	98,12 ± 07,57	98,48 ± 04,43	97,78 ± 04,76	96,69 ± 05,22	95,10 ± 06,93
<b>Selo</b>	33,77 ± 21,79	33,61 ± 21,12	33,03 ± 20,86	32,40 ± 20,55	31,12 ± 20,07
<b>Carimbo</b>	93,15 ± 16,31	93,44 ± 15,41	92,85 ± 15,67	92,09 ± 16,16	90,29 ± 17,18
<b>Ruído</b>	00,20 ± 00,47	00,18 ± 00,44	00,12 ± 00,35	00,08 ± 00,26	00,03 ± 00,12

Estas duas tabelas são apresentadas da mesma forma que os dados da Tabela 4.4 e da Tabela 4.5. Os elementos numéricos representam a média e o desvio padrão das taxas de retorno e do ruído que foram obtidos nos experimentos.

Para facilitar a compreensão dos dados apresentados na Tabela 4.7 e da Tabela 4.8, optou-se por transformar os dados expostos nas tabelas em gráficos. Através dos resultados da Tabela 4.7 e da Tabela 4.8 é construído o gráfico da Figura 4.18, onde o eixo das abscissas é representado pela taxa de retorno do bloco endereço e o eixo das ordenadas é representado pelo ruído. Cada experimento é expresso como um ponto no gráfico da Figura 4.18 de acordo com o ruído e a taxa de retorno do bloco endereço obtido. Os experimentos realizados sobre a base de imagens reais são representados por triângulos, enquanto que os experimentos realizados sobre a base de imagens sintetizadas são representados por quadrados. Os experimentos realizados em cada uma das bases são conectados por retas. Estas conexões são realizadas entre os experimentos com parâmetros vizinhos, isto é, aqueles que tem o valor do parâmetro variado mais próximo. Por exemplo, o ponto do experimento com  $\lambda_1 = 43\%$  é conectado ao ponto do experimento com  $\lambda_1 = 34\%$  e ao ponto do experimento com  $\lambda_1 = 55\%$ . Desta forma é criada uma curva que descreve o comportamento do resultado do algoritmo em função da variação deste parâmetro.



**Figura 4.18:** Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do parâmetro  $\lambda_1$ .

As duas curvas (da base de imagens reais e sintetizadas) apresentam o mesmo comportamento. Um ponto de saturação é visto em  $\lambda_1 = 55\%$ . Esse é o ponto de pico do espaço de resultados (bloco endereço x ruído) em relação ao parâmetro  $\lambda_1$ . Entretanto, este ponto é atingido com um acréscimo significativo de ruído na base de imagens reais. Porém, é possível observar que a medida que a quantidade de coeficientes selecionados diminui (o valor de  $\lambda_1$  diminui) a taxa de retorno do bloco endereço e o ruído também diminuem. Por outro lado, a medida que a quantidade de coeficientes selecionados aumenta (após o ponto de saturação  $\lambda_1 = 55\%$ ) o ruído diminui junto com a taxa de retorno do bloco endereço. A análise do parâmetro  $\lambda_1$  acima do valor de saturação ( $\lambda_1 = 55\%$ ) perde sentido, pois foge das hipóteses levantadas pelo método proposto. Lembrando que  $\lambda_1$  determina a quantidade de coeficientes selecionados, e o passo que envolve o parâmetro  $\lambda_1$  tem como objetivo reduzir a quantidade de coeficientes *Wavelet*, bem como selecionar somente os mais significativos.

Isto, sem levar em conta a correlação existente entre esse parâmetro ( $\lambda_1$ ) e os outros parâmetros existentes no algoritmo. O parâmetro  $\lambda_1$  é o primeiro a ser utilizado na ordem de execução do algoritmo, então a dependência do resultado gerado por ele é alta.

#### 4.5.2 Sensibilidade ao nível de significância ( $\lambda_2$ ) entre distribuições no teste de hipótese

De acordo com a Tabela 4.3, o método proposto alcançou o melhor resultado com  $\lambda_2 = 80\%$ , isto é, o melhor resultado dos experimentos foi obtido para um nível de significância de  $\lambda_2 = 80\%$  utilizado no teste de hipótese da rotulação controlada de janelas de pontos salientes. O conjunto de valores escolhidos para a análise de sensibilidade do parâmetro  $\lambda_2$  é apresentado na Tabela 4.9. Este conjunto de valores foi escolhido indiretamente através do valor de  $Z_{\lambda_2/2}$ . Como na Seção 4.5.1, a escolha dos valores de  $Z_{\lambda_2/2}$  foi feita de forma *ad doc*.

**Tabela 4.9: Valores de  $\lambda_2$  usados na análise da sensibilidade.**

$\lambda_2$ (%)	$\lambda_2 / 2$ (%)	$Z_{\lambda_2/2}$
66,00 (66,30)	33,15	0,96
70,00 (70,16)	35,08	1,04
80,00 (79,94)	39,97	1,28
90,00 (90,00)	44,95	1,64
95,00	47,50	1,96

A Tabela 4.10 e a Tabela 4.11 apresentam os resultados dos experimentos utilizados para avaliar a sensibilidade do parâmetro  $\lambda_2$  na base de imagens reais e sintetizadas, respectivamente.

**Tabela 4.10: Resultados obtidos com a variação de  $\lambda_2$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$\lambda_2 / Z_{\lambda_2/2}$				
	66% / 0,96	70% / 1,04	80% / 1,28	90% / 1,64	95% / 1,96
<b>Bloco Endereço</b>	97,32 ± 05,78	97,23 ± 05,80	97,13 ± 05,86	96,68 ± 05,96	96,17 ± 06,15
<b>Selo</b>	25,46 ± 19,30	25,45 ± 19,30	25,44 ± 19,29	25,41 ± 19,27	25,37 ± 19,24
<b>Carimbo</b>	88,25 ± 19,27	88,20 ± 19,29	88,12 ± 19,36	87,77 ± 19,65	87,30 ± 19,89
<b>Ruído</b>	00,63 ± 00,94	00,61 ± 00,93	00,58 ± 00,88	00,55 ± 00,84	00,53 ± 00,81

**Tabela 4.11: Resultados obtidos com a variação de  $\lambda_2$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$\lambda_2 / Z_{\lambda_2/2}$				
	66% / 0,96	70% / 1,04	80% / 1,28	90% / 1,64	95% / 1,96
<b>Bloco Endereço</b>	98,36 ± 04,58	98,12 ± 04,63	97,78 ± 04,76	96,43 ± 05,61	94,02 ± 08,53
<b>Selo</b>	33,28 ± 20,93	33,16 ± 20,89	33,03 ± 20,86	32,54 ± 20,64	32,01 ± 20,36
<b>Carimbo</b>	93,22 ± 15,62	93,10 ± 15,62	92,85 ± 15,67	91,78 ± 15,96	89,88 ± 16,49
<b>Ruído</b>	00,14 ± 00,37	00,13 ± 00,36	00,12 ± 00,35	00,09 ± 00,30	00,07 ± 00,24

Da mesma forma como foi feito na subseção anterior (Seção 4.5.1), os dados da Tabela 4.10 e da Tabela 4.11 são condensados em um gráfico que é apresentado na Figura 4.19. Observando-se as duas curvas de resultados dos experimentos nas bases de imagens reais e sintetizadas apresentadas no gráfico da Figura 4.19 nota-se que o ruído e a taxa de retorno do bloco endereço variam no mesmo sentido, isto é, quando a taxa de retorno do bloco endereço aumenta (diminui), o ruído também aumenta (diminui). A taxa de retorno do bloco endereço apresenta uma maior variação na base de imagens sintetizadas, ou seja, essa variação depende dos tipos de fundos envelopes, e na base de imagens sintetizadas tem-se menos tipos de fundos. Então, a variação do parâmetro  $\lambda_2$  foi mais sensível nessa base de imagens.

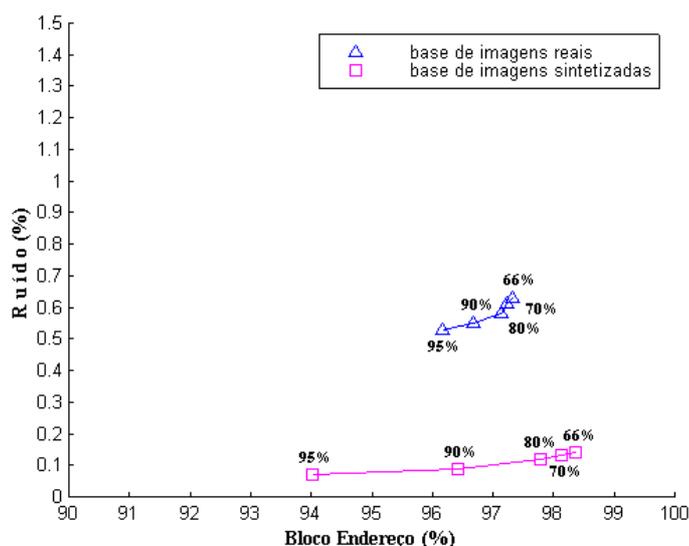


Figura 4.19: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do parâmetro  $\lambda_2$ .

### 4.5.3 Sensibilidade ao fator que delimita a região de existência de objetos de segmentação ( $\lambda_3$ )

O valor de  $\lambda_3$  é 10% na Tabela 4.3, que apresenta o melhor conjunto de parâmetros para os experimentos realizados. O conjunto de valores escolhidos para  $\lambda_3$  é apresentado na Tabela 4.12. Este conjunto de valores foi escolhido indiretamente através do valor de  $Z_{(50\%-\lambda_3)}$ . Como na Seção 4.5.1 e na Seção 4.5.2, a escolha dos valores de  $Z_{(50\%-\lambda_3)}$  foi feita de uma forma *ad hoc*, e os valores deste parâmetro foram os mesmos utilizados na Seção 4.5.2 (os valores de  $Z_{\lambda_2/2}$ ).

Tabela 4.12: Valores de  $\lambda_3$  usados na análise da sensibilidade

$\lambda_3$ (%)	$(50\% - \lambda_3)$ (%)	$Z_{(50\%-\lambda_3)}$
17,00 (16,85)	33,15	0,96
15,00 (14,92)	35,08	1,04
10,00 (10,03)	39,97	1,28
5,00 (5,05)	44,95	1,64
2,50	47,50	1,96

Os resultados dos experimentos realizados nas bases de imagens reais e sintetizadas para avaliar a sensibilidade do parâmetro  $\lambda_3$  são apresentados na Tabela 4.13 e na Tabela 4.14, respectivamente.

**Tabela 4.13: Resultados obtidos com a variação de  $\lambda_3$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$\lambda_3 / Z_{(50\% - \lambda_3)}$				
	17% / 0,96	15% / 1,04	10% / 1,28	5% / 1,64	2,50% / 1,96
<b>Bloco Endereço</b>	98,45 ± 03,47	98,21 ± 03,93	97,13 ± 05,86	94,37 ± 09,75	90,82 ± 13,89
<b>Selo</b>	27,16 ± 20,19	26,73 ± 20,00	25,44 ± 19,29	23,57 ± 18,50	22,04 ± 17,89
<b>Carimbo</b>	91,65 ± 16,67	90,87 ± 17,24	88,12 ± 19,36	83,02 ± 22,41	78,19 ± 24,54
<b>Ruído</b>	01,11 ± 01,55	00,96 ± 01,38	00,58 ± 00,88	00,31 ± 00,50	00,17 ± 00,29

**Tabela 4.14: Resultados obtidos com a variação de  $\lambda_3$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$\lambda_3 / Z_{(50\% - \lambda_3)}$				
	17% / 0,96	15% / 1,04	10% / 1,28	5% / 1,64	2,50% / 1,96
<b>Bloco Endereço</b>	98,46 ± 02,95	98,33 ± 03,33	97,78 ± 04,76	96,23 ± 07,78	94,02 ± 11,27
<b>Selo</b>	35,54 ± 22,26	34,88 ± 21,87	33,03 ± 20,86	30,23 ± 19,48	27,54 ± 18,22
<b>Carimbo</b>	94,93 ± 14,18	94,24 ± 14,50	92,85 ± 15,67	90,26 ± 17,64	87,24 ± 19,75
<b>Ruído</b>	00,35 ± 00,80	00,27 ± 00,64	00,12 ± 00,35	00,04 ± 00,14	00,02 ± 00,07

Novamente, os dados apresentados na Tabela 4.13 e na Tabela 4.14 são plotados no gráfico da Figura 4.20, gerando duas curvas que representam os resultados dos experimentos realizados sobre as bases de imagens.

Essas duas curvas da Figura 4.20 apresentam comportamento semelhante no que diz respeito a forma. A taxa de retorno do bloco endereço e o ruído variam proporcionalmente de maneira direta. Esse parâmetro  $\lambda_3$  determina a localização dos objetos de segmentação no histograma da imagem de acordo com a Figura 3.14 da Seção 3.5. O teste de hipótese que leva em conta  $\lambda_3$  é baseado na hipótese de que os objetos de segmentação são mais escuros que o fundo do envelope.

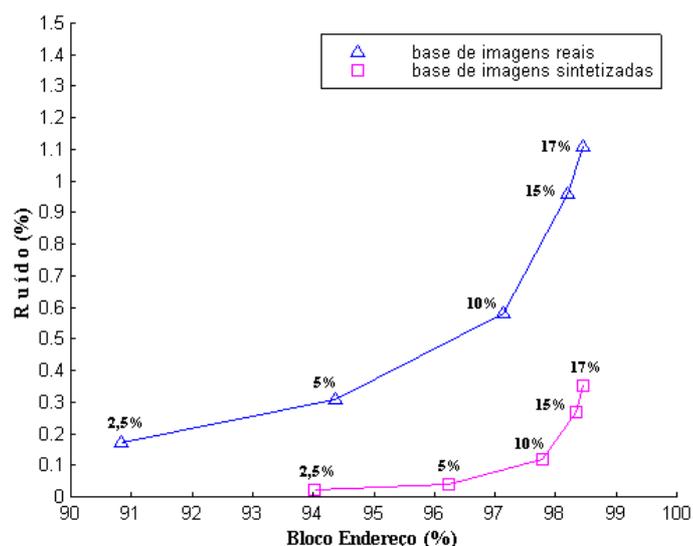


Figura 4.20: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do parâmetro  $\lambda_3$ .

#### 4.5.4 Sensibilidade ao tamanho $k$ da janela de Pontos Salientes

Na Tabela 4.3, que apresenta o conjunto de parâmetros que obteve o melhor resultado nas bases de imagens, o valor do parâmetro  $k$  é 8. O conjunto de valores escolhidos para a análise da sensibilidade do parâmetro  $k$  é mostrado na Tabela 4.15. Esses valores utilizados são múltiplos de 2. Esta escolha foi feita com base na intuição e também pela facilidade de implementação destes valores. A partir do uso destes valores múltiplos de 2, surge a idéia de uma metodologia de segmentação que utilize as propriedades *Wavelet* de multi-escala conjuntamente com tamanhos variados de janelas de pontos salientes – um trabalho futuro.

Tabela 4.15: Tamanho  $k$  de janelas de pontos salientes usados na análise da sensibilidade.

$k$ (tamanho)	$k^2$ (elementos)
2	4
4	16
8	64
16	256
32	1024

A Tabela 4.16 e a Tabela 4.17 mostram os resultados obtidos na análise da sensibilidade do tamanho  $k$  das janelas de pontos salientes nas bases de imagens reais e sintetizadas, respectivamente.

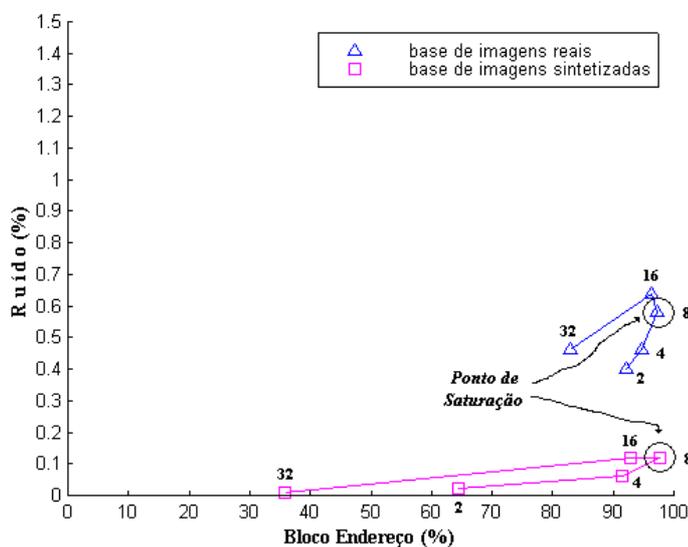
**Tabela 4.16: Resultados obtidos com a variação do tamanho  $k$  da janela de pontos salientes, na base de imagens reais, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$k$				
	2	4	8	16	32
<b>Bloco Endereço</b>	92,14 ± 09,02	94,65 ± 06,90	97,13 ± 05,86	96,42 ± 10,51	82,87 ± 32,91
<b>Selo</b>	24,06 ± 17,95	25,09 ± 19,06	25,44 ± 19,29	25,06 ± 18,72	24,95 ± 19,23
<b>Carimbo</b>	78,04 ± 22,65	84,82 ± 20,24	88,12 ± 19,36	87,90 ± 19,73	78,78 ± 31,94
<b>Ruído</b>	00,40 ± 00,63	00,46 ± 00,76	00,58 ± 00,88	00,64 ± 00,88	00,46 ± 00,78

**Tabela 4.17: Resultados obtidos com a variação do tamanho  $k$  da janela de pontos salientes, na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$k$				
	2	4	8	16	32
<b>Bloco Endereço</b>	64,34 ± 25,16	91,51 ± 10,14	97,78 ± 04,76	92,91 ± 20,99	35,77 ± 45,19
<b>Selo</b>	27,92 ± 18,43	31,30 ± 19,98	33,03 ± 20,86	32,95 ± 21,03	17,02 ± 22,06
<b>Carimbo</b>	56,35 ± 24,57	84,67 ± 17,68	92,85 ± 15,67	89,19 ± 22,87	41,21 ± 45,35
<b>Ruído</b>	00,02 ± 00,08	00,06 ± 00,25	00,12 ± 00,35	00,12 ± 00,37	00,01 ± 00,04

Os dados dos resultados, referentes à taxa de retorno do bloco endereço e ao ruído, realizados para a análise da sensibilidade do parâmetro  $k$  apresentados na Tabela 4.16 e Tabela 4.17 são colocados em um gráfico que está na Figura 4.21.



**Figura 4.21: Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação à variação do tamanho  $k$  da janela de pontos salientes.**

Observando-se as duas curvas da Figura 4.21, nota-se que existe um ponto de saturação nelas. Este ponto ocorre nas duas curvas para o valor do parâmetro  $k = 8$ . A taxa de retorno do bloco endereço variou bruscamente na base de imagens sintetizadas em relação à variação ocorrida na base de imagens reais. Esta variação é dependente dos tipos

de fundos de envelopes existentes bem como das zonas de transição existentes entre os objetos de segmentação e o fundo dos envelopes. A sensibilidade deste parâmetro está correlacionada ao parâmetro  $\lambda_2$  (Teste de hipótese), que determina a separação das regiões de alta e baixa densidade de janelas de pontos salientes; cada janela tem tamanho  $k$  e  $k^2$  elementos.

#### 4.5.5 Sensibilidade ao limitador local ( $LimLocal(\alpha)$ ) gerado pelo arranjo entre os 4 *pixels* filhos referenciados pelo ponto saliente

Os pontos salientes representam na imagem original regiões (de  $2 \times 2$  *pixels*) de alta transição. Essas regiões de transição contêm informações que podem ser usadas como limites ao crescimento do contorno. Os valores (as posições) dos limitadores locais –  $LimLocal(\alpha)$  – definidos aqui são determinados através dos quatro *pixels* filhos que são referenciados pelos pontos salientes. Uma ilustração do posicionamento do limitador local  $LimLocal(\alpha)$  sobre o relevo da imagem do envelope postal pode ser visto na Figura 4.22.

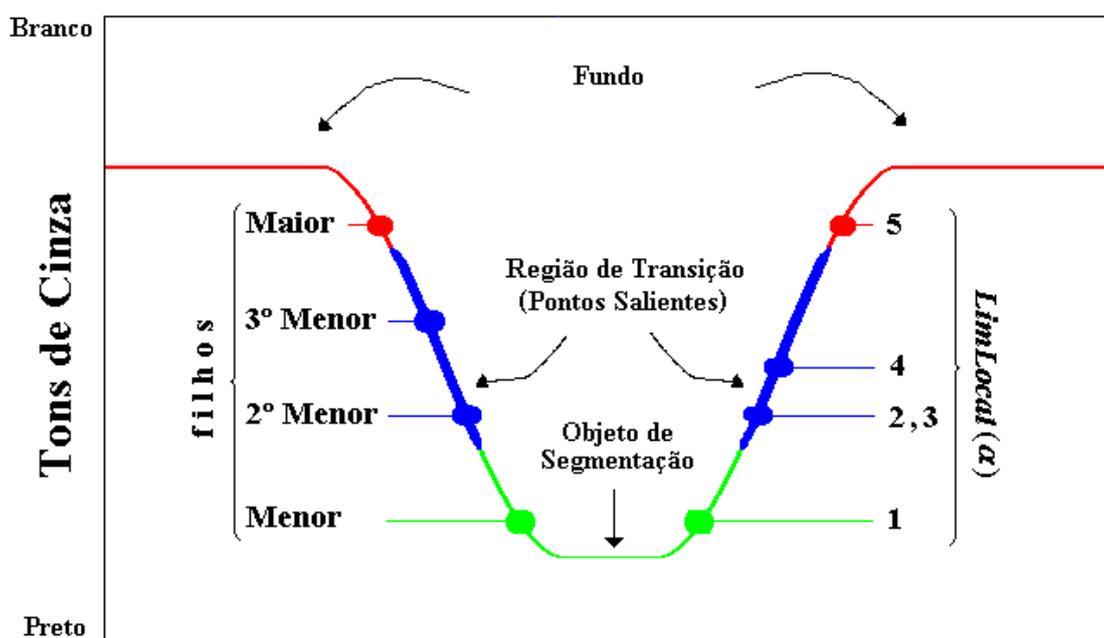


Figura 4.22: Posição do limitador local em relação ao relevo da imagem do envelope postal.

Os limitadores locais são numerados segundo a sua posição – mais provável – no relevo (dos tons de cinza) da imagem do envelope postal, conforme mostra a Tabela 4.18.

Esses limitadores foram apresentados no capítulo anterior (Equações (3.11) à (3.15)). A posição do terceiro limitador proposto, na verdade, não é a mesma do segundo limitador, ela pode estar localizada entre o primeiro e o quarto limitador.

**Tabela 4.18: Os limitadores locais ( $LimLocal(\alpha)$ ) utilizados nos experimentos.**

Tipo	$LimLocal(\alpha)$
1	$I_{SET}(menor)$
2	$I_{SET}(2^\circ menor)$
3	$(I_{SET}(menor) + I_{SET}(2^\circ menor) + I_{SET}(3^\circ menor))/2$
4	$(I_{SET}(i, j) + I_{SET}(i, j + 1) + I_{SET}(i + 1, j) + I_{SET}(i + 1, j + 1))/2$
5	$I_{SET}(maior)$

Os resultados dos experimentos realizados nas bases de imagens reais e sintetizadas para avaliar a sensibilidade do algoritmo em relação ao limitador local  $LimLocal(\alpha)$  são apresentados na Tabela 4.19 e na Tabela 4.20, respectivamente.

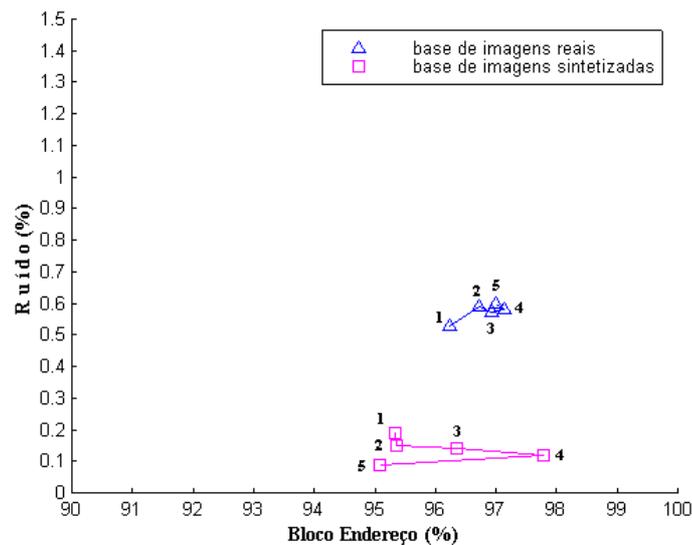
**Tabela 4.19: Resultados obtidos com os arranjos de  $LimLocal(\alpha)$ , na base de imagens reais, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$LimLocal(\alpha)$				
	1	2	3	4	5
<b>Bloco Endereço</b>	96,24 ± 5,87	96,72 ± 05,79	96,93 ± 05,85	97,13 ± 05,86	97,00 ± 06,24
<b>Selo</b>	25,01 ± 18,86	25,39 ± 19,16	25,35 ± 19,17	25,44 ± 19,29	25,28 ± 19,35
<b>Carimbo</b>	87,56 ± 18,81	88,06 ± 19,05	88,02 ± 19,15	88,12 ± 19,36	87,05 ± 20,52
<b>Ruído</b>	00,53 ± 00,73	00,59 ± 00,90	00,57 ± 00,85	00,58 ± 00,88	00,60 ± 00,90

**Tabela 4.20: Resultados obtidos com os arranjos de  $LimLocal(\alpha)$ , na base de imagens sintetizadas, mantendo-se todos os outros parâmetros fixos.**

$(\mu \pm \sigma)$ (%)	$LimLocal(\alpha)$				
	1	2	3	4	5
<b>Bloco Endereço</b>	95,34 ± 04,97	95,35 ± 05,08	96,35 ± 04,73	97,78 ± 04,76	95,07 ± 08,47
<b>Selo</b>	31,82 ± 20,26	32,45 ± 20,49	32,70 ± 20,71	33,03 ± 20,86	32,43 ± 20,49
<b>Carimbo</b>	90,92 ± 15,08	90,44 ± 15,22	91,97 ± 15,17	92,85 ± 15,67	84,98 ± 18,35
<b>Ruído</b>	00,19 ± 00,42	00,15 ± 00,38	00,14 ± 00,37	00,12 ± 00,35	00,09 ± 00,31

Os resultados referentes à taxa de retorno do bloco endereço e ruído da Tabela 4.19 e da Tabela 4.20 são apresentados em um gráfico da Figura 4.23 para facilitar o entendimento dos dados.



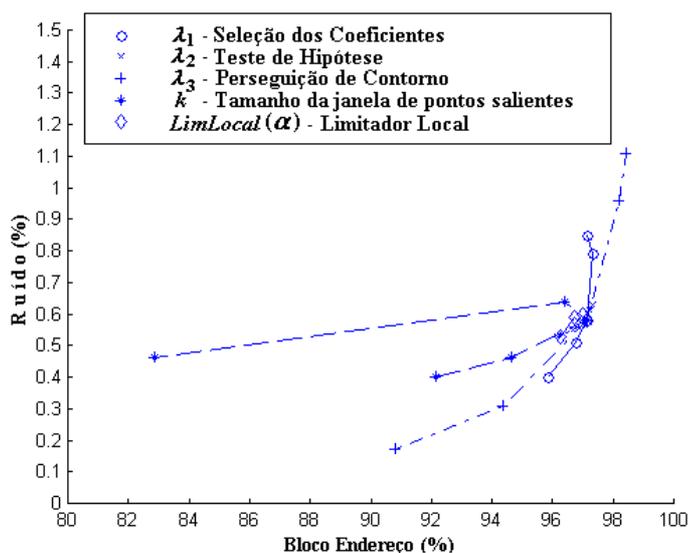
**Figura 4.23:** Gráfico das curvas do resultado dos experimentos do algoritmo de segmentação com relação aos arranjos do limitador local  $LimLocal(\alpha)$ .

Observando-se as duas curvas da Figura 4.23, percebe-se que não houve variação significativa no ruído em relação a todos os arranjos de  $LimLocal(\alpha)$  propostos aqui. Entretanto, a variação do ruído na base de imagens reais não teve um comportamento aproximadamente linear como o observado na base de imagens sintetizadas. Esta variação desordenada do ruído na base de imagens reais pode ser justificada pelos elementos existentes na transição entre os objetos de segmentação e o fundo do envelope ter sido incorporado aos objetos de segmentação ou não (exemplos são as imagens da Figura 4.15–d e da Figura 4.17–d). Portanto, uma análise mais detalhada a esse respeito deve ser realizada para identificar essa hipótese – mais uma investigação futura é proposta. O limitador local que apresentou o melhor resultado na taxa de retorno do bloco endereço (nas duas bases de imagens reais e sintetizadas) foi o arranjo que é baseado na média dos 4 *pixels* filhos referenciados pelo ponto saliente (que é o arranjo Tipo 4 da Tabela 4.18 e, é o valor do parâmetro  $LimLocal(\alpha)$  para o melhor conjunto de parâmetros mostrado na Tabela 4.3).

#### 4.5.6 Resumo da Análise da Sensibilidade

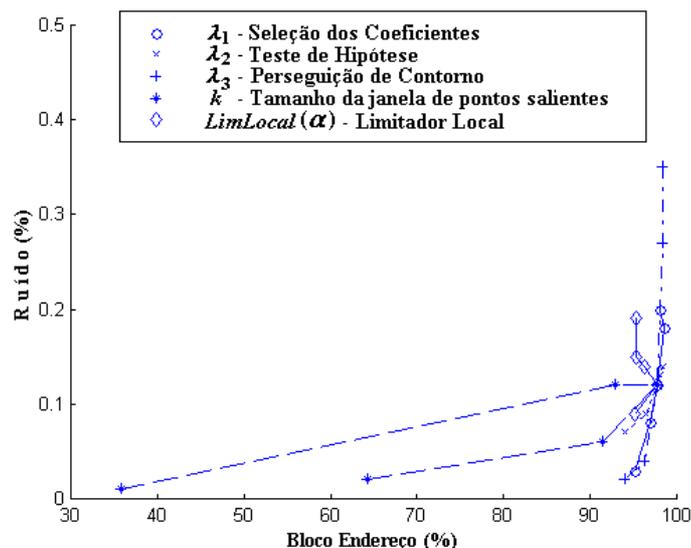
Nas cinco subseções anteriores foram realizadas análises sobre cada um dos parâmetros em separado. Para cada parâmetro colocado em análise foram realizados cinco testes, e, então,

os resultados destes foram plotados em gráficos e analisados. Através desta análise, foi possível estabelecer relações entre o valor do parâmetro e a taxa de retorno do bloco endereço e o ruído. Este número (cinco) de valores de parâmetros utilizados para gerar um espaço de estudo pode não ser tão representativo, mas foi suficiente para descrever a curva que determina o comportamento dos resultados do algoritmo de segmentação em função de cada parâmetro, apesar da alta correlação existente entre os eles.



**Figura 4.24: Resultados dos experimentos utilizados para a análise de sensibilidade do método proposto sobre a base de imagens reais.**

A Figura 4.24 e a Figura 4.25 ilustram todos os testes realizados sobre a base de imagens reais e sintetizadas, respectivamente. Estas figuras visam resumir a análise de sensibilidade executada nesta seção (Seção 4.5). Através das Figura 4.24 e a Figura 4.25 é possível identificar qual é a sensibilidade dos parâmetros no método proposto, mostrar como esses parâmetros funcionam como pontos de controle sobre o algoritmo, e explorar em uma análise futura tais parâmetros. O parâmetro de limitação local  $LimLocal(\alpha)$  não apresentou grande sensibilidade aos possíveis arranjos apresentados aqui. O tamanho  $k$  da janela de pontos salientes apresentou um ponto de saturação ( $k = 8$ ). Os parâmetro  $\lambda_2$  e  $\lambda_3$  mostraram ser proporcionais ao ruído e a taxa de retorno do bloco endereço. Já o parâmetro  $\lambda_1$  apresenta um ponto de saturação, porém com um aumento significativo do ruído.



**Figura 4.25: Resultados dos experimentos utilizados para a análise de sensibilidade do método proposto sobre a base de imagens sintetizadas.**

Aproveitando a análise de sensibilidade realizada aqui é possível estabelecer conclusões, através da observação dos resultados, para o método de segmentação aqui proposto: a medida em que se tem um aumento na taxa de retorno do bloco endereço tem-se também um aumento no ruído, que é indesejável (o ruído pode-se misturar aos objetos de segmentação). Por outro lado: a medida em que se tem uma redução no ruído, tem-se também uma redução na taxa de retorno do bloco endereço, que pode comprometer a identificação do bloco endereço (os objetos de segmentação podem não ser extraídos). Então, o ponto ideal para o método proposto deve levar em conta estas duas restrições, isto é, esse ponto ideal deve ser aquele que seja capaz de permitir que o bloco endereço seja identificado, tanto por uma alta taxa de retorno do bloco endereço, quanto por um baixo ruído.

## 4.6 Considerações sobre a Complexidade Computacional

A complexidade computacional do algoritmo é diretamente proporcional ao número de *pixels* da imagem do envelope postal, ou seja,  $O(n.m)$ , onde  $n$  e  $m$  são a largura e altura da imagem e o produto  $n.m$  é o número de *pixels* da imagem.

Embora esta complexidade computacional seja a mesma em todos os cinco passos do método, a quantidade de informação manipulada em cada um deles varia: o primeiro passo tem como entrada a imagem original ( $n \times m$ ) e como saída os canais de coeficientes *Wavelet* resultantes do algoritmo de decomposição de Mallat que tem dimensão  $n/2 \times m/2$ ; o segundo passo manipula os coeficientes *Wavelet* selecionando os mais significativos e executa uma interseção posicional entre as bandas de alta frequência, reduzindo significativamente o número de informações; o terceiro passo trabalha com uma imagem janelada ou com um histograma de  $(n.m/k^2)$  janelas separando as regiões segundo a concentração de pontos salientes dentro destas janelas; o quarto passo simplesmente projeta os *pixels* filhos dos pontos salientes pertencentes as janelas rotuladas como de alta densidade na imagem original. Porém este passo precisa também zerar os *pixels* que não foram projetados, o que faz com que este passo trabalhe com todos os  $(n \times m)$  *pixels* da imagem; por último o algoritmo de perseguição de contorno, a etapa mais elaborada do método, trabalha com uma quantidade de *pixels* menor, que foram filtrados, e reconstrói o contorno.

**Tabela 4.21: Tempo de processamento e esforço computacional do algoritmo de segmentação, codificado em C++, em um PIII dual de 1.2GHZ com 512MB de memória RAM, sobre a plataforma RedHat Linux.**

Passo	Tempo de Processamento (segundos)		Esforço Computacional de Cada Etapa (%)
	Base de Imagens Reais	Base de Imagens Sintetizadas	
1	3,565 ± 1,207	3,467 ± 1,118	55,00 (55,35)
2	1,083 ± 0,345	1,075 ± 0,343	17,00 (16,99)
3	0,304 ± 0,097	0,303 ± 0,098	05,00 (04,78)
4	0,994 ± 0,316	0,985 ± 0,316	16,00 (15,58)
5	0,469 ± 0,150	0,459 ± 0,150	07,00 (7,30)
Geral	6,736 ± 2,193	6,605 ± 2,168	100,00

Todo este método foi implementado em linguagem C++. Os experimentos foram executados em um ambiente dedicado em um Intel Pentium-III dual de 1.2 GHZ, com 3GB de memória RAM sobre a plataforma RedHat Linux. A Tabela 4.21 ilustra o tempo de processamento e o esforço computacional do algoritmo de segmentação em cada etapa. Este tempo de processamento exibido na Tabela 4.21 é o tempo de execução do algoritmo de segmentação sobre uma única imagem. Em média o tempo total de execução de um experimento na base de imagens reais (200 imagens) é de 22 minutos, enquanto que um experimento na base de imagens sintetizadas (2000 imagens) cerca de 220 minutos. O esforço computacional apresentado na Tabela 4.21 pelo primeiro passo do método

proposto (Transformada *Wavelet*) representa mais da metade (aproximadamente 55%) de todo o processo. No entanto, a implementação da transformada *Wavelet* segundo o algoritmo de Mallat não foi otimizada e este esforço poderia ser reduzido em cerca de três vezes, reduzindo, dessa forma, significativamente o tempo de processamento. A análise de sensibilidade realizada na Seção 4.5 consumiu cerca de 50 horas de processamento neste ambiente.

## 4.7 Comentários Finais

O objetivo principal do método proposto, que é separar os objetos existentes (bloco endereço, selo e carimbo) no envelope postal do fundo da imagem para localização do bloco endereço, foi atingido e apresentou resultados promissores nos experimentos. O tamanho da base foi suficiente para avaliar os elementos necessários existentes nas imagens dos envelopes postais. O método proposto se mostrou robusto no sentido de que a base de imagens reais continha tanto fundos brancos como fundos complexos. A análise da sensibilidade dos resultados apresentada é promissora. No entanto, faz-se necessário uma análise que possa modelar o alto grau de dependência existente entre estes parâmetros, criado através do encadeamento existente entre os 5 passos principais da metodologia proposta. Algumas análises e parte dos resultados aqui apresentados foram publicados por Menoti *et al* em [MENOTI *et al* 2003a] e em [MENOTI *et al* 2003b].

## Capítulo 5

### 5. Conclusões e Trabalhos Futuros

Neste capítulo são apresentadas as contribuições originais, as conclusões adquiridas durante o trabalho, sugestões para trabalhos futuros e alguns anseios.

As contribuições originais deste trabalho de pesquisa podem ser resumidas em:

- Detecção de pontos salientes através de evidências localizados através dos canais de alta frequência da transformada *Wavelet* discreta;
- Segmentação de regiões salientes similares por teste de hipótese estatístico baseado nos próprios dados;
- Projeção reversa dos pontos salientes e perseguição de contorno para inclusão dos pontos no espaço anterior;
- O método como um todo para a segmentação do bloco endereço e ajustes dos parâmetros relacionados ao algoritmo.
- Metodologia para avaliação quantitativa/objetiva da segmentação baseada em uma abordagem *ground truth (pixel a pixel)*. Esta metodologia consiste em utilizar imagens de referência como a segmentação desejada comparando-a com a segmentação obtida.

O trabalho proposto tem como objetivo a segmentação de imagens de envelopes postais para a localização do bloco endereço utilizando características do espaço *Wavelet*. Depois de apresentada a metodologia de segmentação e os experimentos para validá-la, estas são as conclusões atingidas:

- O objetivo principal do método proposto; a segmentação dos envelopes postais para a localização do bloco endereço foi atingida e apresentou resultados promissores. Alcançou-se através do método de segmentação cerca de 97% dos *pixels* do bloco endereço, na média, com menos de 1% de ruído. O método foi robusto no sentido de que foi capaz de segmentar envelopes com vários tipos de diagramações, fundos de envelopes, tamanho e tipos de regiões escritas;
- A maneira proposta para analisar o resultado da segmentação permite avaliar, de forma objetiva (quantitativa), os objetos extraídos. Entretanto, existe uma dificuldade dessa metodologia em determinar a posição do ruído. Esta dificuldade pode ser superada com a detecção de retângulos de contorno e submissão desses a um sistema de reconhecimento. A metodologia de avaliação proposta aqui mostrou-se mais completa que avaliações subjetivas (de forma visual) como a maioria dos trabalhos relacionados, apresentados no Capítulo 2;
- O tamanho da janela de pontos salientes que apresentou o melhor desempenho em relação a taxa de retorno do bloco endereço e ao ruído foi  $k = 8$ . As bases de imagens continham objetos de segmentação de vários tamanhos e espessuras, e esse foi o valor que se mostrou robusto, neste sentido;
- A análise de sensibilidade realizada mostrou a influência existente, de forma isolada, de cada parâmetro;
- A quantidade, bem como a variedade, de imagens que compõem a base de imagens reais foi suficiente para validar, nos experimentos, as hipóteses e suposições levantadas durante a elaboração do método proposto. A base de imagens sintetizadas veio concretizar os resultados obtidos na base de imagens reais. Outra conclusão importante em relação às bases de imagens é que a maneira como a base de imagens sintetizadas foi construída não teve influência nos resultados dos experimentos apresentados nessa base com

relação aos resultados apresentados pela base de imagens reais. Ou seja, não houve grande divergência entre os resultados dos experimentos realizados nas duas bases.

Os trabalhos futuros propostos nesta área de pesquisa são os seguintes:

- Analisar simultaneamente o comportamento dos parâmetros  $\lambda_1$ ,  $\lambda_2$  e  $k$ , que determinam o funcionamento da identificação de pontos salientes ( $\lambda_1$ : quantidade de pontos salientes selecionados) e a rotulação das janelas de pontos salientes ( $\lambda_2$ : nível de significância no teste de hipótese, e  $k$ : tamanho da janela de pontos salientes);
- O algoritmo de rotulação de janelas de pontos salientes, apresentado na Seção 3.3, em regiões de alta e baixa densidade, pode ser estendido para que ele rotule as janelas em regiões de várias densidades;
- Um estudo sobre o funcionamento das razões  $\beta$  (Equação (3.16)) e  $\varepsilon$  (Equação (3.17)), definidas na Seção 3.5, que foram definidas empiricamente, para realizarem restrições locais ao algoritmo de perseguição de contorno deve ser executado;
- Analisar os processos de identificação de pontos salientes e de rotulação das janelas de pontos salientes em várias escalas, aproveitando as propriedades da transformada *Wavelet*, utilizando modelos piramidais e em árvore;
- Analisar os efeitos dos cinco arranjos propostos para o limitador local  $LimLocal(\alpha)$  sobre os resultados da taxa de retorno e do ruído;
- Um procedimento de extração de componentes (em forma de caixas retangulares) deve ser aplicado às imagens segmentadas para determinar os componentes candidatos a bloco endereço.

- Analisar o encadeamento dos parâmetros do método proposto, com experimentos sobre as bases de imagens;
- Estender a metodologia de segmentação para documentos complexos e imagens com cenas naturais;

Após a realização deste trabalho de pesquisa, surgem alguns anseios:

- Criar um sistema que contenha todas as fases necessárias ao processo de automação postal para os CB, desde a aquisição da imagem do envelope postal até o ponto final, onde as informações pertinentes a essa automação sejam geradas. Este sistema visa vincular os resultados de pesquisa com aplicações necessárias ao mercado;
- Uma forma de distinguir e identificar seres humanos é a impressão digital do polegar direito. Outras técnicas estão sendo estudadas para esta tarefa, como os vasos sanguíneos presentes na íris do olho. A abordagem de segmentação de objetos presentes em envelopes postais proposta neste trabalho poderia ser utilizada para extrair estes vasos sanguíneos. Porque os objetos de segmentação, particularmente o bloco endereço e o carimbo, deste trabalho têm a mesma natureza dos vasos sanguíneos.

## 6. Referências Bibliográficas

[CHENG *et al* 1997] CHENG, H.; BOUMAN, C. A.; ALLEBACH, J. P. “Multiscale Document Segmentation”, in *Proceedings IS&T 50th Annual Conference*, Cambridge, EUA, 18 a 23 de Maio de 1997, páginas 417-425.

[CHENG, BOUMAN 1998] CHENG, H.; BOUMAN, C.A. “Trainable context model for multiscale segmentation”, In *Proceedings IEEE of the International Conference on Image Processing*, Chicago, EUA, Outubro de 1998, páginas 4-7.

[CHENG, BOUMAN 2001] CHENG, H.; BOUMAN, C. A. “Multiscale Bayesian Segmentation Using a Trainable Context Model”, *IEEE Transactions on Image Processing*, volume 10, número 10, Abril de 2001, páginas 511-525.

[CHOI, BARANIUK 2001] CHOI, H.; BARANIUK, R. G. “Multiscale image segmentation using wavelet-domain hidden markov models”, *IEEE Transactions on Image Processing*, volume 10, número 9, Setembro de 2001, páginas 1309-1321.

[ECT 2001] ECT – Empresa Brasileira de Correios e Telégrafos, Correios do Brasil, Banco de Imagens, Departamento de Engenharia, Divisão de Desenvolvimento Tecnológico, Convênio PUCPR-ECT, 2001.

[ECT 2003] ECT – Empresa Brasileira de Correios e Telégrafos, Correios do Brasil, [www.correios.com.br](http://www.correios.com.br), , acessado em 8 de Abril de 2003.

[FACON 2001] FACON, J. “*Metodologia de Avaliação de Abordagens de Segmentação de Imagens*”, Pontifícia Universidade Católica do Paraná, Programa de Pós-Graduação em Informática Aplicada, Tese de Professor Titular, Março de 2001.

[HARALICK 1994] HARALICK, R. M. “Document Image Understanding: Geometric and Logical Layout”, *CVPR94: Proceedings of Computer Vision and Pattern Recognition Conference*, IEEE, 1994, páginas 385-390.

[JAIN, BHATTACHARJEE 1992a] JAIN, A. K.; BHATTACHARJEE, S. K. “Address Block Location on Envelopes Using Gabor Filters”, *Pattern Recognition*, volume 25, número 12, 1992, páginas 1459-1477.

[JAIN, BHATTACHARJEE 1992b] JAIN, A. K.; BHATTACHARJEE, S. K. “Address Block Location on envelopes using Gabor filters: supervised method”, *In International Conference on Pattern Recognition*, 1992, páginas 264-267.

[MALLAT 1989] MALLAT, S. “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, volume 11, número 7, Julho de 1989, páginas 674-693.

[MENOTI *et al* 2003] MENOTI, D.; BORGES, D. L.; FACON, J.; BRITTO JR, A. S. “Segmentation of Postal Envelopes for Address Block Location: an approach based on features selection in wavelet space”, *in Proceedings of the 7<sup>th</sup> International Conference on Document Analysis and Recognition (ICDAR) 2003*, Edimburgo, Escócia, aceito para publicação, 3 a 6 de Agosto, IEEE, 2003.

[MENOTI *et al* 2003] MENOTI, D.; BORGES, D. L.; BRITTO JR, A. S. “Salient Features and Hypothesis Testing: evaluation a novel approach for segmentation and address block location”, *in Workshop on Document Image Analysis and Retrieval (DIAR) of Computer Vision and Pattern Recognition (CVPR) 2003*, Madison, Wisconsin, EUA, aceito para publicação, 21 de Junho, IEEE, 2003.

[PHPM 2003] *Postal History and Postal Mechanisation*, <http://www.argonet.co.uk/users/devekeyb/Phil.html>, acessado em 7 de Abril de 2003.

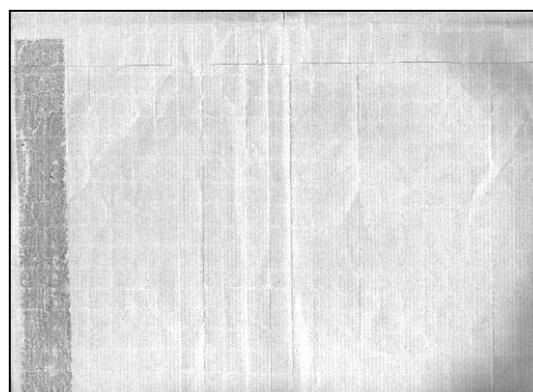
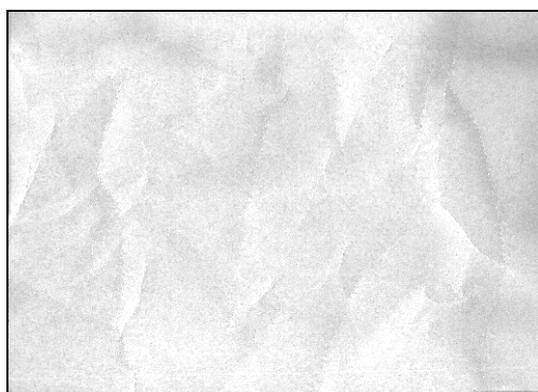
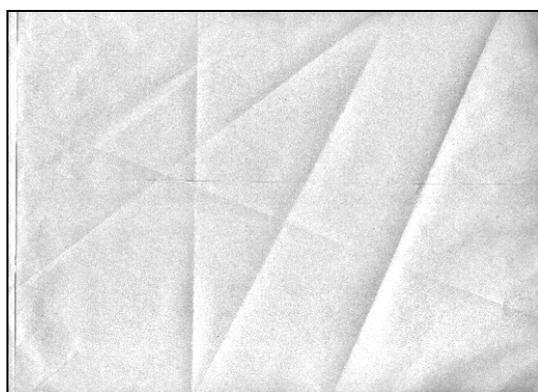
[PRESS *et al* 1996] PRESS, W. H.; TEUKOLSKY, S.A.; VETTERLING, W. T.; FLANNERY, B. P. “*Numerical Recipes in C: the art of scientific computing*” – Segunda Edição: Cambridge University Press, Reino Unido, 1996, 994p.

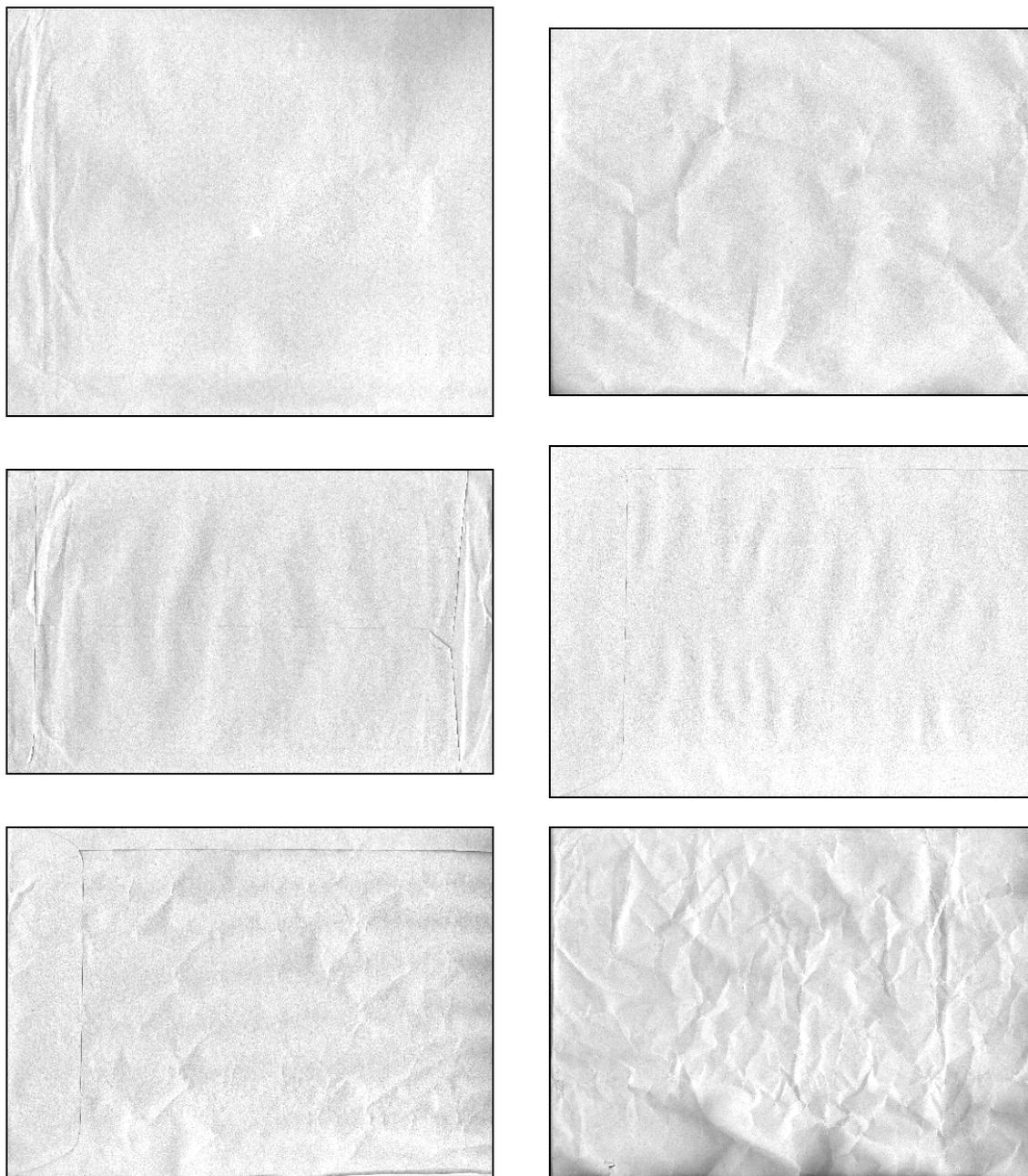
- [SRIHARI, KUEBERT 1997] SRIHARI, N.; KUEBERT, E. J. “Integration of Hand-Written Address Interpretation Technology into the United States Postal Service Remote Computer Reader System”, *In Proceedings of the Fourth International Conference on Document Analysis and Recognition (ICDAR)*, Abril de 1997, páginas 892-896.
- [STRANG, NGUYEN 1996] STRANG, G.; NGUYEN, T. “*Wavelets and Filter Banks*”, Wellesley-Cambridge Press, EUA, 1996.
- [TOU, GONZALEZ 1974] TOU, J. T.; GONZALEZ, R. C.; “*Pattern recognition Principles*”, Addison-Wesley Publishing Company, Reading, EUA, 1974.
- [WOLF, NIEMANN 1997] WOLF, M.; NIEMANN, H. “Form-Based Localization of the Destination Address Block on Complex Envelopes”, *In Proceedings of the Fourth International Conference on Document Analysis and Recognition (ICDAR)*, Abril de 1997, páginas 908-913.
- [WOLF *et al* 1997] WOLF, M.; NIEMANN, H.; SCHMIDT, W. “Fast Address Block Location on Handwritten and Machine Printed Mail-piece Images”, *In Proceedings of the Fourth International Conference on Document Analysis and Recognition (ICDAR)*, Abril de 1997, páginas 753-757.
- [WU *et al* 1997] WU, V.; MANMATHA, R.; RISEMAN, E. “Automatic text detection and recognition”, *In Proceedings of Image Understanding Workshop*, 1997, páginas 707-712.
- [YU *et al* 1997] YU, B.; JAIN, A.; MOHIUDDIN, M. “Address block location on complex mail pieces”, *In Proceedings of the Fourth International Conference on Document Analysis and Recognition (ICDAR)*, Abril de 1997, páginas 897-901.

## Apêndices

### A. Tipos de fundos de envelopes complexos sem objetos de segmentação

As imagens apresentadas nesta seção correspondem aos tipos de fundos de envelopes complexos sem objetos de segmentação que foram utilizados para gerar a base de imagens sintetizadas. Estes fundos são considerados complexos porque, apresentam uma variação significativa dos tons cinza e, também apresentam irregularidades no seu corpo. Na verdade estes fundos correspondem à face reversa de alguns envelopes de correspondências comerciais. A face reversa foi escolhida, pois a mesma não contém nenhum objeto que se assemelhe aos objetos do envelope que poderiam vir a prejudicar (ou interferir) na avaliação do método proposto.





**Figura A.1: Tipos de Fundos de envelopes complexos utilizados para a construção da base de imagens sintetizadas.**