

Renata Nagima Imada

*Reconhecimento de contorno de edifício em  
imagens de alta resolução usando os  
momentos complexos de Zernike*

Presidente Prudente - SP, Brasil

Outubro de 2014

Renata Nagima Imada

*Reconhecimento de contorno de edifício em  
imagens de alta resolução usando os  
momentos complexos de Zernike*

Dissertação apresentada ao Programa de Pós-graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da UNESP como requisito para a obtenção do Título de Mestre em Matemática Aplicada e Computacional.

Orientador:

Prof. Dr. Aluir Porfírio Dal Poz

PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA E COMPUTACIONAL  
FACULDADE DE CIÊNCIAS E TECNOLOGIA DE PRESIDENTE PRUDENTE  
UNIVERSIDADE ESTADUAL PAULISTA

Presidente Prudente - SP, Brasil

Outubro de 2014

## FICHA CATALOGRÁFICA

I29r Imada, Renata Nagima.  
Reconhecimento de contorno de edifício em imagens de alta resolução usando os momentos complexos de Zernike / Renata Nagima Imada. - Presidente Prudente : [s.n], 2014  
xi, 71 f. : il.

Orientador: Aluir Porfírio Dal Poz  
Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia  
Inclui bibliografia

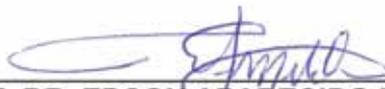
1. Contorno de edifício. 2. Momentos de Zernike. 3. Segmentação de Imagem. I. Imada, Renata Nagima. II. Dal Poz, Aluir Porfírio. III. Universidade Estadual Paulista. Faculdade de Ciências e Tecnologia. IV. Título.

**BANCA EXAMINADORA**



---

**PROF. DR. ALUIR PORFIRIO DAL POZ**  
ORIENTADOR



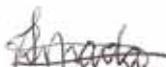
---

**PROF. DR. EDSON APARECIDO MITSHITA**  
UFPR



---

**PROF. DR. AYLTON PAGAMISSE**  
UNESP/FCT



---

**RENATA NAGIMA IMADA**

Presidente Prudente (SP), 24 de outubro de 2014.

Resultado:

Aprovado

*Aos meus queridos pais,  
Nelson e Rosa,  
dedico!*

# *Agradecimentos*

Agradeço primeiramente a Deus, por seu amor incondicional, permitindo-me vivenciar este momento de aprendizagem, por me guiar e dar forças para enfrentar as dificuldades encontradas. Obrigada também minha Mãe Maria por sua preciosa intercessão.

Agradeço imensamente à minha família. Em especial, meus pais Nelson e Rosa, razão do meu existir e que formaram a pessoa que sou hoje, sempre me ensinando e orientando. Meu irmão Rogerio, que sempre me ajudou quando precisei, principalmente ao me levar e buscar tantas vezes na faculdade. Passamos por um problema de saúde muito difícil durante o desenvolvimento deste trabalho, mas tudo o que vivemos serviu para fortalecer ainda mais nosso amor e união. Muito obrigada pelo apoio de todos.

Agradeço também ao meu namorado Vinicius, meu melhor amigo e companheiro de estudos durante toda a graduação. Obrigada pelo amor, carinho e compreensão, por sempre me incentivar e acreditar em mim, não me deixando desanimar.

Muito obrigada ao meu orientador Prof. Dr. Aluir Porfírio Dal Poz pela orientação, compreensão, discussões e conhecimentos compartilhados para meu crescimento científico.

Estendo meus agradecimentos a todos os professores que passaram por minha vida acadêmica, obrigada pela inspiração e contribuições para minha formação.

Não posso deixar de agradecer também aos amigos da graduação, de todas as turmas do PósMac que tive contato e do Grupo de Pesquisa em Fotogrametria. São muitos para citar todos, mas em especial agradeço ao Wesley, Clóvis, Vanessa, Érico e Tatiana, que me ajudaram com os programas, implementações, materiais e dados utilizados no trabalho. Com certeza, a presença de todos tornou este período mais agradável, obrigada por compartilhar conhecimentos, expectativas, ansiedades e risos.

Gostaria de agradecer à CAPES e à FAPESP pelo auxílio financeiro para a realização desta pesquisa.

Enfim, agradeço a todos que de alguma maneira estiveram ao meu lado e me ajudaram a chegar até aqui.

*“Tudo posso  
naquele que me fortalece!”*  
***Fl 4,13***

# *Resumo*

Nesta pesquisa foi estudado um método de reconhecimento de contornos de telhado de edifícios em imagens digitais de alta resolução, que classifica-os com relação à sua forma. O método baseia-se nos momentos de Zernike, que são baseados nos polinômios ortogonais de Zernike, em que cria-se um vetor de características para cada região da imagem, que deve ser previamente segmentada de maneira que seus objetos sejam divididos em diferentes regiões. Este método para a descrição de forma baseia-se na área do objeto de interesse e possui a característica dos momentos serem invariantes em relação às transformações geométricas de rotação, translação e escala, que o torna atrativo para o problema de análise de imagem proposto. Desse modo, foi criada uma base de dados contendo esboços (ou modelos) de possíveis aparições de contornos de telhado de edifício numa dada cena, para que seja associado também um vetor de características de Zernike para estes esboços. Assim, a distância euclidiana entre este vetor e o vetor de características calculado a partir de uma região segmentada na imagem, permite dizer se a região dada corresponde a um contorno de edifício ou a outro objeto. A capacidade de discriminação do método proposto entre diferentes formas de edifícios, e também entre formas de edifícios e não edifícios foi avaliada experimentalmente e mostrou resultados positivos.

Palavras-Chave: *Imagens de alta resolução, segmentação de imagem, contorno de edifícios, momentos de Zernike, caracterização de forma.*

# *Abstract*

## **Recognition of building roof contour in high-resolution images using Zernike moments**

In this research, a method of recognition of building roof contours in high-resolution digital images which classifies them with respect to their form was studied. The method is based on Zernike moments, which are based on orthogonal Zernike polynomials and it creates a feature vector for each image region. The image segmentation has to be made first to define different regions for its objects. This method for shape analysis is based on the object area of interest and the moments has the characteristic of being invariant under geometric transformations of rotation, translation and scaling, this makes it attractive to the proposed image analysis problem. Thus, a database containing sketches (or models) of possible appearances of building roof contours in a given scene was created, so a Zernike feature vector was also associated for these sketches. Therefore, the Euclidean distance between this vector and the feature vector calculated from a segmented region in the image lets say if the given region corresponds to a building contour or other object. The capacity of the proposed method in discriminating different building shapes and also in discriminating building shapes from non-building shapes was evaluated experimentally and it showed positive results.

Key-Words: *High-resolution images, image segmentation, building roof contour, Zernike moments, shape characterization.*

# *Sumário*

<b>Lista de Figuras</b>	p. iv
<b>Lista de Tabelas</b>	p. vi
<b>1 INTRODUÇÃO</b>	p. 1
1.1 Considerações Iniciais . . . . .	p. 1
1.2 Objetivos . . . . .	p. 4
1.3 Estrutura da dissertação . . . . .	p. 4
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	p. 5
2.1 Fundamentos de Processamento Digital de Imagens . . . . .	p. 5
2.1.1 Estrutura e Resolução de Imagens Digitais . . . . .	p. 5
2.1.2 Histograma de uma Imagem . . . . .	p. 7
2.1.2.1 Histograma Unidimensional . . . . .	p. 8
2.1.2.2 Histograma Cumulativo . . . . .	p. 9
2.1.3 Filtragem Espacial . . . . .	p. 10
2.1.3.1 Filtragem com a Janela Móvel ou Máscara . . . . .	p. 11
2.1.3.2 Filtros Passa Baixas ou de Suavização . . . . .	p. 12
2.1.3.3 Filtros Passa Altas ou de Realce de Bordas . . . . .	p. 16
2.1.4 A Cor no Processamento Digital de Imagens . . . . .	p. 17
2.1.4.1 O Espaço RGB . . . . .	p. 18
2.1.4.2 O Modelo HSI . . . . .	p. 19
2.2 Segmentação de Imagens . . . . .	p. 21

2.2.1	Regiões <i>versus</i> Bordas . . . . .	p. 21
2.2.2	Segmentação de Imagem por Regiões . . . . .	p. 22
2.2.3	Métodos Pontuais (pixel-a-pixel) . . . . .	p. 23
2.2.3.1	Método do Limiar Percentual . . . . .	p. 23
2.2.3.2	Método da Moda . . . . .	p. 24
2.2.3.3	Método da Limiarização Iterativa . . . . .	p. 25
2.2.3.4	Método da Limiarização Adaptativa . . . . .	p. 25
2.2.3.5	Algumas Limitações . . . . .	p. 26
2.2.4	Métodos de Análise de Vizinhança . . . . .	p. 26
2.2.4.1	Divisão e Fusão . . . . .	p. 27
2.2.4.2	Crescimento de Regiões . . . . .	p. 28
2.2.5	Métodos de Perseguição do Contorno . . . . .	p. 28
2.3	Polinômios Ortogonais em Várias Variáveis . . . . .	p. 29
2.3.1	Polinômio de Zernike . . . . .	p. 30
2.4	Momentos de uma Imagem Digital . . . . .	p. 32
2.4.1	Momentos de Zernike . . . . .	p. 33
2.5	Descritor de forma de Zernike . . . . .	p. 37
<b>3</b>	<b>METODOLOGIA</b> . . . . .	p. 39
3.1	Recursos Computacionais e Dados Utilizados . . . . .	p. 39
3.2	Método . . . . .	p. 40
3.2.1	Reconhecimento de Contornos de Edifício usando Momentos de Zernike . . . . .	p. 40
3.2.2	Índices de Avaliação dos Resultados . . . . .	p. 45
<b>4</b>	<b>RESULTADOS</b> . . . . .	p. 47
4.1	Experimentos para Escolha do Método de Segmentação . . . . .	p. 48
4.2	Verificação da Invariância por Rotação, Escala e Translação . . . . .	p. 51

4.3	Cena Simulada . . . . .	p. 56
4.4	Cenas Reais . . . . .	p. 58
4.4.1	Recorte 1 . . . . .	p. 58
4.4.2	Recorte 2 . . . . .	p. 60
4.4.3	Recorte 3 . . . . .	p. 61
4.4.4	Comparação com Outros Métodos . . . . .	p. 64
<b>5</b>	<b>CONCLUSÃO</b>	p. 66
	<b>Referências</b>	p. 69

# *Lista de Figuras*

1	Exemplo de uma imagem e sua forma digital. . . . .	p. 6
2	Comparação de resolução radiométrica. . . . .	p. 7
3	Distribuição de níveis de cinza de uma imagem através de seu histograma. . . . .	p. 8
4	Comparação do contraste de imagens a partir de histogramas. . . . .	p. 9
5	Imagem de uma plantação de cana e seu histograma. . . . .	p. 9
6	Imagem de uma região costeira e seu histograma. . . . .	p. 10
7	Comparação entre dois tipos de histograma. . . . .	p. 10
8	Esquema da máscara para filtragem de uma imagem. . . . .	p. 12
9	Resultados de filtros da média. . . . .	p. 13
10	Função Gaussiana bidimensional com média 1, centrada em (0,0) e $\sigma = 1$ . . . . .	p. 14
11	Exemplo da representação de uma imagem para aplicação de filtros passa baixas. . . . .	p. 15
12	Resultados de filtros da média e mediana. . . . .	p. 15
13	Resultado do filtro Laplaciano. . . . .	p. 16
14	Espectro eletromagnético. . . . .	p. 17
15	Esquema do cubo de cores RGB. . . . .	p. 18
16	Esquema da relação entre os espaços RGB e HSI. . . . .	p. 20
17	Exemplo do uso do método do limiar percentual. . . . .	p. 24
18	Histogramas de imagens que possuem objetos e fundo com diferentes valores de intensidade. . . . .	p. 24
19	Esquema do processo de limiarização adaptativa. . . . .	p. 26
20	Mapeamento de uma imagem $N \times N$ sobre o disco unitário. . . . .	p. 35
21	Fluxograma do funcionamento do método proposto. . . . .	p. 41

22	Esboços de contornos de edifício contidos no banco de dados. . . . .	p. 44
23	Imagem em escala de cinza. . . . .	p. 48
24	Histograma da Figura 23. . . . .	p. 48
25	Segmentação com dois limiares dados. . . . .	p. 49
26	Segmentação com limiares definidos automaticamente. . . . .	p. 49
27	Segmentação baseada nos contornos. . . . .	p. 50
28	Segmentação através do método de crescimento de regiões. . . . .	p. 51
29	Esboços de contornos de edifício desenhados manualmente. . . . .	p. 51
30	Esboços de contornos de edifício em forma de L. . . . .	p. 52
31	Esboços de contornos de edifício em forma retangular. . . . .	p. 52
32	Esboços de contornos de edifício em forma de quadrado. . . . .	p. 53
33	Gráfico das normas das transformações do edifício em forma de L. . . . .	p. 54
34	Gráfico das normas das transformações do edifício retangular. . . . .	p. 54
35	Gráfico das normas das transformações do edifício em forma de quadrado. . . . .	p. 55
36	Cena simulada com esboços de edifícios. . . . .	p. 56
37	Segmentação da cena simulada. . . . .	p. 57
38	Recorte 1 e sua transformação em escala de cinza. . . . .	p. 59
39	Segmentação do Recorte 1. . . . .	p. 59
40	Recorte 2 e sua transformação em escala de cinza. . . . .	p. 60
41	Segmentação do Recorte 2. . . . .	p. 60
42	Recorte 3 e sua transformação em escala de cinza. . . . .	p. 61
43	Segmentação do Recorte 3. . . . .	p. 62
44	Recorte 3 com contornos de edifícios delimitados manualmente e sua segmentação. . . . .	p. 63

## *Lista de Tabelas*

1	Normas e distâncias das transformações do edifício em forma de L . . . .	p. 53
2	Normas e distâncias das transformações do edifício retangular . . . . .	p. 54
3	Normas e distâncias das transformações do edifício em forma de quadrado	p. 55
4	Média e Desvio-padrão das normas dos vetores de características . . . . .	p. 55
5	Reconhecimento das regiões da cena simulada . . . . .	p. 57
6	Reconhecimento das regiões do Recorte 1 . . . . .	p. 59
7	Reconhecimento das regiões do Recorte 2 . . . . .	p. 60
8	Reconhecimento das regiões do Recorte 3 com segmentação automática	p. 62
9	Reconhecimento das regiões do Recorte 3 com os contornos de edifícios delimitados manualmente . . . . .	p. 63
10	Resultados dos experimentos . . . . .	p. 67

# ***1 INTRODUÇÃO***

## **1.1 Considerações Iniciais**

O desenvolvimento e a manutenção de um Sistema de Informações Geográficas (SIG) dependem da aquisição e atualização de dados espaciais. Assim, a extração de feições em imagens digitais vem sendo alvo de várias pesquisas desde a década de 1960. As edificações urbanas são objetos topográficos de grande importância para o mapeamento cadastral em escala grande e para tarefas específicas como o planejamento urbano e a análise ambiental.

É fundamental o aumento do nível de automação dos processos de compilação de edificações a partir de imagens de alta resolução, trazendo duas implicações importantes: maior agilidade e menor custo na compilação de grande massa de dados capturados por sensores diversos (imagens aéreas e de satélite, dados de varredura a laser, etc.). Isso também é importante para viabilizar ciclos mais curtos de atualização de bases de dados topográficos de edificações (SOWMYA; TRINDER, 2000).

As imagens de alta resolução facilitam a identificação de edificações, mas também impõem desafios aos métodos clássicos de análise de imagem, tais como os voltados para a classificação de imagens ao nível do pixel e para a segmentação por regiões. De fato, com o aumento da resolução espacial os edifícios podem se manifestar geometricamente de forma mais intrincada. Pode ocorrer também que diferentes materiais apareçam com assinatura espectral similar (por exemplo, telhados e solo exposto se manifestam de forma similar em imagens RGB).

Desta discussão decorre que uma melhor exploração do potencial das imagens de alta resolução depende da modelagem mais eficiente das características geométricas do objeto de interesse. As características geométricas, relacionais e espectrais de objetos de interesse podem ser melhor exploradas através de métodos de classificação baseados em objetos, podendo-se citar como exemplo Bouziani, Goita e He (2010), Bruzzone e Carlin

(2006), Hofmann (2003), Shackelford e Davis (2003). Uma deficiência compartilhada por métodos dessa categoria está relacionada com a modelagem das características da forma dos objetos de interesse. As características mais comumente utilizadas são a área, o perímetro, o invólucro convexo e a compacidade. Entretanto, estas características não são suficientes para uma completa discriminação entre objetos, além de serem sensíveis a deformações nos contornos dos objetos e não invariantes em relação a transformações geométricas (LIAO, 1993).

Numa visão estatística, os momentos são a expectativa estatística de certas funções de uma variável aleatória. Seguindo a mesma ideia, mas com uma visão diferente, o resultado de momentos dá medidas que contêm informações sobre a forma ou distribuição de  $f(x)$ . Isto faz com que momentos sejam úteis para a análise de formas no processamento de imagem, para os quais  $f(x, y)$  é a função da imagem, que pode ser uma função de nível de cinza ou binária (0 para o fundo, 1 para os objetos existentes na cena).

Teague (1980), Chong, Raveendran e Mukundan (2003), Liao (1993), Celebi e Aslandogan (2005) apresentam diferentes trabalhos sobre análise de imagens através de momentos. Mahi, Isabaten e Serief (2014) mostram um trabalho em que momentos de Zernike são utilizados para obter informações sobre a forma dos objetos e detectar os edifícios em imagens de satélite de alta resolução. Após a segmentação feita pelo método de média-móvel, a classificação é feita através de *support vector machines* (SVMs). Os momentos de Zernike também podem ser utilizados no diagnóstico de massas encontradas em imagens de mamografia (TAHMASBI; SAKI; SHOKOUHI, 2011; TAHMASBI et al., 2011; FELIPE; OLIOTI; TRAINA, 2005), bem como podem ser utilizados em oftalmologia (JANKOV et al., 2002; BISNETO et al., 2007).

Tomando o exposto acima como motivação para esta pesquisa, propõe-se estudar e avaliar experimentalmente o problema de caracterização de forma de contorno de edifício usando os momentos complexos de Zernike. Esses momentos baseiam-se na área do objeto de interesse e possuem duas características que os tornam altamente atrativos para o problema proposto (LIAO, 1993): 1) é invariante em relação a transformações geométricas de rotação, translação e escala; e 2) é pouco sensível a deformações no contorno. Na verdade, como mostra o trabalho de Khotanzad e Hong (1990), a invariância em termos de translação e escala não é uma propriedade original dos momentos de Zernike, mas ela é obtida a partir de uma transformação prévia da imagem de entrada numa outra, tendo por base os momentos geométricos. Estas propriedades de invariância implicam numa flexibilidade e nos motivam a propor uma base de dados contendo esboços de possíveis

aparções de contornos de edifício numa dada cena e associar a cada esboço um vetor de características baseados nos momentos de Zernike.

Neste trabalho, os momentos de Zernike também são utilizados para descrever a forma de objetos obtidos a partir de uma prévia segmentação da imagem. A fim de avaliar a capacidade deste método na discriminação entre diferentes formas de edifícios, e também entre formas de edifícios e não edifícios, propõe-se utilizar uma métrica (a distância Euclidiana) para medir a similaridade entre um vetor de característica contendo momentos de Zernike para uma região da imagem e um outro vetor do mesmo tipo, de cada esboço de edifício da base de dados.

Outras abordagens também são encontradas na literatura para a extração de edifícios. Chen et al. (2012) apresenta um método de detecção de edifícios em uma área urbana que combina dados LASER e imagem de satélite QuickBird. Um Modelo Digital de Superfície normalizado (MDSn) é gerado a partir dos dados LASER, obtendo os candidatos a representar edifícios. Em seguida, a vegetação é filtrada através da imagem usando o Índice de Vegetação por Diferença Normalizada (NDVI). Por fim, essas informações são combinadas e encontrados os edifícios. Da mesma forma, Marcato (2013) também combina dados LASER e imagem de alta resolução. Neste trabalho, propõe-se a reconstrução de contornos de telhados de edifícios extraídos de dados LASER, usando imagens aéreas de alta resolução e campos randômicos de Markov (MRF - *Markov Random Field*).

Thomaz (2012) apresenta um método de extração de contornos de telhados convexos de edifícios a partir de um MDSn, em que a modelagem do contorno é feita através do modelo de *snake balloon*, onde um funcional de energia é otimizado através do algoritmo de Programação Dinâmica (PD).

O trabalho de Bhadauria, Bhadauria e Kumar (2013) propõe um método em que o objetivo é extrair edifícios de imagens aéreas e de satélite usando hipóteses. Essas hipóteses são geradas a partir do algoritmo de detecção de bordas de Canny, em que a transformada de Hough produz segmentos de linha de acordo com essas bordas extraídas. Estes segmentos geram os possíveis edifícios que são verificados através das informações obtidas pela segmentação da representação HSI da imagem.

Pakizeh e Palhang (2010) sugere um método de detecção de edifícios a partir de imagens aéreas usando Transformada de Hough e informações de intensidade. Primeiro, a localização dos centros dos edifícios é estimada pela segmentação da imagem através de suas informações de intensidade e, em seguida, a aparência dos edifícios nos centros estimados é verificada pela transformada de Hough.

Singh et al. (2012) expõe um método para extração de edifício de imagens multiespectrais de alta resolução utilizando segmentação baseada no NDVI para eliminar as regiões naturais da cena. Em seguida, operações morfológicas são usadas para separar estradas de edifícios, com base em propriedades espaciais, pois estradas são regiões alongadas e com área maior do que edifícios e geralmente edifícios tem os telhados retangulares.

Além da relevância científica, pode-se apontar que o tema da pesquisa possui também relevância tecnológica no contexto de captura de informações espaciais para alimentar/atualizar bases de dados de cadastros urbanos em grande escala, que são fundamentais para muitas ações públicas. Neste sentido, espera-se contribuir positivamente com o desenvolvimento do método proposto.

## 1.2 Objetivos

O objetivo principal deste trabalho é estudar e avaliar o poder de discriminação de formas de contornos de edifícios usando características calculadas com base nos momentos de Zernike.

Especificamente, os objetivos são:

- Estudar diferentes métodos de segmentação por região e escolher um método adequado para segmentar contornos de edifícios;
- Estudar e implementar computacionalmente o método de momentos de Zernike;
- Criar um banco de dados de esboços de contorno de edifício.

## 1.3 Estrutura da dissertação

Esta dissertação está estruturada em cinco capítulos. O Capítulo 2 apresenta a fundamentação teórica, que inclui uma breve fundamentação do processamento digital de imagens, princípio e métodos de segmentação de imagem por regiões, polinômios ortogonais, momentos de uma imagem, em especial, os momentos de Zernike e o descritor de forma baseado nestes momentos. No capítulo 3 é apresentada a Metodologia, onde são explanados quais são os materiais e o método utilizados na pesquisa. O capítulo 4 descreve a realização e análise dos experimentos. As conclusões e recomendações, onde será apresentada a conclusão do trabalho, dificuldades encontradas e possíveis trabalhos futuros são apresentados no capítulo 5.

## 2 *FUNDAMENTAÇÃO TEÓRICA*

Neste capítulo são estudados conceitos importantes para o desenvolvimento do trabalho proposto. Fundamentos de processamento digital de imagens são essenciais para qualquer trabalho com imagem, pois propiciam uma base para este tipo de estudo. Quanto aos princípios de segmentação de imagens, é possível conhecer diferentes métodos, que proporcionaram suporte para os trabalhos práticos. Também foi feito um estudo dos momentos complexos de Zernike que serviram de base para a construção do descritor de forma utilizado para o reconhecimento de contornos de edifício.

### 2.1 Fundamentos de Processamento Digital de Imagens

Uma imagem digital contém inúmeras informações, cujo sistema visual humano não é capaz de processar totalmente. O processamento digital de imagens visa facilitar a identificação e a extração de informação contida nas imagens, e assim, realizar interpretações apropriadas. Para isso, existem técnicas para capturar, representar e transformar imagens que melhoram a qualidade visual de certos aspectos estruturais, facilitando a percepção humana e a interpretação automática por meio de máquinas (PEDRINI; SCHWARTZ, 2008).

#### 2.1.1 Estrutura e Resolução de Imagens Digitais

A imagem digital pode ser vista como uma matriz de dimensão  $x \times y$ , pois está localizada num sistema de coordenadas de  $x$  linhas e  $y$  colunas, onde a origem é definida, por convenção, como sendo o canto superior esquerdo. Cada elemento dessa matriz recebe o nome de **pixel**, que possui um atributo numérico que indica o **nível de cinza** dessa célula, conhecido por DN, do inglês “*digital number*” (CRÓSTA, 1992).

Pode ser visto na Figura 1 uma imagem bem simples com sua forma digital, representada por uma matriz numérica, onde cada pixel possui uma localização definida por linha e coluna e o nível de cinza correspondente, que neste caso varia de 0 (preto) a 255 (branco) e é proporcional à tonalidade da imagem. Por exemplo, o pixel da 1ª linha/1ª coluna da imagem a esquerda é quase preto e seu DN é 20, enquanto que o pixel da 3ª linha/1ª coluna possui um DN igual a 230, ou seja, é quase branco.

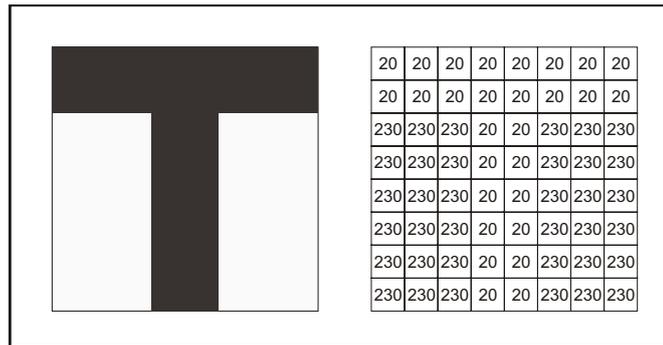


Figura 1: Exemplo de uma imagem e sua forma digital.

Fonte: Adaptado de Crósta (1992).

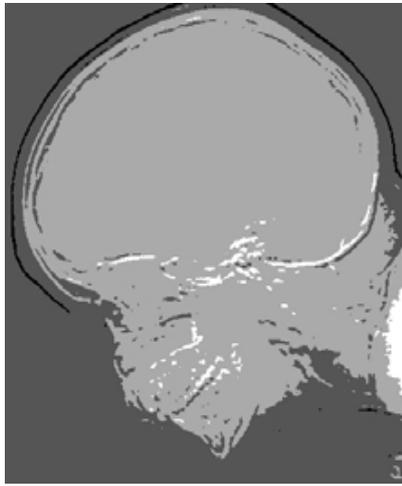
A qualidade ou resolução de uma imagem está relacionada com três componentes: resolução espacial, resolução espectral e resolução radiométrica, além de depender da forma como foi capturada.

Segundo Gonzalez e Woods (2008), basicamente, a resolução **espacial** é uma medida do menor detalhe discernível em uma imagem. Assim, quanto maior a densidade de pixels em uma imagem, maior será a resolução da imagem (PEDRINI; SCHWARTZ, 2008). Pode-se utilizar o GSD (*Ground Sample Distance*) para quantificar essa resolução, que representa o tamanho real que um pixel representa em função da resolução de uma imagem. Por exemplo, uma imagem com GSD de 0,1m apresentará detalhes visuais mais nítidos do que uma imagem com GSD de 2m, pois para representar uma mesma área serão necessários uma maior quantidade de pixels.

O conceito de resolução **espectral** é inerente às imagens multiespectrais de sensoramento remoto e é definida pelo número de bandas espectrais de um sistema sensor (que captura a imagem) e pela amplitude do intervalo de comprimento de onda coberto por cada banda. Dessa forma, quanto menor a amplitude desse intervalo e maior o número de bandas, a resolução espectral do sensor é maior. Um objeto pode ser representado em apenas uma banda espectral através de uma fotografia em preto e branco, enquanto que se a foto é colorida, o mesmo objeto é representado em três bandas espectrais (vermelha,

verde e azul) que, quando combinadas por superposição, mostram o objeto em cores.

A resolução **radiométrica**, também conhecida por resolução de intensidade é dada pela quantidade de níveis de cinza. Quanto maior o número de níveis, maior é a resolução radiométrica. Por exemplo, uma imagem com 64 níveis de cinza entre o branco e o preto terá mais detalhes perceptíveis que uma imagem com apenas 4 níveis e assim, terá melhor resolução radiométrica, como pode ser visto na Figura 2. O número de níveis é expresso em função do número de “bits” necessários para armazenar o valor do nível máximo em forma digital. O valor em bits é uma potência de 2, ou seja, uma imagem de 8 bits, por exemplo, armazena  $2^8 = 256$  níveis de cinza (0 - preto, 255 - branco).



(a) Imagem de 2 bits (apenas 4 níveis de cinza).



(b) Mesma imagem, de 6 bits (64 níveis de cinza).

Figura 2: Comparação de resolução radiométrica.

Fonte: Gonzalez e Woods (2008).

### 2.1.2 Histograma de uma Imagem

Para representar a distribuição dos níveis de cinza (ou DN's) de uma imagem, comumente utiliza-se o **histograma**, que também é conhecido como distribuição de intensidades, pois informa quantos pixels na imagem possuem o mesmo valor de DN e o faz para todos os valores possíveis. Sendo que o histograma não informa, por exemplo, a localização de um pixel com determinado DN em relação a outro pixel com outra intensidade, isto é, ele não contém informações espaciais sobre a imagem.

### 2.1.2.1 Histograma Unidimensional

Pode-se usar tabelas ou gráficos para representar o histograma unidimensional, ou de apenas uma banda, que possui o número de pixels atribuído a cada um dos possíveis níveis de cinza, como mostra a Figura 3, onde o eixo horizontal representa a escala de níveis de cinza e o eixo vertical indica a frequência de pixels em cada um desses tons de cinza.

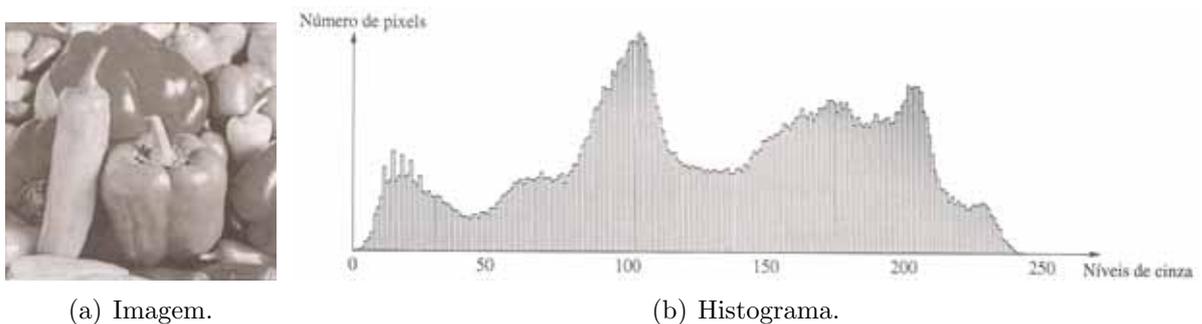


Figura 3: Distribuição de níveis de cinza de uma imagem através de seu histograma.  
Fonte: Pedrini e Schwartz (2008).

Informações como os valores mínimo e máximo, a intensidade média, a variância e o desvio padrão dos níveis de cinza podem ser obtidas a partir do histograma, além disso, observando-o também pode-se avaliar o contraste da imagem (PEDRINI; SCHWARTZ, 2008). Como pode ser observado nas Figuras 4a-c, que apresentam escalas estreitas de níveis de cinza, correspondendo a imagens de baixo contraste, enquanto que na Figura 4d os valores de cinza no histograma estão melhor distribuídos ao longo de toda a escala, logo, corresponde a uma imagem de alto contraste.

Quando uma imagem contém uma área homogênea (uma monocultura, por exemplo), com baixo contraste, seu histograma terá um único pico (Figura 5), enquanto imagens que contêm diferentes tipos de cobertura superficial (uma região costeira, por exemplo, com áreas de água e solo, Figura 6) mostrarão vários picos e podem estar claramente separados no histograma. Neste sentido, observar a forma do histograma de uma imagem cria uma espécie de aproximação sobre as características espectrais da área dessa imagem. Sendo assim, a análise e compreensão do histograma de uma imagem é fundamental para a aplicação de funções de realce ou aumento de contraste e também de segmentação de imagens, cujos métodos serão estudados adiante.

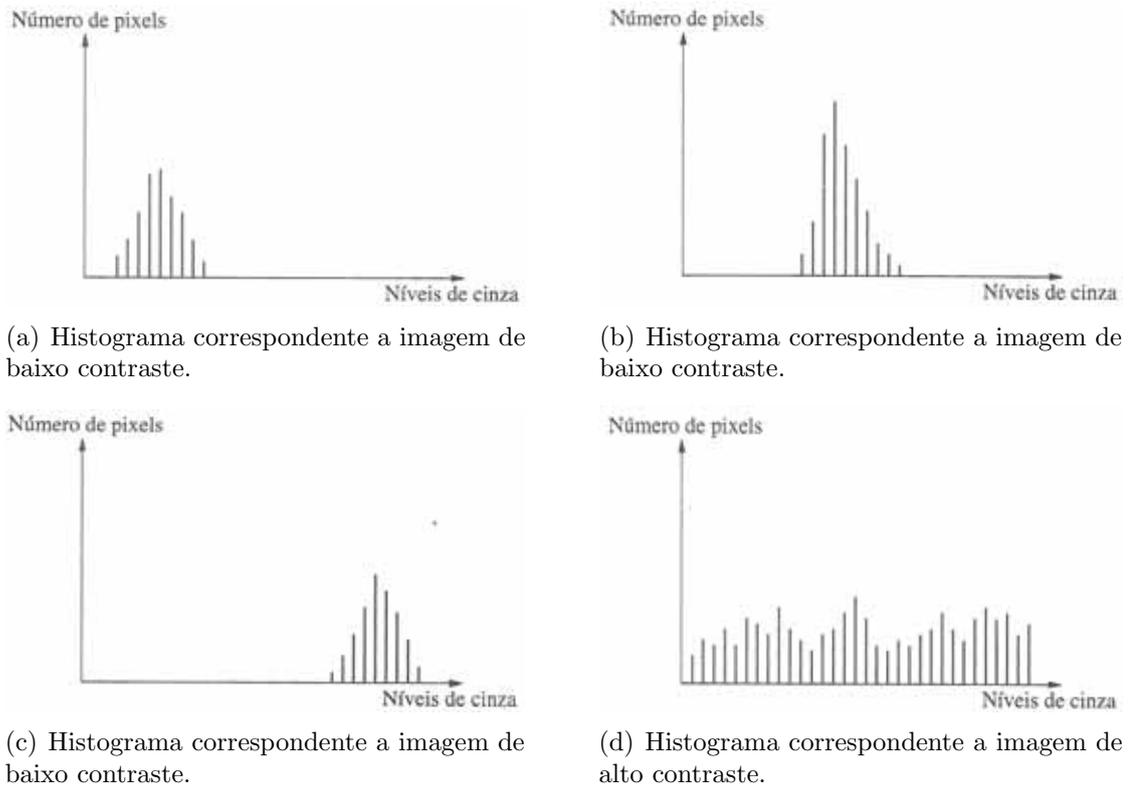
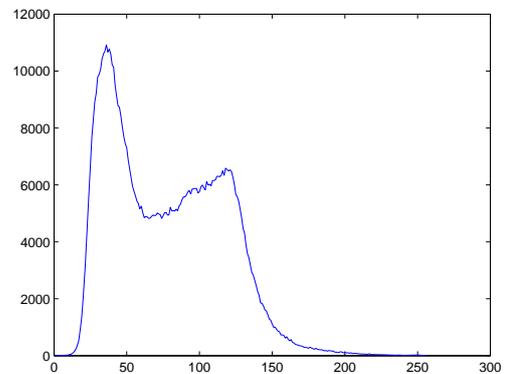


Figura 4: Comparação do contraste de imagens a partir de histogramas.

Fonte: Pedrini e Schwartz (2008).



(a)



(b)

Figura 5: Imagem de uma plantação de cana e seu histograma.

### 2.1.2.2 Histograma Cumulativo

Outra forma de representar a distribuição de intensidades de uma imagem é através do histograma cumulativo, que mostra a frequência cumulativa dos pixels com determinado valor de nível de cinza contra todos os valores de cinza, onde a frequência cumulativa, para cada valor de DN, é a soma da frequência desse DN e as frequências de todos os DNs

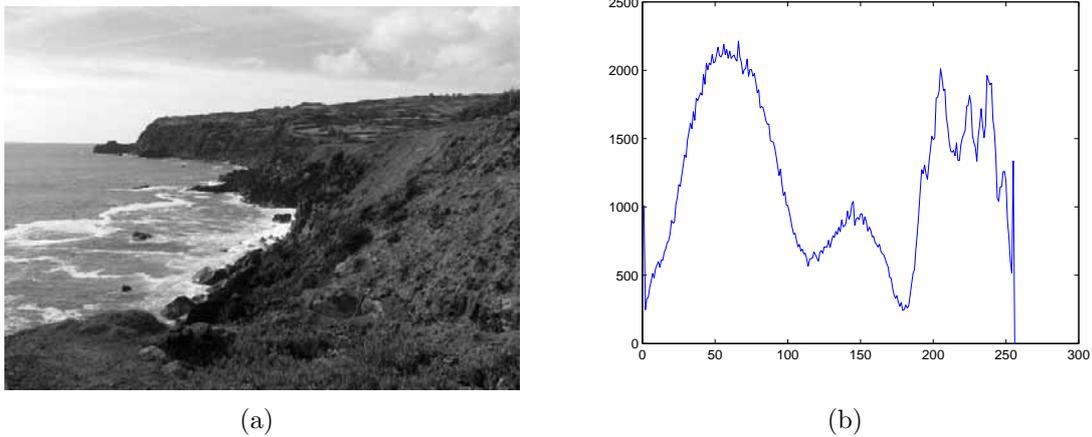
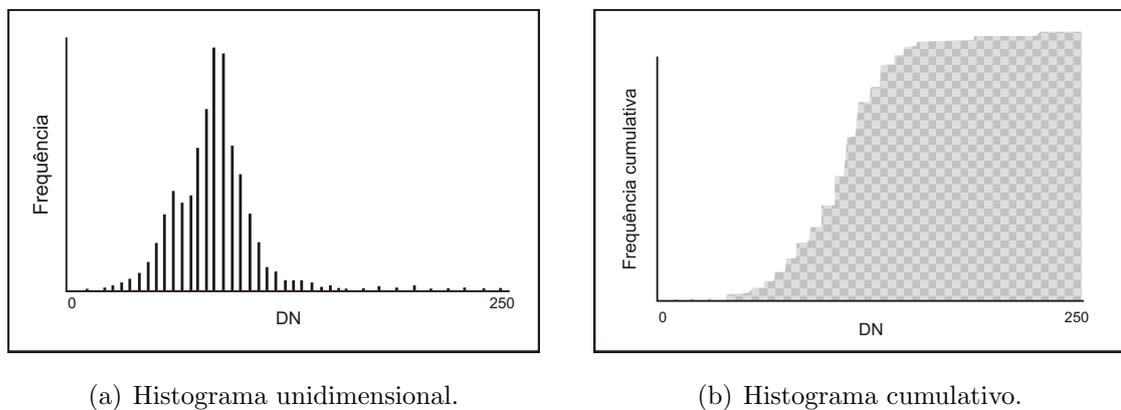


Figura 6: Imagem de uma região costeira e seu histograma.

abaixo dele. Neste caso, a inclinação da curva do gráfico é a medida do contraste, sendo que as heterogeneidades se mostram como quebras na curva.

Na Figura 7 é possível observar os dois tipos de histograma já citados.



(a) Histograma unidimensional.

(b) Histograma cumulativo.

Figura 7: Comparação entre dois tipos de histograma.

Fonte: Crósta (1992).

### 2.1.3 Filtragem Espacial

O valor dos níveis de cinza de uma imagem variam dependendo de sua localização na imagem. Em limites entre áreas com diferentes níveis de cinza, como bordas e linhas, o valor dos DN's variam bastante em pequenas áreas da imagem, assim, são chamadas de feições ou componentes de **alta frequência** e estão relacionadas com a informação de detalhes da imagem. Já as áreas em que os níveis de cinza variam mais uniformemente

são chamadas de feições ou componentes de **baixa frequência**.

Utiliza-se técnicas de filtragem espacial de frequências para que informações contidas numa imagem sejam melhoradas e o resultado desse processo seja mais apropriado, para uma especificidade, do que a imagem original. Para isso, essas técnicas atenuam ou eliminam as feições de alta, média ou baixa frequência, operando na vizinhança de cada pixel.

Os métodos de filtragem podem ser implementados no domínio espacial ou no domínio da frequência de uma imagem. Adiante são estudados os métodos no domínio espacial, que utilizam do processo de convolução: filtros **passa baixas** e **passa altas**. Os filtros **passa banda** ou **passa faixa** são capazes de remover regiões selecionadas entre alta e baixa frequência, mas é mais comum ser utilizado em filtrações no domínio das frequências e não será abordado neste trabalho.

### 2.1.3.1 Filtragem com a Janela Móvel ou Máscara

Os filtros de convolução em uma imagem trabalham com uma **máscara**, também chamada de **janela móvel**. Por exemplo, uma máscara  $M$  de dimensões  $3 \times 3$  possui valores pré-definidos nas 9 células (pesos) e é, primeiramente, colocada sobre o canto superior esquerdo da imagem de entrada  $A$ , multiplicando cada nível de cinza das 9 células de  $A$  pelo respectivo valor da célula da máscara; o próximo passo é substituir o valor do pixel central dessa área da imagem  $A$  pela soma dos 9 valores e, assim, definir o novo valor do pixel correspondente na imagem de saída  $B$  (no caso, o pixel da  $2^a$  linha/ $2^a$  coluna).

Em seguida, a janela deve ser movida um pixel ao longo da linha da imagem e realizar as mesmas operações novamente, para então determinar o valor do pixel da  $2^a$  linha/ $3^a$  coluna da imagem  $B$ . Dessa maneira, a operação deve ser feita pixel-a-pixel para a imagem toda. A Figura 8 ilustra este exemplo.

Em outras palavras, seja  $m(x, y)$  uma máscara com dimensões  $(2i + 1) \times (2j + 1)$ , onde  $i$  e  $j$  são inteiros positivos, a expressão matemática da convolução é dada por (GONZALEZ; WOODS, 2008):

$$g(x, y) = \sum_{s=-i}^i \sum_{t=-j}^j p(i + s, j + t) m(s, t) \quad (2.1)$$

onde os valores de  $p$  são os pixels da imagem centrados no pixel  $(x, y)$  e  $g(x, y)$  são os pixels da imagem resultante.

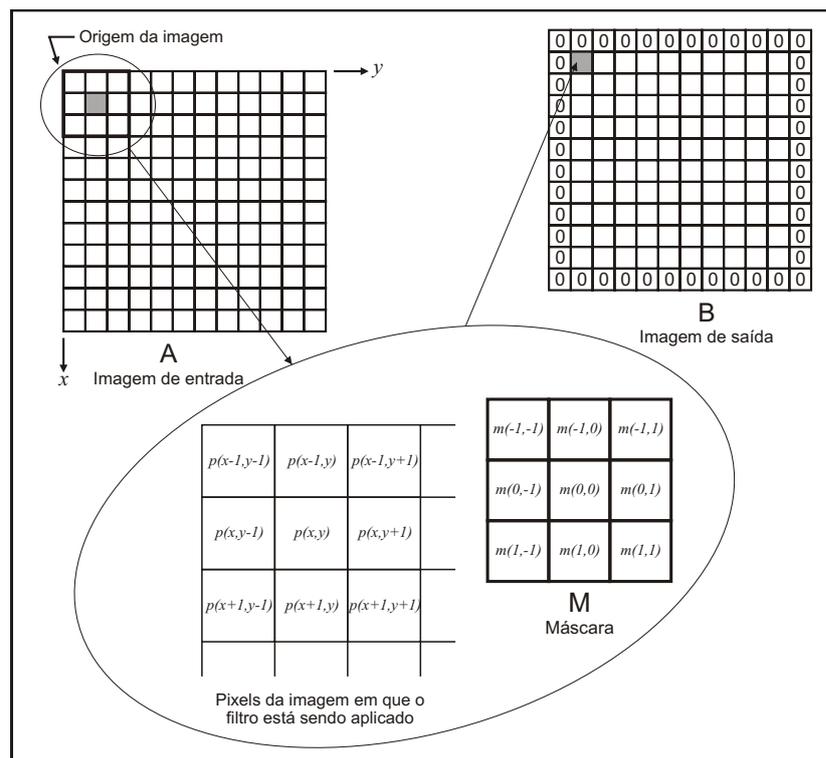


Figura 8: Esquema da máscara para filtragem de uma imagem.  
Fonte: Gonzalez e Woods (2008).

Observe na Figura 8 que a imagem de saída produzida pela filtragem possui os valores dos pixels de suas bordas iguais a zero, devido ao *efeito de bordas*, que acontece porque a primeira posição em que é possível aplicar uma máscara de dimensão  $3 \times 3$  é no pixel da 2ª linha/2ª coluna, ao invés do pixel da 1ª linha/1ª coluna. Isto ocorre em todos os pixels da borda, que tem então seus valores substituídos por zero. Mas note que, se a máscara possui  $5 \times 5$  elementos, então as bordas terão 2 linhas e 2 colunas.

### 2.1.3.2 Filtros Passa Baixas ou de Suavização

Os filtros do tipo **passa baixas** ou de **suavização** recebem este nome porque deixam “passar” apenas as baixas frequências da imagem, eliminando as altas. Esses filtros são eficientes para remoção de ruídos e excesso de detalhes que não sejam importantes na imagem. No entanto, por perderem essas informações de detalhe, “desfocam” a imagem, sendo que este efeito aumenta proporcionalmente à dimensão da máscara usada, isto é, uma máscara  $5 \times 5$  produzirá uma imagem mais desfocada do que uma máscara  $3 \times 3$  equivalente. Nessa família de filtros, a soma dos pesos utilizados na máscara é sempre igual a 1. A seguir, são apresentados alguns tipos comuns desses filtros.

- **Filtro da Média**

O filtro de média é um dos tipos mais simples e ele opera substituindo o valor do pixel original pela média aritmética deste pixel e seus vizinhos. Este filtro tem a desvantagem de afetar áreas sem ruídos, muitas vezes homogeneizando a imagem onde não se deseja. Por exemplo, uma máscara  $3 \times 3$  é da seguinte forma (CRÓSTA, 1992):

$1/9$	$1/9$	$1/9$
$1/9$	$1/9$	$1/9$
$1/9$	$1/9$	$1/9$

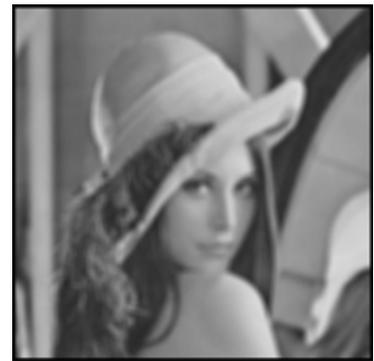
Na Figura 9 tem-se um exemplo da filtragem de uma imagem, utilizando filtro de média com máscaras  $3 \times 3$  e  $11 \times 11$ , onde é possível perceber o efeito de borda e como a imagem ficou borrada, principalmente no caso da máscara maior.



(a) Imagem original.



(b) Imagem filtrada com máscara  $3 \times 3$ .



(c) Imagem filtrada com máscara  $11 \times 11$ .

Figura 9: Resultados de filtros da média.

- **Filtro da Média Ponderada**

Esse tipo de filtro de suavização é semelhante ao anterior, porém, a média utilizada é ponderada, onde os pesos da máscara são definidos em função de sua distância do peso central, assim, o filtro de média ponderada causa um efeito de suavização menor do que no filtro de média, visto que o peso do pixel central é maior. Uma máscara  $3 \times 3$  deste tipo de filtro apresenta os seguintes pesos (CRÓSTA, 1992):

$1/16$	$1/8$	$1/16$
$1/8$	$1/4$	$1/8$
$1/16$	$1/8$	$1/16$

- **Filtro Gaussiano**

De acordo com Pedrini e Schwartz (2008), os coeficientes da máscara dos filtros Gaussianos são derivados de uma função Gaussiana bidimensional, sendo que a função Gaussiana discreta com média um e desvio padrão  $\sigma$  é definida abaixo e um gráfico dessa função pode ser visto na Figura 10.

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp \frac{-(x^2 + y^2)}{2\sigma^2} \quad (2.2)$$

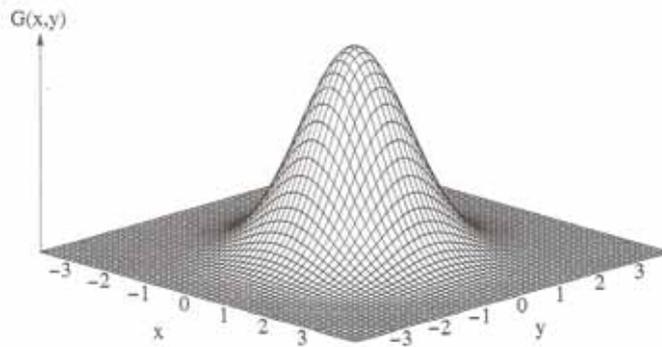


Figura 10: Função Gaussiana bidimensional com média 1, centrada em (0,0) e  $\sigma = 1$ .

Fonte: Pedrini e Schwartz (2008).

Algumas características importantes dos filtros Gaussianos são que o grau de suavização realizado pelo filtro é o mesmo em todas as direções (isotrópico), sendo que quanto maior o valor de  $\sigma$ , maior será este grau de suavização. Além disso, a suavização da imagem é realizada por meio da substituição de cada pixel por uma média ponderada dos pixels vizinhos, de tal maneira que o peso dado a um vizinho decresce com a distância do pixel central.

Os filtros descritos a seguir são não-lineares, então não existe uma formulação matemática linear que descreva o seu processo de aplicação. Eles baseiam-se diretamente nos valores dos pixels da vizinhança considerada.

- **Filtro da Moda**

Neste tipo de filtro, o pixel central da área da imagem trabalhada é substituído pelo seu vizinho mais comum, ou seja, por aquele valor de DN que mais aparece dentro da dimensão da máscara pré-definida. O filtro de moda é útil na “limpeza” de pixels isolados nas imagens.

- **Filtro da Mediana**

Neste caso, o pixel central é substituído pela mediana dos seus vizinhos, isto é, pelo valor intermediário do conjunto formado pelos DN's dos vizinhos do pixel central, sendo que a quantidade de elementos desse conjunto depende da dimensão da máscara. Por exemplo, a mediana do conjunto de 9 elementos (máscara  $3 \times 3$ ) será o quinto elemento do conjunto que deve estar ordenado em ordem crescente. Assim, a implementação desse filtro é um pouco complexa, devido a essa ordenação, mas é um dos filtros passa baixas que melhor preserva as bordas da imagem, não “borrando-a” tanto.

Considerando os valores de nível de cinza de uma imagem dados na Figura 11, tem-se que o pixel central da área sombreada será substituído pelos seguintes valores em cada caso de filtro: 26 no filtro da média, 27 no filtro da média ponderada e no filtro da mediana e 28 no filtro da moda.

25	27	32	20	21	19	24
23	27	19	25	22	18	25
27	26	24	26	22	22	26
23	25	28	32	27	25	27
19	22	28	19	29	27	25
17	19	25	27	30	31	22
20	21	25	23	21	19	17

Figura 11: Exemplo da representação de uma imagem para aplicação de filtros passa baixas.

Na Figura 12 tem-se um exemplo da filtragem de uma imagem utilizando filtro da média e da mediana, observa-se que o filtro da mediana conservou melhor as bordas da imagem, obtendo melhores resultados.



(a) Imagem original.



(b) Imagem filtrada com filtro da média.



(c) Imagem filtrada com filtro da mediana.

Figura 12: Resultados de filtros da média e mediana.

### 2.1.3.3 Filtros Passa Altas ou de Realce de Bordas

Os filtros passa altas ou de realce de bordas fazem o inverso dos filtros de suavização, pois eliminam as feições de baixa frequência e deixam “passar” somente as de alta frequência, que geralmente expressam bordas ou limites entre áreas de diferentes valores de intensidade, aumentando a nitidez dessas regiões.

Neste tipo de filtro, a dimensão da máscara também influencia no resultado da imagem de saída, no entanto, quanto menor a máscara, os detalhes a serem realçados na imagem serão menores.

#### • Filtros Laplacianos

Geralmente, a soma dos pesos utilizados na máscara é igual a zero e esses filtros são úteis na detecção de bordas, pois por serem regiões de mudanças abruptas em uma imagem, as bordas, são realçadas com a aplicação do Laplaciano, enquanto que os níveis de cinza em regiões homogêneas são atenuadas. Existem vários tipos de filtros Laplacianos, estes são alguns deles (CRÓSTA, 1992; GONZALEZ; WOODS, 2008):

0	1	0	1	1	1	1	-2	1	0	-1	0	-1	-1	-1
1	-4	1	1	-8	1	-2	4	-2	-1	4	-1	-1	8	-1
0	1	0	1	1	1	1	-2	1	0	-1	0	-1	-1	-1

A Figura 13 apresenta um exemplo da aplicação da primeira máscara do Laplaciano apresentada acima.



(a) Imagem original.



(b) Imagem com aplicação do Laplaciano.

Figura 13: Resultado do filtro Laplaciano.

- **Filtros Laplacianos + Original**

Imagens resultantes da filtragem passa altas geralmente são pobres em contraste. Adicionar a imagem filtrada à imagem original é uma maneira de manter o contraste original e obter também o realce de bordas. Para isso, basta adicionar a unidade ao peso central da máscara, desse modo, alguns dos exemplos acima ficariam da seguinte forma (CRÓSTA, 1992):

1	-2	1
-2	5	-2
1	-2	1

0	-1	0
-1	5	-1
0	-1	0

-1	-1	-1
-1	9	-1
-1	-1	-1

Para melhores resultados, ao invés da unidade, pode-se dar um valor muito maior ao peso central da máscara, dependendo do que se deseja.

## 2.1.4 A Cor no Processamento Digital de Imagens

A utilização de cor na análise de imagens simplifica a identificação de objetos e a interpretação de uma cena. Além disso, o ser humano consegue reconhecer milhares de tons e intensidades de cores, em comparação com apenas algumas dezenas de níveis de cinza (GONZALEZ; WOODS, 2008; PEDRINI; SCHWARTZ, 2008).

Em 1666, Isaac Newton descobriu que quando um feixe de luz solar atravessa um prisma de vidro, o feixe de luz emergente consiste em um espectro contínuo de cores variando do violeta ao vermelho (PEDRINI; SCHWARTZ, 2008). Esse intervalo de luz visível corresponde a uma pequena região do espectro eletromagnético cujo comprimento de onda é capaz de sensibilizar o olho humano, conforme ilustra a Figura 14.

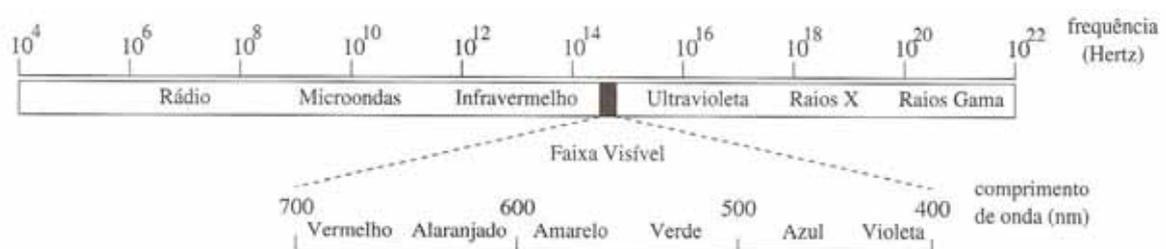


Figura 14: Espectro eletromagnético.  
Fonte: Pedrini e Schwartz (2008).

Segundo Gonzalez e Woods (2008), um corpo que reflete luz de forma balanceada em todos os comprimentos de onda visíveis é percebido como branco pelo observador,

enquanto que um corpo que favorece a refletância em uma faixa limitada do espectro visível exibe alguns tons de cores, como por exemplo, os objetos verdes que refletem luz com comprimentos de onda primariamente no intervalo de 500 a 570 nm e absorvem a maior parte da energia de outros comprimentos de onda.

#### 2.1.4.1 O Espaço RGB

De acordo com Pedrini e Schwartz (2008) foi o trabalho de Thomas Young (1773-1829) que propôs que as cores são percebidas pelo olho humano como combinações das chamadas cores *primárias*: vermelho, verde e azul (**RGB**), sendo que a mistura em quantidades iguais dessas três cores gera a cor branca. Essas cores primárias podem ser adicionadas duas a duas para produzir as cores *secundárias*: ciano, magenta e amarelo (**CMY**), que por sua vez, podem ser subtraídos da cor branca e obter os valores RGB, além disso, a adição das cores secundárias produz o preto (GONZALEZ; WOODS, 2008). Então, através de qualquer um desses conjuntos de cores, é possível obter várias tonalidades de cores. Sendo que, por exemplo, os monitores de televisão usam o modelo RGB e as impressoras coloridas utilizam o modelo CMY.

A Figura 15 mostra o cubo RGB, onde cada eixo representa uma cor RGB e sua intensidade aumenta com a distância da origem. Assim, uma cor **C** pode ser vista como um ponto nesse sistema de coordenadas, dado pelos coeficientes que definem o quanto de cada cor é necessária para produzir a cor desejada, em que  $C = r.R + g.G + b.B$ , onde  $R, G, B$  são as três cores primárias e  $r, g, b$  são os coeficientes, que geralmente são limitados ao intervalo  $0 \leq r, g, b \leq 1$ .

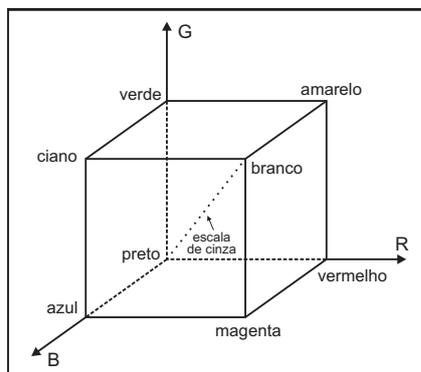


Figura 15: Esquema do cubo de cores RGB.

Fonte: Crósta (1992).

Note que a cor preta é representada pela origem e a diagonal pontilhada tem as tonalidades de cinza, onde os valores  $r, g, b$  são iguais, chamado de eixo acromático.

### 2.1.4.2 O Modelo HSI

O modelo HSI é outra forma de representação de cores, que são definidas por atributos: Matiz (*Hue-H*), Saturação (*Saturation-S*) e Intensidade (*Intensity-I*), sendo que esses atributos podem ser analisados individualmente, diferente do modelo RGB, onde as componentes são interligadas.

*Matiz* de um objeto é o comprimento de onda dominante em uma mistura de ondas de luz que ele reflete ou emite, assim, esse atributo representa a cor desse objeto. A *saturação* refere-se à pureza de uma cor, ou seja, à quantidade de luz branca misturada com um matiz, por exemplo, cores puras são completamente saturadas, enquanto que cor-de-rosa (vermelho e branco) e lilás (violeta e branco) são menos saturadas. Quanto à *intensidade*, é o atributo responsável pela sensação de brilho de uma cor sobre o olho humano.

Desse modo, o modelo de cores HSI é útil para algoritmos de processamento de imagens baseados em algumas propriedades do sistema visual humano, como sistemas para inspeção de qualidade de produtos coloridos.

A Figura 16 mostra a relação entre os espaços RGB e HSI, que pode ser representado graficamente por um cone. Assim, o modelo HSI usa coordenadas cilíndricas polares para representar as cores, enquanto que o sistema RGB utiliza coordenadas cartesianas. A partir dessas coordenadas, é possível converter imagens no formato RGB para o modelo HSI, ou vice-versa, através de transformações matemáticas.

A distância de um ponto do cone até a origem, que é a mesma para os dois sistemas e retrata a cor preta, representa a intensidade ou brilho da cor; já a distância radial do ponto até o eixo central do cone, que coincide com o eixo acromático, representada a saturação. O valor do matiz é um ângulo que varia de 0 a 360° e é representado como uma sequência radial ao redor do eixo de intensidades e dos círculos de saturação.

Para transformar cores no formato RGB para o modelo HSI, assumindo que os componentes R, G e B estão normalizados no intervalo entre 0 e 1, utiliza-se (GONZALEZ; WOODS, 2008):

$$H = \begin{cases} \theta, & \text{se } B \leq G \\ 360 - \theta, & \text{se } B > G \end{cases} \quad (2.3)$$

$$S = 1 - \frac{3}{R + G + B} \min\{R, G, B\} \quad (2.4)$$

$$I = \frac{R + G + B}{3} \quad (2.5)$$

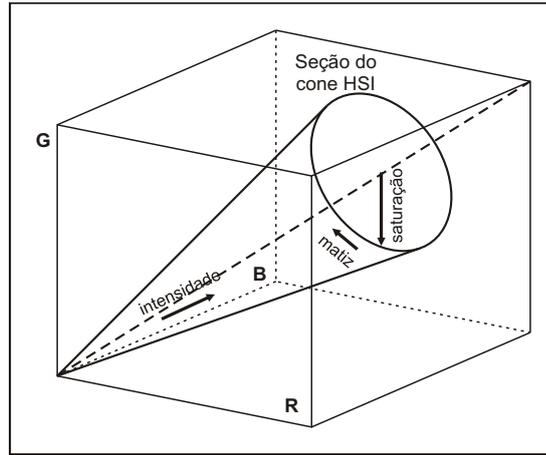


Figura 16: Esquema da relação entre os espaços RGB e HSI.  
Fonte: Crósta (1992).

onde

$$\theta = \arccos \left( \frac{(R - G) + (R - B)}{2\sqrt{(R - G)^2 + (R - B)(G - B)}} \right) \quad (2.6)$$

Existem três casos na conversão de cores no modelo HSI, cujos valores estão no intervalo  $[0, 1]$ , para RGB (GONZALEZ; WOODS, 2008).

Se  $0^\circ \leq H < 120^\circ$ , então:

$$B = I(1 - S) \quad (2.7)$$

$$R = I \left[ 1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (2.8)$$

$$G = 3I - (R + B) \quad (2.9)$$

Se  $120^\circ \leq H < 240^\circ$ , fazemos  $H = H - 120^\circ$  e então:

$$R = I(1 - S) \quad (2.10)$$

$$G = I \left[ 1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (2.11)$$

$$B = 3I - (R + G) \quad (2.12)$$

Se  $240^\circ \leq H \leq 360^\circ$ , fazemos  $H = H - 240^\circ$  e então:

$$G = I(1 - S) \quad (2.13)$$

$$B = I \left[ 1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (2.14)$$

$$R = 3I - (B + G) \quad (2.15)$$

## 2.2 Segmentação de Imagens

### 2.2.1 Regiões *versus* Bordas

Um grupo de pixels conectados com características semelhantes forma uma região numa imagem, que podem corresponder a objetos, que por sua vez, podem conter diferentes regiões correspondentes às suas partes.

Para interpretar precisamente uma imagem é preciso que ela seja particionada em regiões que representam seus diferentes objetos ou partes de objetos. Encontram-se essas partições a partir dos níveis de cinza da imagem e a este processo dá-se o nome de **segmentação**.

Segundo Jain, Kasturi e Schunck (1995), existem duas abordagens para segmentar uma imagem em regiões: agregação de pixels que compõem as regiões e agregação de pixels que compõem os contornos das regiões.

Na **agregação de pixels que compõem as regiões**, como o próprio nome já diz, uma região é composta pelo agrupamento de todos os pixels correspondentes a um objeto, sendo que algum critério é estabelecido para distinguir estes pixels do resto da imagem. Dois pixels podem ser agrupados se possuem valores de intensidade parecidos e se estão próximos um do outro, assim, dois princípios são fundamentais: a *similaridade de valor* e a *proximidade espacial*, que geralmente são baseadas na diferença entre os níveis de cinza e na distância Euclidiana, respectivamente.

De acordo com Pedrini e Schwartz (2008), outras propriedades também podem ser consideradas, como cor, informação semântica ou textura.

A outra abordagem, de **agregação de pixels que compõem os contornos das regiões**, procura os pixels que estão sobre os limites das regiões, chamados de bordas. Como esses pixels estão no contorno e as regiões ficam do outro lado, eles podem ter níveis de cinza diferentes, assim, muitos detectores de bordas medem a diferença de intensidade entre pixels vizinhos. Dessa forma, uma região é delimitada por suas bordas.

A princípio, as duas categorias de segmentação deveriam produzir os mesmos resultados, pois bordas podem ser encontradas a partir de regiões através de algoritmos de perseguição de contorno e as regiões podem ser obtidas de bordas usando algoritmos de preenchimento de regiões. Infelizmente, é raro obter resultados corretos. Além disso, devido a ruídos e outros fatores, as duas abordagens não produzem informações perfeitas (JAIN; KASTURI; SCHUNCK, 1995).

## 2.2.2 Segmentação de Imagem por Regiões

Seja  $I$  uma imagem, o problema de segmentação pode ser visto como uma partição da imagem  $I$  em  $n$  regiões  $R_1, R_2, \dots, R_n$ , tal que (JAIN; KASTURI; SCHUNCK, 1995; PEDRINI; SCHWARTZ, 2008):

- (a)  $\bigcup_{i=1}^n R_i = I$ ;
- (b)  $R_i \cap R_j = \emptyset$  para todo  $i$  e  $j$ ,  $i \neq j$ ;
- (c)  $P(R_i) = VERDADEIRO$  para  $i = 1, 2, \dots, n$ ;
- (d)  $P(R_i \cup R_j) = FALSO$  para  $i \neq j$  e  $R_i$  adjacente a  $R_j$ .

onde  $P(R_i)$  é um atributo de homogeneidade para os pontos do conjunto  $R_i$ , ou seja, é uma sentença que se refere a determinadas propriedades que servem para agrupar os pixels da imagem em uma região, tais como intensidade média, variância, textura ou cor.

A primeira condição quer dizer que cada pixel deve pertencer a uma região, desse modo, a união de todas as regiões é a própria imagem. A condição (b) indica que as regiões são distintas, logo, a interseção de duas regiões deve ser o conjunto vazio. A condição (c) estabelece que o pixel deve satisfazer à sentença  $P(R_i)$  para pertencer à região  $R_i$  e a condição (d) indica que a fusão de duas regiões adjacentes não satisfaz a mesma sentença  $P$ , isto é, elas não podem ser unidas em uma única região.

Uma forma simples de segmentação é dividir a imagem em dois conjuntos, convertendo-a em uma imagem binária, onde um dos conjuntos representa os objetos e o outro é reservado para o fundo. No entanto, este conceito pode ser generalizado para segmentar imagens com mais de dois níveis, através de algoritmos de limiarização.

Para que o processo de segmentação seja mais robusto, ao invés de fixar um valor para o limiar, o algoritmo deveria selecionar automaticamente um limiar apropriado, sem intervenção humana, levando em consideração conhecimentos sobre o ambiente e os objetos contidos na cena, tais como as características da intensidade dos objetos, o tamanho dos objetos, os fragmentos da imagem ocupados pelos objetos e o número de diferentes tipos de objetos aparecendo na imagem (JAIN; KASTURI; SCHUNCK, 1995).

Algoritmos de limiarização automática analisam a distribuição dos níveis de cinza, geralmente utilizando o histograma da imagem e, dependendo da aplicação, selecionam o limiar mais apropriado.

Desse modo, a seguir, serão discutidos alguns métodos de segmentação de imagens por regiões.

### 2.2.3 Métodos Pontuais (pixel-a-pixel)

De acordo com Jähne (2002), estes métodos são conceitualmente os mais simples. Eles dependem apenas de informações tomadas sobre pixels individuais. O problema básico consiste em determinar um ou mais limiares para classificar os pixels da imagem em regiões homogêneas.

Segue abaixo vários métodos para a determinação automática do limiar a partir de um histograma da imagem de entrada apresentados por Jain, Kasturi e Schunck (1995), considerando a convenção de que objetos são escuros e o fundo é claro, assim, os pixels com níveis de cinza abaixo de um certo limiar pertencem aos objetos enquanto pixels com DN's acima do limiar são do fundo da imagem. Porém, esses algoritmos podem ser modificados para outros casos, como por exemplo, objetos claros e fundo escuro ou objetos com nível de cinza médio e fundo claro.

#### 2.2.3.1 Método do Limiar Percentual

O método do limiar percentual depende do conhecimento sobre a área ocupada ou do tamanho de um objeto particular na imagem.

Suponha que o objeto ocupa cerca de  $p\%$  da área da imagem, então um ou mais limiares  $T$  podem ser tomados de maneira que  $p\%$  dos pixels sejam atribuídos ao objeto, conforme mostra o exemplo da Figura 17, onde a área sombreada no histograma corresponde à  $p\%$  da área da imagem. Assim,  $T$  é selecionado tal que  $p\%$  dos pixels tenham intensidade abaixo de  $T$  e sejam classificados como objeto.

O uso desta técnica é limitado e de difícil aplicação prática, pois poucos casos permitem uma estimativa da área.

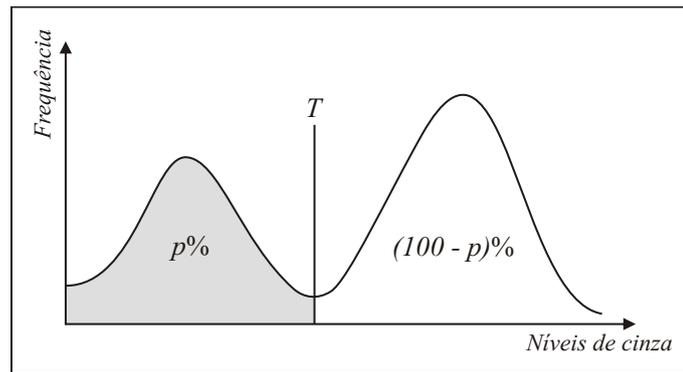


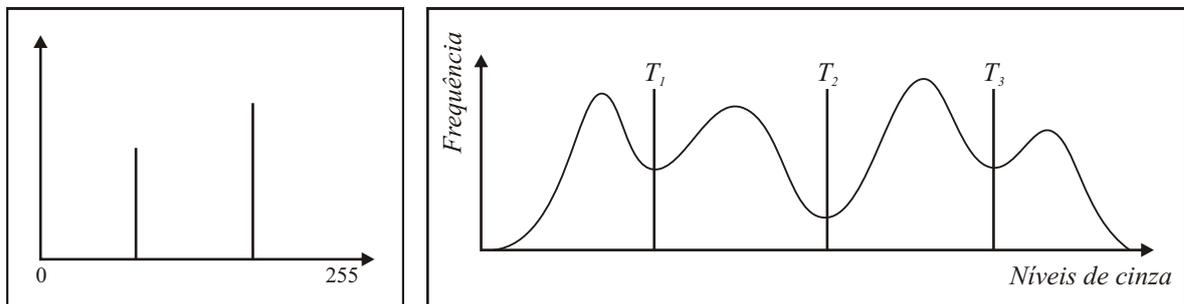
Figura 17: Exemplo do uso do método do limiar percentual.

Fonte: Adaptado de Jain, Kasturi e Schunck (1995).

### 2.2.3.2 Método da Moda

No método da moda, diferentes objetos numa imagem possuem níveis de cinza distintos, inclusive em relação ao fundo. Assim, o histograma correspondente da imagem terá forma multimodal.

Num caso ideal, suponha que uma imagem tenha apenas um objeto, cujos pixels tenham o mesmo nível de cinza e o fundo tenha outro nível de cinza, o histograma dessa imagem será duas linhas verticais, como mostra a Figura 18a, e o limiar pode ser tomado em qualquer lugar entre os picos. Em geral, os picos não são tão bem separados, então o histograma também possuirá vales entre as modas, que podem ser tomados como os limiares para a segmentação dos objetos, conforme se observa na Figura 18b.



(a) Caso ideal.

(b) Caso geral.

Figura 18: Histogramas de imagens que possuem objetos e fundo com diferentes valores de intensidade.

Fonte: Jain, Kasturi e Schunck (1995).

Um maneira de determinar os picos e vales em um histograma, é encontrando dois máximos locais que estejam a alguma distância mínima, assim, o menor ponto entre estes máximos será o vale que é considerado um bom limiar. Este processo pode ser generalizado para imagens com vários objetos. Por exemplo, se a imagem possui  $n$  objetos e o fundo com diferentes níveis de cinza, seu histograma terá  $n + 1$  picos e as posições dos vales  $T_1, T_2, \dots, T_n$  podem ser determinadas e tomadas como limiares para separar os objetos e o fundo.

### 2.2.3.3 Método da Limiarização Iterativa

Este método se inicia com uma estimativa inicial do limiar (por exemplo, o nível de cinza médio da imagem), para depois ser refinado progressivamente de modo que a imagem seja melhor segmentada.

Este processo de refinamento é realizado da seguinte maneira. A partir da estimativa inicial do limiar  $T$ , a imagem é dividida em duas regiões  $R_1$  e  $R_2$ . Daí, é preciso calcular a média dos níveis de cinza  $\mu_1$  e  $\mu_2$  das duas partições, então o novo limiar é definido por

$$T = \frac{1}{2}(\mu_1 + \mu_2). \quad (2.16)$$

Com o novo limiar, os cálculos devem ser repetidos até que os valores das médias  $\mu_1$  e  $\mu_2$  não mudem em sucessivas iterações.

### 2.2.3.4 Método da Limiarização Adaptativa

Devido às sombras ou à direção da iluminação, imagens podem ter iluminação desigual e o mesmo limiar não é adequado para ser usado na imagem inteira, logo, os métodos anteriores não são apropriados para este caso.

Visando minimizar este problema, uma alternativa é subdividir a imagem em várias janelas e analisar o histograma de cada uma, obtendo um limiar para cada janela.

Em outras palavras, divide-se a imagem em  $m \times m$  subimagens e um limiar  $T_{ij}$  é encontrado para cada partição baseando-se no histograma da  $ij$ -ésima subimagem ( $1 \leq i, j \leq m$ ), conforme mostra o exemplo da Figura 19. Por fim, a segmentação da imagem é dada pela união das regiões dessas subimagens.

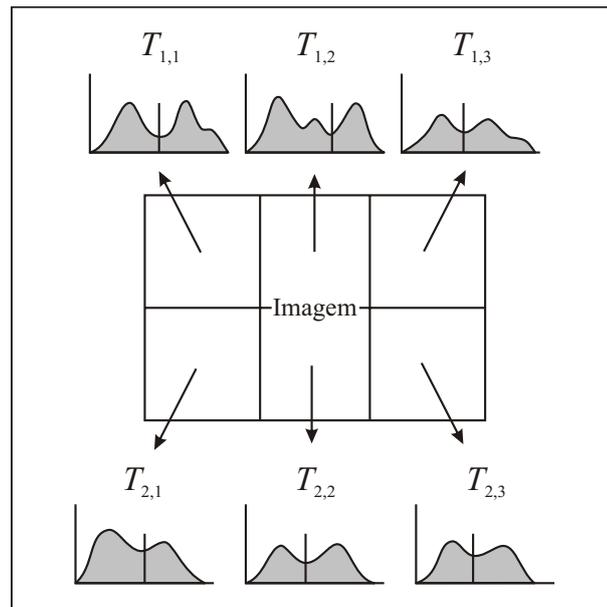


Figura 19: Esquema do processo de limiarização adaptativa.  
 Fonte: Jain, Kasturi e Schunck (1995).

### 2.2.3.5 Algumas Limitações

O histograma de uma imagem descreve a distribuição global de valores de intensidades, por exemplo, não é possível distinguir entre uma distribuição aleatória de pontos preto e branco e uma imagem que é metade branco e metade preto somente através de seus histogramas. Assim, uma limitação básica destes métodos que se baseiam no histograma para selecionar um ou mais limiares vem do fato de ignorarem o princípio de proximidade espacial, pois pontos de um mesmo objeto geralmente são espacialmente próximos. Logo, o fato do histograma não levar em consideração informações espaciais sobre os níveis de cinza na imagem, limita a aplicabilidade dos métodos citados acima em cenas complexas.

### 2.2.4 Métodos de Análise de Vizinhança

A segmentação baseada somente na exploração pontual da imagem, característica básica dos métodos descritos acima, pode resultar em muitas regiões extras, fragmentando regiões correspondentes a um mesmo objeto presente na imagem, até mesmo em casos em que os objetos aparecem claramente homogêneos. Outra consequência da exploração pontual da imagem pode ser o isolamento de regiões que correspondem a mais de um objeto.

Os métodos descritos abaixo, segundo Gonzalez e Woods (2008) e Jain, Kasturi e

Schunck (1995), visam solucionar esse problema e exploram também as informações espaciais sobre os pixels da imagem.

#### 2.2.4.1 Divisão e Fusão

O processo de  **fusão**  combina regiões que são consideradas semelhantes, sendo que a operação mais importante em um algoritmo desse tipo é como determinar a similaridade entre duas regiões. Por exemplo, considere três regiões adjacentes  $A$ ,  $B$  e  $C$  em uma imagem, o atributo de similaridade pode determinar separadamente que  $A$  e  $B$  são similares e que  $B$  e  $C$  são similares, por consequência, as três regiões serão unidas e se tornarão apenas uma, mesmo que  $A$  e  $C$  não sejam semelhantes.

Abordagens para decidir a similaridade de regiões adjacentes podem se basear: nos valores de níveis de cinza, onde se as médias das intensidades das regiões não se diferem por um valor pré-determinado, então elas são consideradas similares; na fragilidade das bordas entre regiões ou nas características estatísticas das regiões adjacentes.

Deve haver a  **divisão**  de uma região se alguma propriedade nessa região não é constante. O problema está em decidir quando isso acontece e como dividir uma região de tal maneira que a propriedade para cada componente resultante seja constante. Uma das formas é verificar a variância da intensidade dos níveis de cinza na região. Outra dificuldade é decidir onde dividir a região, em que pode-se considerar as medidas da robustez das bordas no interior da região ou então divide-se a região em um número fixo de regiões com o mesmo tamanho. Essas operações podem ser usadas juntas, de maneira que uma segmentação realizada por um método pontual é refinada por uma sucessão de divisões e fusões. Segue abaixo um esquema para realizar esse processo (JAIN; KASTURI; SCHUNCK, 1995).

1. Pré-segmentar a imagem usando um método pontual.
2. Considere uma região  $R$  qualquer. Se  $P(R) = FALSO$ , então a região deve ser dividida em quatro sub-regiões.
3. Tome duas ou mais regiões vizinhas  $R_1, R_2, \dots, R_n$ . Caso  $P(R_1 \cup R_2 \cup \dots \cup R_n) = VERDADEIRO$ , mesclar as  $n$  regiões numa única.
4. Repetir essas etapas até que não haja mais divisões e fusões.

### 2.2.4.2 Crescimento de Regiões

Este método é normalmente mais eficiente para aplicação em casos envolvendo ligeiras variações do nível de cinza nas regiões.

Como o próprio nome indica, a segmentação baseada em crescimento de regiões é um procedimento que agrupa pixels em regiões com base em critérios predefinidos para o crescimento (GONZALEZ; WOODS, 2008). O algoritmo parte de regiões sementes que crescem para formar regiões maiores que satisfaçam algum atributo de homogeneidade, sendo estas as principais dificuldades no crescimento de regiões.

Segundo Pedrini e Schwartz (2008), os pixels sementes podem ser escolhidos de maneira aleatória, determinística ou selecionados pelo usuário. Além disso, a escolha de um conjunto de um ou mais pontos de partida pode ser baseada na natureza do problema (GONZALEZ; WOODS, 2008).

Uma maneira de definir o atributo de homogeneidade é calculando a diferença absoluta entre as intensidades da vizinhança de um pixel e verificando se é menor que um dado limiar.

Levando em consideração a hipótese de que as regiões de uma imagem podem ser modeladas por funções simples, primeiramente, a imagem é dividida em algumas regiões sementes e funções são ajustadas a cada uma delas. Pixels serão agregados a essas regiões se forem compatíveis com as respectivas funções ajustadas. Assim, outra sentença de vizinhança poderia ser definido como ‘um dado pixel está próximo o suficiente da função que modela uma dada região?’.

### 2.2.5 Métodos de Perseguição do Contorno

De acordo com Jähne (2002), diferente dos métodos citados acima, a segmentação baseada nas bordas dos objetos não pode analisar vários pixels paralelamente, pois o passo seguinte depende dos resultados dos passos anteriores. Este método também leva em conta que objetos são formados por pixels adjacentes.

Após a utilização de um algoritmo para detecção de bordas, a imagem é varrida linha a linha até encontrar um pixel de borda. Quando este pixel é encontrado, um algoritmo de perseguição de contorno deve seguir este pixel a fim de traçar o contorno da região de um objeto, permitindo retornar a este mesmo pixel. Daí, uma nova pesquisa deve ser feita e o processo realizado novamente.

## 2.3 Polinômios Ortogonais em Várias Variáveis

As seguintes definições são baseadas em Dunkl e Xu (2001).

Seja  $\mathbb{N}_0$  o conjunto dos números inteiros não-negativos. Definimos um multi-índice  $\alpha$  por  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ , em que  $\mathbb{N}_0^d$  é o conjunto das  $d$ -uplas cujas componentes pertencem a  $\mathbb{N}_0$ .

Tomando  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$  e  $\alpha \in \mathbb{N}_0^d$ , podemos definir um monômio  $\mathbf{x}^\alpha$  por

$$\mathbf{x}^\alpha = x_1^{\alpha_1} \dots x_d^{\alpha_d}, \quad (2.17)$$

sendo que  $x_1, \dots, x_d$  são as variáveis e o número  $|\alpha| = \alpha_1 + \dots + \alpha_d$  é o grau total do monômio  $\mathbf{x}^\alpha$ .

Assim, um polinômio  $P$  em  $d$  variáveis pode ser definido como uma combinação linear de monômios,

$$P(\mathbf{x}) = \sum_{\alpha} c_{\alpha} \mathbf{x}^{\alpha}, \quad (2.18)$$

em que cada coeficiente  $c_{\alpha}$  é uma constante que pertence a um corpo  $\mathbb{K}$ .

Dessa forma, o grau do polinômio é o maior grau total dos monômios e denotamos por  $\Pi^d$  o conjunto dos polinômios de  $d$  variáveis e por  $\Pi_n^d$  o conjunto dos polinômios de  $d$  variáveis e grau no máximo  $n$ . Posteriormente, trabalharemos com o espaço dos polinômios de duas variáveis  $\Pi_n^2$  e grau no máximo  $n$ . Por exemplo, a forma de um polinômio de duas variáveis e grau 3 é a seguinte:

$$P(x, y) = c_0 x^3 + c_1 x^2 y + c_2 x y^2 + c_3 y^3 + c_4 x^2 + c_5 x y + c_6 y^2 + c_7 x + c_8 y + c_9$$

onde  $c_i$  pertence a um corpo  $\mathbb{K}$ ,  $i = 1, 2, \dots, 9$ .

Dois polinômios  $P$  e  $Q$  são ditos ortogonais entre si se  $\langle P, Q \rangle = 0$ , onde  $\langle \cdot, \cdot \rangle$  é o produto interno definido em  $\Pi^d$ . Um polinômio  $P$  é chamado de polinômio ortogonal se  $P$  é ortogonal a todos polinômios de grau menor, ou seja,  $\langle P, Q \rangle = 0$ ,  $\forall Q \in \Pi^d$ , com grau  $Q < \text{grau } P$ .

Se o produto interno é dado em termos de uma função peso  $W$ ,

$$\langle P, Q \rangle = \int_{\Omega} P(x)Q(x)W(x)dx, \quad (2.19)$$

onde  $\Omega$  é o domínio em  $\mathbb{R}^d$ , dizemos que os polinômios são ortogonais com relação à função peso  $W$ .

Denotamos por  $\mathcal{V}_n^d$  o espaço dos polinômios ortogonais de grau exatamente  $n$ , isto é,

$$\mathcal{V}_n^d = \{P \in \Pi_n^d : \langle P, Q \rangle = 0, \forall Q \in \Pi_{n-1}^d\}.$$

### 2.3.1 Polinômio de Zernike

Segundo Flusser, Suk e Zitová (2009), os polinômios ortogonais em um disco são funções intrinsecamente 2D. Eles são construídos como produtos de um fator radial (geralmente um polinômio ortogonal 1D) e um fator angular que é geralmente uma função do tipo harmônica.

As próximas definições foram baseadas em Chong, Raveendran e Mukundan (2003), Khotanzad e Hong (1990), Kim e Lee (2003), Li, Lee e Pun (2009).

Por volta de 1934, Fritz Zernike introduziu um conjunto de polinômios complexos, em duas variáveis, que formam um conjunto ortogonal completo no interior de um círculo unitário, sendo, portanto, limitado pela curva  $x^2 + y^2 = 1$ . Estes polinômios, então, são chamados de polinômios de Zernike e sua forma em coordenadas polares é:

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{im\theta} \quad (2.20)$$

onde  $n \geq 0$  e  $m$  é um inteiro (negativo ou positivo) tais que  $|m| \leq n$  e  $n - |m|$  é par,  $n$  e  $m$  são denominados *ordem* e *repetição*, respectivamente. Além disso,  $\rho$  é a norma do vetor que parte da origem até o pixel  $(x, y)$  e  $\theta$  é o ângulo entre esse vetor e o eixo  $x$  no sentido anti-horário.

O polinômio radial  $R_{nm}$  é definido como:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)! \rho^{n-2s}}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \quad (2.21)$$

Se  $n - |m|$  é ímpar, então  $R_{nm}(\rho) = 0$ . E note que a função radial é simétrica com relação à repetição, isto é,  $R_{nm}(\rho) = R_{n,-m}(\rho)$ .

O polinômio radial pode ser reescrito como:

$$R_{nm}(\rho) = \sum_{k=|m|, |m|+2, \dots}^n B_{nmk} \rho^k \quad (2.22)$$

em que

$$B_{nmk} = \frac{(-1)^{\frac{n-k}{2}} \left(\frac{n+k}{2}\right)!}{\left(\frac{n-k}{2}\right)! \left(\frac{k+m}{2}\right)! \left(\frac{k-m}{2}\right)!} \quad (2.23)$$

Segundo Prata e Rusch (1989), considerando  $m \geq 0$ , outra representação para o polinômio radial pode ser obtida em termos do polinômio de Jacobi  $P_{n'}^{(\alpha, \beta)}(x)$ :

$$P_{n'}^{(\alpha, \beta)}(x) = (-1)^{n'} \frac{R_{nm}(\rho)}{\rho^\alpha}, \quad (2.24)$$

com  $x = 1 - 2\rho^2$ ,  $\alpha = m$ ,  $\beta = 0$  e  $n' = \frac{n-m}{2}$ .

Desse modo, tem-se as seguintes relações de recorrência (PRATA; RUSCH, 1989):

$$\rho R_{nm}(\rho) = \frac{1}{2(n+1)} [(n+m+2)R_{n+1, m+1}(\rho) + (n-m)R_{n-1, m+1}(\rho)] \quad (2.25)$$

$$\left. \begin{aligned} R_{n+2, m}(\rho) &= \frac{n+2}{(n+2)^2 - m^2} \left\{ \left[ 4(n+1)\rho^2 - \frac{(n+m)^2}{n} - \frac{(n-m+2)^2}{n+2} \right] \right. \\ &\left. R_{nm}(\rho) - \frac{n^2 - m^2}{n} R_{n-2, m}(\rho) \right\} \end{aligned} \right\} \quad (2.26)$$

Outra fórmula de recorrência para o polinômio radial, de acordo com Chong, Raveendran e Mukundan (2003) e Flusser, Suk e Zitová (2009), é dada por:

$$R_{nm}(\rho) = \frac{2\rho n}{n+m} R_{n-1, m-1}(\rho) - \frac{n-m}{n+m} R_{n-2, m}(\rho). \quad (2.27)$$

O cálculo dos polinômios de Zernike por essa fórmula deve começar com

$$R_{nn}(\rho) = \rho^n, \quad R_{n, -n}(\rho) = \rho^n, \quad n = 0, 1, \dots \quad (2.28)$$

e então os polinômios de Zernike podem ser calculados para  $n = 2, 3, \dots$ ;  $m = -n+2, -n+4, \dots, n-2$ .

Além disso, os polinômios radiais satisfazem a relação:

$$\int_0^1 R_{nm}(\rho) R_{lm}(\rho) \rho d\rho = \frac{1}{2(n+1)} \delta_{nl} \quad (2.29)$$

onde  $\delta_{nl}$  é o delta de Kronecker, ou seja,  $\delta_{nl} = \begin{cases} 1, & n = l \\ 0, & \text{caso contrário} \end{cases}$ .

Assim, os polinômios de Zernike são ortogonais porque satisfazem:

$$\int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) V_{n'm'}(\rho, \theta) \rho d\rho d\theta = \frac{\pi}{n+1} \delta_{nn'} \delta_{mm'} \quad (2.30)$$

em que o símbolo \* denota o complexo conjugado.

## 2.4 Momentos de uma Imagem Digital

Segundo Flusser, Suk e Zitová (2009), momentos são quantidades escalares utilizadas para caracterizar uma função e para capturar suas características significativas. Eles têm sido largamente utilizados durante centenas de anos em Estatística para a descrição da forma de uma função de densidade de probabilidade. Do ponto de vista matemático, os momentos são “projeções” de uma função em uma base polinomial.

As características capturadas pelos momentos conseguem informação global de uma imagem, sendo que momentos de análise de imagem foram originalmente introduzidos na década de 1960 (NIXON; AGUADO, 2008).

O *momento geral*  $M_{pq}$  de uma imagem  $f(x, y)$  é definido como (FLUSSER; SUK; ZITOVÁ, 2009):

$$M_{pq} = \int \int_D p_{pq}(x, y) f(x, y) dx dy \quad (2.31)$$

onde  $p$  e  $q$  são inteiros não-negativos,  $r = p + q$  é chamado a *ordem* do momento e  $p_{00}(x, y), p_{10}(x, y), \dots, p_{kj}(x, y), \dots$  são funções de base polinomiais definidas em  $D$ .

Os *momentos geométricos* ou *regulares* são o tipo mais popular e são definidos como:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (2.32)$$

onde  $m_{pq}$  é o momento de ordem  $p + q$  e  $f(x, y)$  é a função da imagem. Para imagens digitais, as integrais são substituídas por somatórios.

Outra escolha popular são os *momentos complexos*, onde a base polinomial é  $p_{kj} = (x + iy)^k (x - iy)^j$ . Assim, os momentos são definidos como:

$$c_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + iy)^p (x - iy)^q f(x, y) dx dy \quad (2.33)$$

O momento complexo pode ser expresso em coordenadas polares  $(\rho, \theta)$ :

$$\begin{aligned} x &= \rho \cos \theta & \rho &= \sqrt{x^2 + y^2} \\ y &= \rho \sin \theta & \theta &= \arctan(y/x) \end{aligned} \quad (2.34)$$

então a equação (2.33) fica na forma:

$$c_{pq} = \int_0^\infty \int_0^{2\pi} \rho^{p+q+1} e^{i(p-q)\theta} f(\rho, \theta) d\rho d\theta. \quad (2.35)$$

Momentos geométricos e momentos complexos carregam a mesma quantidade de informação.

Se a base polinomial  $p_{kj}(x, y)$  é ortogonal, isto é, se seus elementos satisfazem a condição de ortogonalidade

$$\int \int_{\Omega} p_{pq}(x, y) \cdot p_{nm}(x, y) dx dy = 0 \quad (2.36)$$

ou a ortogonalidade ponderada

$$\int \int_{\Omega} w(x, y) \cdot p_{pq}(x, y) \cdot p_{nm}(x, y) dx dy = 0 \quad (2.37)$$

para algum índice  $p \neq n$  ou  $q \neq m$ , falamos sobre os *momentos ortogonais* e  $\Omega$  é a área de ortogonalidade.

Existem vários tipos de momentos ortogonais, alguns deles são: Momentos de Legendre, Momentos de Chebyshev, Momentos de Zernike, Momentos de Fourier-Mellin, entre outros.

De acordo com Flusser, Suk e Zitová (2009), em teoria, todas as bases polinomiais de mesmo grau são equivalentes porque geram o mesmo espaço de funções. Qualquer momento no que diz respeito a uma determinada base pode ser expressa em termos de momentos em relação a qualquer outra base. Deste ponto de vista, momentos ortogonais de qualquer tipo são equivalentes aos momentos geométricos.

Entretanto, uma diferença significativa aparece quando se considera a estabilidade e as questões computacionais em um domínio discreto. Momentos geométricos necessitam de alta precisão computacional e momentos ortogonais podem capturar as características da imagem de uma maneira melhor, não redundante. Eles também têm a vantagem de exigir menor precisão computacional porque podemos avaliá-los usando as relações de recorrência.

### 2.4.1 Momentos de Zernike

Ao implementar momentos baseados em polinômios ortogonais sobre um disco, a imagem deve ser mapeada em um disco de ortogonalidade que cria certos problemas de

re-amostragem. Por outro lado, os momentos ortogonais podem ser facilmente utilizados para a construção de rotação invariante porque eles mudam de uma maneira simples (FLUSSER; SUK; ZITOVÁ, 2009).

Segundo Flusser, Suk e Zitová (2009), momentos ortogonais em um disco geralmente são construídos na forma:

$$v_{pq} = n_{pq} \int_0^{2\pi} \int_0^1 R_{pq}(\rho) e^{-iq\theta} f(\rho, \theta) \rho d\rho d\theta, p = 0, 1, 2, \dots; q = -p, \dots, p \quad (2.38)$$

em que  $n_{pq}$  é algum fator de normalização,  $\rho, \theta$  são coordenadas polares,  $R_{pq}$  é a parte radial e  $e^{iq\theta}$  é a parte angular do respectivo polinômio. A área de ortogonalidade é um disco unitário  $\Omega = \{(x, y) | x^2 + y^2 \leq 1\}$ . Para calcular os momentos, a imagem  $f$  deve ser redimensionada de maneira que esteja contida em  $\Omega$ .

Os momentos de Zernike de ordem  $n$  com repetição  $m$  são definidos como (CELEBI; ASLANDOGAN, 2005; KHOTANZAD; HONG, 1990; KIM; LEE, 2003):

$$Z_{nm} = \frac{n+1}{\pi} \int \int_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(x, y) dx dy \quad (2.39)$$

ou em coordenadas polares (FLUSSER; SUK; ZITOVÁ, 2009):

$$Z_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta) V_{nm}^*(\rho, \theta) \rho d\rho d\theta \quad (2.40)$$

lembrando que  $|m| \leq n$  e  $n - |m|$  é par.

Momentos de Zernike foram introduzidos em análise de imagens na década de 1980 por Teague (1980).

Imagens digitais devem ser mapeadas no disco unitário antes dos momentos de Zernike serem calculados, ou seja, o centro da imagem ou do objeto de interesse é considerado como o centro do disco unitário ( $x^2 + y^2 \leq 1$ ) e as coordenadas do pixel são mapeadas para este disco e os pixels fora dele não são usados no cálculo, conforme ilustra a Figura 20.

Para realizar esse mapeamento, o raio  $\rho$  e o ângulo  $\theta$  no pixel  $(x, y)$  é dado por (HWANG; KI, 2006):

$$\rho = \frac{\sqrt{(2x - N + 1)^2 + (N - 1 - 2y)^2}}{N}, \theta = \arctan\left(\frac{N - 1 - 2y}{2x - N + 1}\right) \quad (2.41)$$

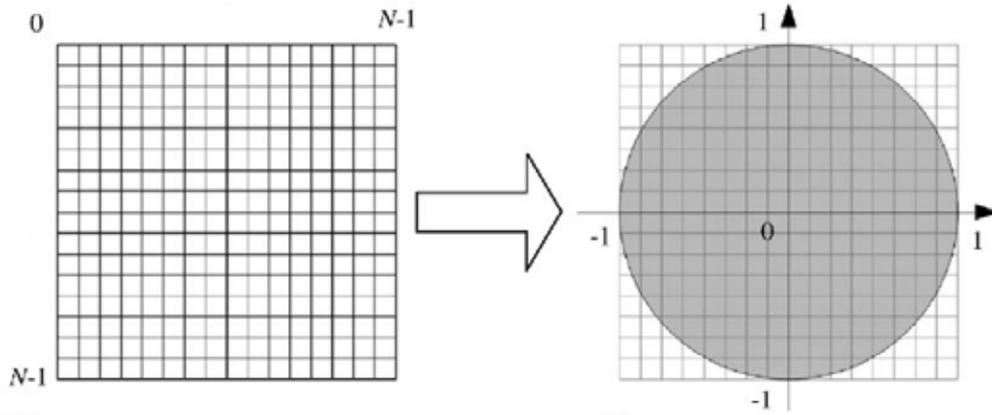


Figura 20: Mapeamento de uma imagem  $N \times N$  sobre o disco unitário.

Se a imagem tem tamanho  $N \times N$ , então as integrais são substituídas por somatórios e tem-se (KHOTANZAD; HONG, 1990; HWANG; KI, 2006):

$$Z_{nm} = \frac{n+1}{\lambda_N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) V_{nm}^*(\rho, \theta) \quad (2.42)$$

em que  $0 \leq \rho \leq 1$  e  $\lambda_N$  é um fator de normalização dado pelo número de pixels localizados no disco unitário.

Os coeficientes  $B_{nmk}$  definidos pela equação (2.23) podem ser usados para expressar os momentos de Zernike usando momentos geométricos (FLUSSER; SUK; ZITOVÁ, 2009):

$$Z_{nm} = \frac{n+1}{\pi} \sum_{k=|m|, |m|+2, \dots}^n \sum_{j=0}^{\frac{k-|m|}{2}} \sum_{l=0}^{|m|} \binom{\frac{k-m}{2}}{j} \binom{|m|}{l} w^l B_{nmk} m_{k-2j-l, 2j+l} \quad (2.43)$$

em que

$$w = \begin{cases} -i, & m > 0 \\ i, & m \leq 0. \end{cases}$$

Os momentos de Zernike são invariantes em relação à transformação de rotação mas, como será mostrado abaixo, podem também ser invariantes em relação às transformações de escala e translação se a imagem de entrada passar por uma manipulação prévia baseada na teoria dos momentos geométricos (KHOTANZAD; HONG, 1990).

Considere uma imagem rotacionada por um ângulo  $\alpha$ , o momento de Zernike  $Z_{nm}^r$  da imagem rotacionada é dado por (CELEBI; ASLANDOGAN, 2005; KHOTANZAD; HONG, 1990; KIM; LEE, 2003):

$$Z_{nm}^r = Z_{nm} e^{-im\alpha}. \quad (2.44)$$

De fato, se a imagem rotacionada é denotada por  $f^r(x, y)$  então sua relação com a imagem original  $f(x, y)$  no mesmo sistema de coordenadas polares é dada por  $f^r(\rho, \theta) = f(\rho, \theta - \alpha)$ . Desse modo, obtemos:

$$\begin{aligned} Z_{nm}^r &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f^r(\rho, \theta) V_{nm}^*(\rho, \theta) \rho d\rho d\theta = \\ &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta - \alpha) R_{nm}(\rho) e^{-im\theta} \rho d\rho d\theta \end{aligned}$$

Fazendo a mudança de variável  $\theta_1 = \theta - \alpha$ , temos:

$$\begin{aligned} Z_{nm}^r &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta_1) R_{nm}(\rho) e^{-im(\theta_1 + \alpha)} \rho d\rho d\theta_1 = \\ &= \left[ \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta_1) R_{nm}(\rho) e^{-im\theta_1} \rho d\rho d\theta_1 \right] e^{-im\alpha} = \\ &= Z_{nm} e^{-im\alpha}. \end{aligned}$$

Isto significa que a magnitude dos momentos de Zernike de uma imagem rotacionada permanece a mesma após a rotação, isto é,  $|Z_{nm}^r| = |Z_{nm} e^{-im\alpha}| = |Z_{nm}|$ . Dessa forma, a magnitude dos momentos de Zernike pode ser tomada como uma característica de rotação invariante da imagem.

Apenas lembrando que  $e^{-im\alpha} = \cos(m\alpha) - i \sin(m\alpha)$ , logo,  
 $|e^{-im\alpha}| = \sqrt{\cos^2(m\alpha) + \sin^2(m\alpha)} = 1$ .

As invariâncias por mudança de escala e por translação podem ser obtidas através de um processo de normalização da imagem usando os momentos regulares ou geométricos de uma imagem.

A invariância por translação é obtida através da transformação da imagem de entrada em uma nova, movendo a origem da imagem original e fazendo com que os primeiros momentos,  $m_{01}$  e  $m_{10}$ , sejam ambos iguais a zero (KHOTANZAD; HONG, 1990). Isto é feito através da transformação da imagem original  $f(x, y)$  em outra,  $f(x + \bar{x}, y + \bar{y})$ , onde  $\bar{x}$  e  $\bar{y}$  são a localização do centro de gravidade ou centróide da imagem, calculado por:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (2.45)$$

Em outras palavras, a origem é movida para o centróide da imagem antes do cálculo dos momentos.

A invariância em relação à mudança de uma mesma escala em todas as direções é

garantida pela ampliação ou redução de cada forma, tal que seu momento geométrico de ordem zero  $m_{00}$  seja igual a um valor pré-determinado  $\beta$ . Note que  $m_{00}$  é a “massa” da imagem e, para imagens binárias,  $m_{00}$  é a área do objeto (FLUSSER; SUK; ZITOVÁ, 2009).

Seja  $f(\frac{x}{a}, \frac{y}{a})$  a função que representa uma versão em escala da função da imagem  $f(x, y)$ . Então seu momento regular,  $m'_{pq}$ , é dado por:

$$m'_{pq} = a^{p+q+2}m_{pq} \quad (2.46)$$

De fato, tem-se (KHOTANZAD; HONG, 1990):

$$\begin{aligned} m'_{pq} &= \int_x \int_y x^p y^q f\left(\frac{x}{a}, \frac{y}{a}\right) dx dy = \int_x \int_y a^p x^p a^q y^q f(x, y) a^2 dx dy = \\ &= \int_x \int_y a^{p+q+2} x^p y^q f(x, y) dx dy = a^{p+q+2} \int_x \int_y x^p y^q f(x, y) dx dy = \\ &= a^{p+q+2} m_{pq} \end{aligned}$$

Como o objetivo é obter  $m'_{00} = \beta$ , pode-se fazer  $a = \sqrt{\frac{\beta}{m_{00}}}$  e, substituindo em  $m'_{00}$ , temos  $m'_{00} = a^2 m_{00} = \beta$ . Assim, a invariância em relação à mudança de uma mesma escala em todas as direções (similaridade) é alcançada transformando a função da imagem original em uma nova função  $f(\frac{x}{a}, \frac{y}{a})$ , com  $a = \sqrt{\frac{\beta}{m_{00}}}$ .

Resumindo, uma imagem  $f(x, y)$  pode ser normalizada, preservando as propriedades de invariância à translação e à escala, transformando-a em (KHOTANZAD; HONG, 1990; KIM; LEE, 2003):

$$g(x, y) = f\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}\right) \quad (2.47)$$

em que  $(\bar{x}, \bar{y})$  é o centróide de  $f(x, y)$  e  $a = \sqrt{\frac{\beta}{m_{00}}}$ , com  $\beta$  um valor pré-determinado.

Os erros envolvidos na implementação discreta podem ser reduzidos por interpolação. Se a coordenada  $(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y})$  não coincide com uma localização da malha da imagem de entrada, o valor da função da imagem  $f$  pode ser interpolado a partir dos valores dos quatro pixels adjacentes.

## 2.5 Descritor de forma de Zernike

Note que  $Z_{nm}^* = Z_{n,-m}$ , então  $|Z_{nm}| = |Z_{n,-m}|$ . Logo, pode-se concentrar em definir os momentos de Zernike por  $|Z_{nm}|$  com  $m \geq 0$ .

Para a construção do descritor de forma com base nos momentos de Zernike, um vetor

de características é construído com esses momentos para a imagem de entrada:

$$\vec{v} = (v_1, v_2, \dots, v_L) = (|Z_{00}|, |Z_{11}|, |Z_{20}|, |Z_{22}|, |Z_{31}|, \dots, |Z_{n_{max}, n_{max}}|) \quad (2.48)$$

onde  $L$  é a dimensão do vetor e  $n_{max}$  é a ordem máxima que será calculada para os momentos de Zernike, onde se procura a ordem ótima que gera a imagem reconstruída da forma mais fiel possível.

Vale lembrar que descritores de forma podem ser amplamente categorizados como *descritores baseados no contorno*, que usam informações apenas dos limites da forma, ignorando o conteúdo do interior, enquanto que *descritores baseados na região* exploram tanto os limites quanto os pixels da região interna da forma, sendo portanto aplicáveis para formas genéricas (CELEBI; ASLANDOGAN, 2005).

Sendo assim, descritores baseados em momentos são um exemplo de descritor de forma baseado na região. Em particular, estudamos o descritor baseado nos momentos de Zernike que, como mostrado, possui características atrativas para sua utilização.

## 3 METODOLOGIA

Para melhor compreensão dos procedimentos realizados no trabalho, este capítulo é dedicado a descrever brevemente os *softwares* utilizados no desenvolvimento do trabalho e quais os dados usados nos experimentos. O funcionamento do método proposto para o reconhecimento de contornos de edifício usando os momentos de Zernike também é exposto, bem como a maneira como os resultados foram avaliados.

### 3.1 Recursos Computacionais e Dados Utilizados

Para a realização da pesquisa, foram utilizados os *softwares* HALCON, MATLAB e Photoshop.

O HALCON é um *software* abrangente para visão de máquina com um ambiente de desenvolvimento integrado (IDE), que reúne características e ferramentas para aumentar a produtividade dos desenvolvedores de *software*, bem como a qualidade desses produtos. O HALCON pode ser tanto utilizado como um aplicativo para uma infinidade de tarefas ou pode ser utilizado como uma poderosa biblioteca de classes ou funções para programação em C e C++. Em outras palavras, neste *software* estão disponíveis inúmeros métodos de processamento e análise de imagens, que podem, conforme dito acima, ser utilizados via aplicativos fechados ou via programas desenvolvidos em C/C++ pelo usuário, nos quais são anexadas as classes ou funções.

O MATLAB é um *software* de alto nível e um ambiente interativo para computação numérica, visualização e programação. Pode-se analisar dados, desenvolver algoritmos e criar modelos e aplicações. A linguagem, as ferramentas e as funções matemáticas embutidas permitem explorar múltiplas abordagens e chegar a uma solução mais rápida do que com planilhas ou linguagens de programação tradicionais como C/C++ ou Java. O MATLAB pode ser usado para uma variedade de aplicações, incluindo o processamento de sinais, imagem e vídeo, sistemas de controle, finanças computacionais e biologia computacional.

O Photoshop é, basicamente, um *software* para edição de imagens, com diversas ferramentas como o recorte de elementos, adição de filtros e efeitos.

Para a análise com dados reais, foi utilizada uma imagem aérea RGB de alta resolução obtida a partir da digitalização do diapositivo colorido original que corresponde à área do Campus III da Universidade Federal do Paraná (UFPR), possui dimensão de  $9728 \times 9728$  pixels e resolução espacial de aproximadamente 0,25m.

Três recortes foram definidos dessa imagem e as dimensões de cada recorte são:

- Recorte 1:  $269 \times 215$  pixels;
- Recorte 2:  $417 \times 159$  pixels;
- Recorte 3:  $406 \times 396$  pixels.

Além dos dados reais, com o intuito de verificar a eficácia do método proposto, alguns dados foram gerados manualmente. Estes dados consistem no esboço de alguns tipos de contornos de telhado de edifício (em forma de L, retangular e em forma de quadrado).

## 3.2 Método

### 3.2.1 Reconhecimento de Contornos de Edifício usando Momentos de Zernike

A Figura 21 mostra um fluxograma do funcionamento do método de reconhecimento de contornos de edifício usando os momentos de Zernike, sendo que cada etapa será detalhada em seguida.

Primeiramente, é preciso transformar a imagem original colorida em uma imagem em escala de cinza. Para isso, utilizou-se o índice de artificialidade NandA (*Natural and Artificial*), dado por (VALE, 2007):

$$NandA = G - (R + B) \quad (3.1)$$

em que R, G e B correspondem às respectivas bandas vermelho (**R**ed), verde (**G**reen) e azul (**B**lue).

Este índice visa realçar a vegetação, deixando-a em tons mais claros enquanto que as edificações e rodovias são levadas para tons mais escuros. Quanto à tonalidade do solo

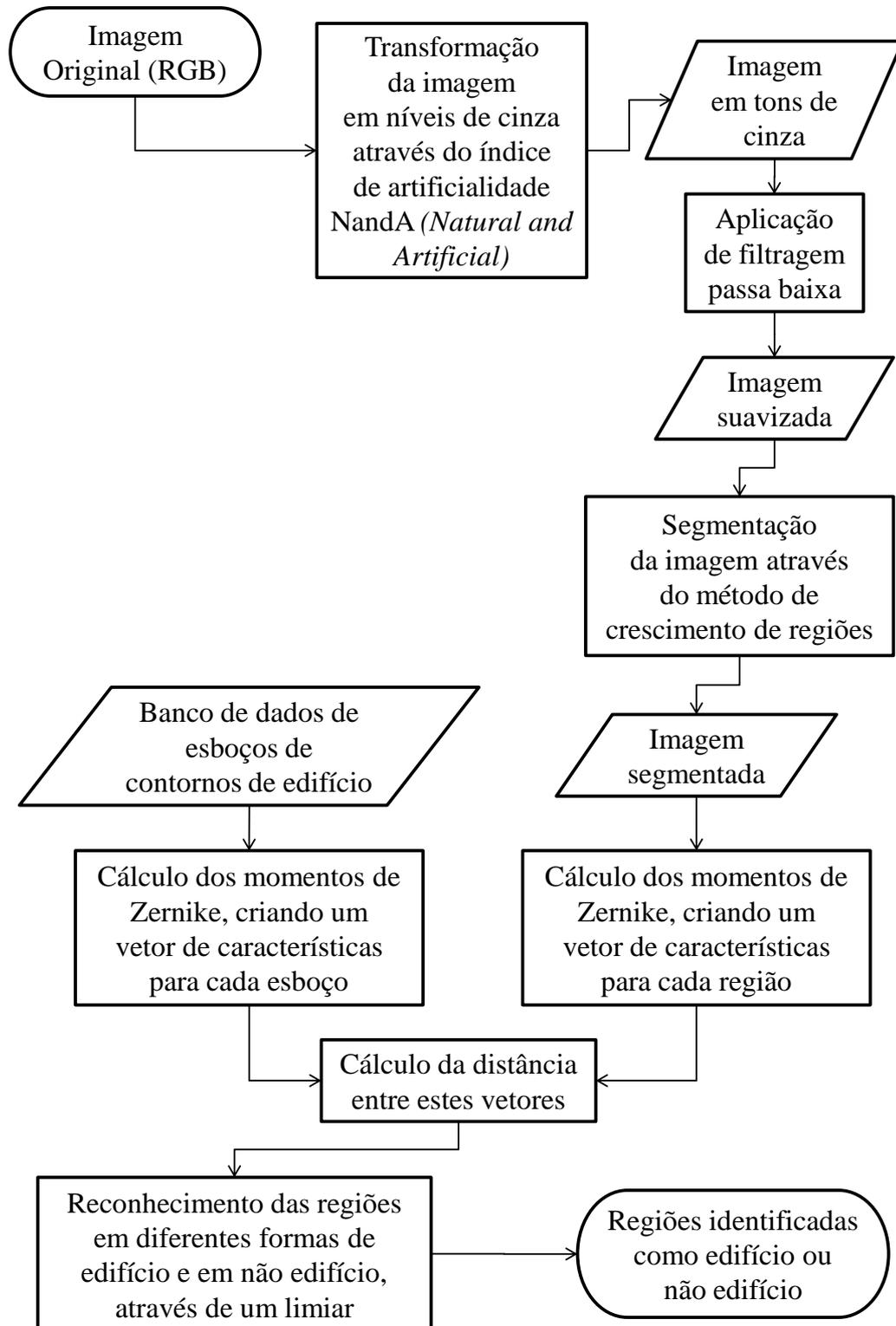


Figura 21: Fluxograma do funcionamento do método proposto.

exposto, como ruas e calçadas, ela dependerá do tipo, que em geral fica numa tonalidade constante e intermediária entre o valor da vegetação e o das edificações o que, dependendo da situação, poderá gerar alguns problemas no processo de segmentação.

Como o objeto de interesse deste estudo são as edificações, foi utilizada a inversão da imagem produzida pelo NandA, de maneira que a vegetação fique em tons mais escuros e as edificações sejam realçadas em tons mais claros. Por consequência, o objetivo desse procedimento é eliminar alguns falsos positivos e aumentar a qualidade da segmentação.

Depois de transformar a imagem original em níveis de cinza, é necessária a utilização de alguma técnica de filtragem da imagem, a fim de obter melhores resultados, principalmente para eliminação de ruído. O filtro da Moda foi o escolhido nos testes realizados, pois foi o método que apresentou melhores resultados.

Posteriormente, deve-se segmentar por regiões a imagem a ser trabalhada, em que são geradas regiões (segmentos) da imagem a serem submetidas à análise de forma com base no método dos momentos de Zernike. E assim, através da segmentação, é possível melhor interpretar a imagem, que é particionada em regiões que representam seus diferentes objetos ou partes de objetos.

Foram avaliados vários métodos de segmentação de imagens disponíveis no aplicativo HALCON da MVTech. Vários testes com diferentes métodos de limiarização e o método de crescimento de regiões usando diferentes limiares para a segmentação das imagens foram realizados. Estes testes poderão ser vistos no próximo capítulo. Visualmente, o método de crescimento de regiões apresentou melhores resultados e foi o escolhido para realizar os experimentos.

Basicamente, o método de crescimento de regiões presente no aplicativo, através da função *regiongrowing*, parte de regiões sementes (retângulos determinados automaticamente pelo método, em que o usuário define apenas o tamanho dessas regiões) que crescem para formar regiões maiores. Assim, duas regiões sementes vão pertencer ao mesmo segmento se a diferença entre o nível de cinza de seus pontos centrais for menor ou igual a um dado limiar (tolerância), sendo que apenas as regiões com um tamanho mínimo (quantidade de pixels) são retornadas na imagem de saída.

Nos experimentos realizados, as regiões sementes foram definidas e fixadas como retângulos com dimensão de um pixel, ou seja, o método iniciou analisando pixel a pixel como uma região para depois, através do limiar entre os níveis de cinza (que foi modificado para cada caso), determinar a formação final de cada região. E ainda, como  $30m^2$

é a área mínima para que uma região seja considerada como uma possível edificação e a resolução espacial da imagem utilizada é de aproximadamente 0,25m, então a quantidade mínima de pixels para que uma região fosse retornada no processo de segmentação foi fixada em 480.

Com o intuito de refinar o resultado do processo de segmentação das imagens, foi usada uma técnica existente no HALCON, chamada *fill\_up\_shape*, para o preenchimento das regiões com “buracos”. Este operador se baseia na área de cada região, em que se definem valores mínimo e máximo para a área da região e, assim, preenche todos os pixels no interior da região que tiver a área entre esses limiares. Desse modo, o *fill\_up\_shape* não leva em consideração o tamanho dos buracos, mas sim a área das regiões. Então foram definidos limiar mínimo 1 e limiar máximo 1500, de maneira que todas as regiões encontradas pelo método de segmentação sejam compreendidas pelo operador e tenham seus buracos preenchidos.

Todo este processo de segmentação foi onde se encontrou maior dificuldade no trabalho, pois os métodos não foram capazes de segmentar fielmente as regiões das imagens, o que levou a algumas falhas. A etapa de segmentação da imagem é normalmente um grande obstáculo a ser resolvido em qualquer método de reconhecimento de objetos baseado em regiões.

Após a segmentação da imagem, foi feita a descrição de forma de regiões através dos momentos complexos de Zernike, onde cada região foi analisada individualmente como uma imagem binária (preto e branco), pois o que interessa para a análise é a forma dessa região. Para isso, primeiramente, deve-se redimensionar todas as regiões de interesse de modo que tenham as mesmas dimensões, sejam quadradas (do tipo  $N \times N$ ) e possuam a mesma área (momento geométrico de ordem zero  $m_{00}$  seja igual a um valor pré-determinado  $\beta$ ). Posteriormente, cada região é transladada para o centro da imagem e, assim, obtém-se as invariâncias por escala e translação como descrito no Capítulo 2. Realizadas essas transformações, são criados vetores de características para cada região. Sendo assim, o mapeamento para o disco unitário, necessário para o cálculo dos momentos de Zernike, é feito na região de interesse e não na imagem inteira.

Além disso, foi construído um banco de dados contendo esboços ou modelos de contornos de telhados, passíveis de serem encontrados numa dada cena e que podem servir de referência para o reconhecimento de edificações em uma imagem digital. As propriedades de invariância do método dos momentos de Zernike implicam em grande flexibilidade na construção desses esboços. Porém, é importante ressaltar que a invariância é com relação

a uma mesma escala em todas direções (invariância por similaridade). Então o método não é capaz de identificar qualquer tipo de forma retangular, por exemplo, com apenas um esboço de um tipo de retângulo no banco de dados. Isto quer dizer que não é possível criar um banco de dados que atenda a todas as variações possíveis nas dimensões dos esboços de contorno de edifício em forma de L e em forma retangular. No entanto, buscou-se construir esboços o bastante, mesmo que alguns sejam parecidos um com o outro, para se obter resultados satisfatórios. O banco de dados utilizado no trabalho pode ser visto na Figura 22.

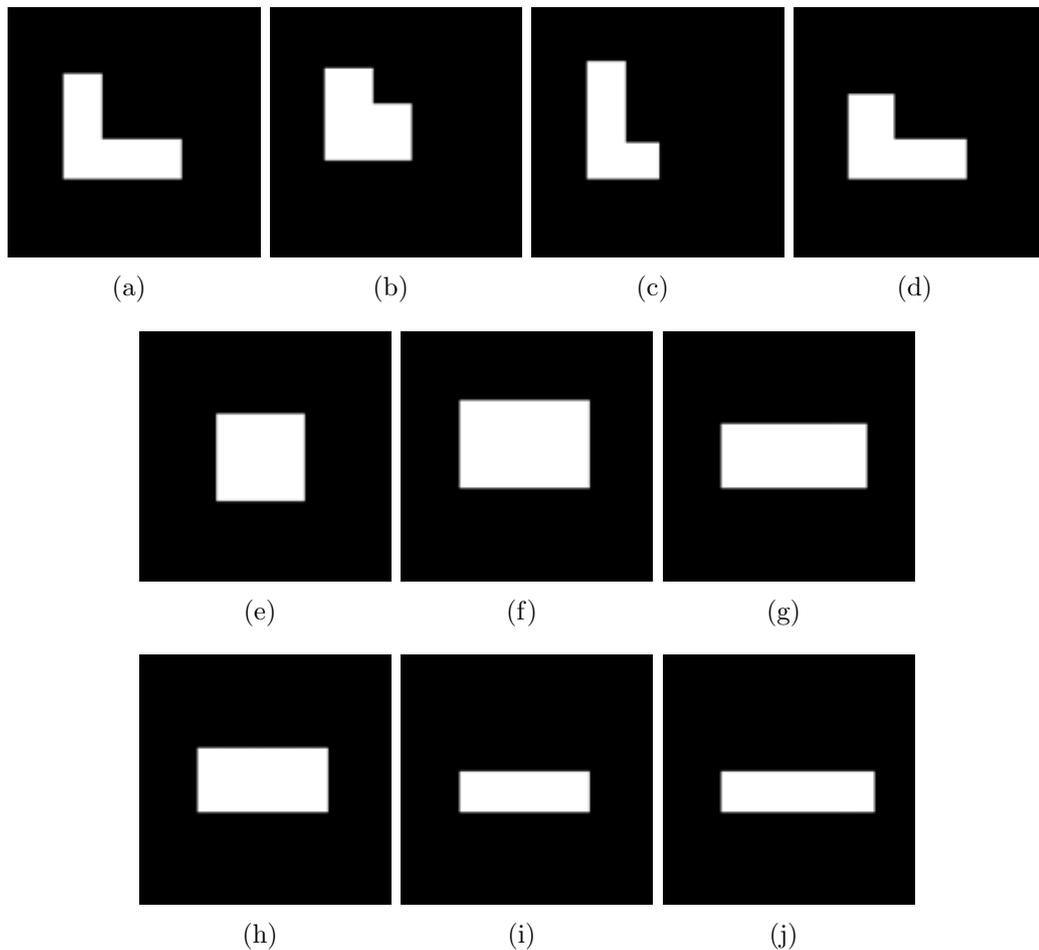


Figura 22: Esboços de contornos de edifício contidos no banco de dados.

Outra motivação para a criação deste banco de dados e o uso de momentos de Zernike no reconhecimento de edifícios é o fato de que eles são robustos a ruído e pequenas variações na forma (CELEBI; ASLANDOGAN, 2005; KIM; KIM, 2000). Desse modo, mesmo que a região segmentada tenha algumas irregularidades nos contornos, o método irá reconhecer a forma mais próxima para essa região.

Para avaliar a capacidade dos descritores de Zernike na discriminação entre diferentes

formas de contornos de edifícios e entre formas de contornos de edifícios e de outros objetos, o método separa as regiões segmentadas em duas classes: regiões correspondentes a edifícios e regiões correspondentes a outros objetos. Para isso, o método tem por base a distância Euclidiana entre o vetor de características de Zernike calculado para uma dada região extraída da imagem e os vetores de características de Zernike associados a cada esboço do banco de dados, utilizando-se a mesma ordem máxima para os momentos de Zernike em ambos os casos.

Quando a região segmentada na imagem for um edifício, a menor distância Euclidiana deve permitir a identificação do tipo de edifício. Entretanto, quando não se tratar de edifício, o desafio é encontrar um limiar para a distância entre os vetores que permita rotular a região como outro objeto.

Outra dificuldade é determinar a ordem ótima dos vetores de característica, isto é, a ordem máxima dos momentos de Zernike que permite bem caracterizar as regiões para o sucesso do método. O valor desta ordem foi definido experimentalmente através dos testes realizados. Sendo que, a cada caso, os modelos do banco de dados são redimensionados da mesma maneira que as regiões extraídas da imagem, pois os esboços precisam ter a mesma dimensão e possuir a mesma área que as regiões segmentadas.

As implementações foram realizadas no aplicativo MATLAB, sendo que o algoritmo disponibilizado em <http://www.utdallas.edu/~a.tahmasbi/Zernike.html>, página do estudante de doutorado no Departamento de Engenharia Elétrica da Universidade do Texas, auxiliou o desenvolvimento da implementação completa (TAHMASBI; SAKI; SHOKOUHI, 2011; TAHMASBI et al., 2011).

### 3.2.2 Índices de Avaliação dos Resultados

Como meio de avaliação da eficácia do método proposto, foram feitas comparações entre os resultados obtidos com o método de extração e os correspondentes resultados obtidos visualmente, através da verificação de edifícios não identificados (falsos negativos), edifícios identificados incorretamente (falsos positivos) e pela razão de extração de edifícios (REE), proposta por Rüter, Martine e Mtalo (2002) e dada por:

$$REE = \frac{EC}{EC + EI} \cdot 100, \quad (3.2)$$

em que EC é o número de estruturas identificadas corretamente pelo método e EI é o número de estruturas identificadas incorretamente.

Além disso, também foi verificada a acurácia de detecção de edifícios (ADE), definida por Mahi, Isabaten e Serief (2014) como:

$$ADE = \frac{EC}{TE} \cdot 100, \quad (3.3)$$

sendo TE o número total de edifícios existentes na cena.

Cabe ressaltar que as porcentagens de falsos positivos e a REE são complementares (a soma de ambos é 100%) e ainda, os falsos negativos e a ADE também são complementares.

## 4 *RESULTADOS*

Neste capítulo são apresentados os experimentos realizados no decorrer do desenvolvimento do trabalho. De início, foram feitos vários testes no aplicativo HALCON com diferentes métodos de segmentação, a fim de determinar qual o método mais apropriado a ser utilizado.

Definido o método de segmentação a ser usado, foram realizados experimentos para o reconhecimento de forma de contornos de edifício. Primeiramente, com o intuito de verificar que os momentos de Zernike são invariantes com relação às transformações de rotação, translação e escala (similaridade), foram feitos vários testes em esboços desenhados manualmente e que passaram por essas transformações. Neste sentido, também foi realizado um experimento com uma cena simulada para verificar a eficácia do método, sendo que esta cena consiste em edifícios do banco de dados e os desenhados manualmente. Assim, por ser um caso ideal, é de se esperar que a segmentação da imagem divida corretamente todas as regiões. Por consequência, o método proposto deve obter resultados muito satisfatórios.

Por fim, foram realizados experimentos com três recortes de dados reais. Sendo que em um dos experimentos houve a comparação do desempenho do método quando a segmentação foi realizada automaticamente pelo aplicativo HALCON e quando houve a delimitação e o preenchimento manual dos contornos de edifício presentes na cena.

Vale ressaltar que os resultados dos experimentos com a cena simulada e os dados reais foram avaliados numericamente pela razão de extração de edifícios e a acurácia de detecção de edifícios, em que se analisa os edifícios detectados corretamente pelo método e a quantidade de edifícios identificados dentre os existentes na cena.

## 4.1 Experimentos para Escolha do Método de Segmentação

Foram realizados alguns testes com diferentes métodos de segmentação presentes no aplicativo HALCON e serão apresentados a seguir. A Figura 23 mostra a imagem utilizada nos testes, já transformada em escala de cinza através do índice de artificialidade Nanda, bem como seu histograma é apresentado na Figura 24.



Figura 23: Imagem em escala de cinza.

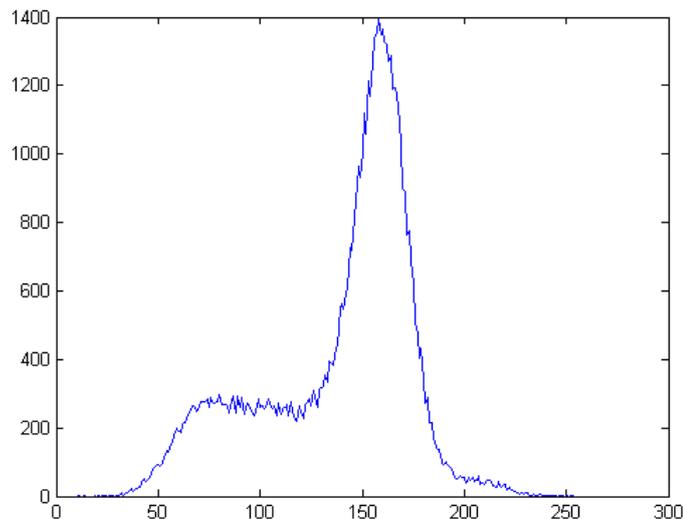


Figura 24: Histograma da Figura 23.

Analisando este histograma, observa-se que existe um pico, supostamente correspondente aos pixels das edificações, por estarem em tons mais claros e cobrirem maior parte da imagem. Isto nos motiva a usar a função *threshold*, encontrada no aplicativo HALCON, que segmenta uma imagem usando um limiar global. Esta função seleciona os pixels da imagem de entrada cujos valores de nível de cinza estão entre os valores mínimo e máximo dados pelo usuário, retornando apenas uma região. Na Figura 25 pode ser visto o resultado dessa segmentação, com os valores de nível de cinza mínimo 125 e máximo 190, em que a cor vermelha representa a região dada.

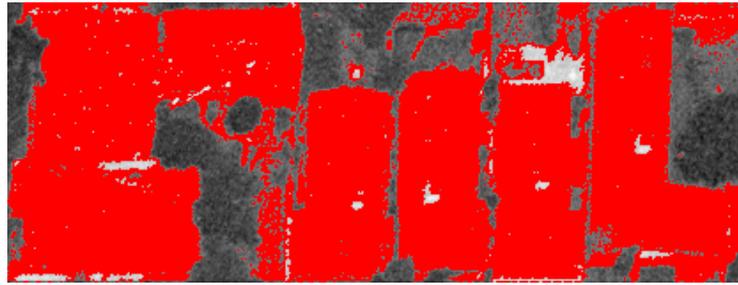


Figura 25: Segmentação com dois limiares dados.

A função *threshold* opera usando o método da Moda, mas como encontram-se vários edifícios na imagem com pouca variação nos níveis de cinza de um para o outro, não é possível segmentar cada edifício em diferentes regiões. Sendo assim, o uso deste método não se mostra interessante. O mesmo aconteceu com a função *auto\_threshold*.

A *auto\_threshold* segmenta a imagem usando múltiplos limiares determinados automaticamente a partir de seu histograma. Os mínimos relevantes são extraídos a partir do histograma da imagem e são usados sucessivamente como parâmetros para uma operação de limiarização. Para cada intervalo de valores de nível de cinza, uma região é gerada. Assim, o número de regiões retornadas pelo método é o número de mínimos mais um. A Figura 26 apresenta o resultado da segmentação da Figura 23, onde as diferentes cores indicam cada região.



Figura 26: Segmentação com limiares definidos automaticamente.

Observa-se que o método separa o fundo da imagem e seus objetos, mas a utilização dele também não se mostra interessante, pois novamente, os edifícios foram classificados em uma única região. Além disso, percebe-se que há regiões bastante fragmentadas, com alguns pixels isolados. Assim, pode-se constatar a limitação destes métodos que não levam em consideração a localização espacial dos pixels para determinar os limiares, pois são métodos pontuais que se baseiam apenas no histograma da imagem. Dependendo do caso, estes métodos são interessantes para separar o fundo da imagem dos objetos nela contidos.

Outro método testado baseia-se nos contornos dos objetos, em que encontra-se a função *hysteresis\_threshold* no HALCON. Ela é baseada em contornos e realiza uma operação de limiar histerese em uma imagem. Ele utiliza os parâmetros *baixo*, *alto* e *comprimento máximo*, sendo que os pixels da imagem com nível de cinza maior ou igual a *alto* são aceitos como pontos “seguros”. Por outro lado, todos os pixels com valores menores que *baixo* são rejeitados. Os pixels com nível de cinza entre estes dois limiares são ditos pontos “potenciais” e são aceitos se eles são conectados a pontos “seguros” por um caminho de pontos “potenciais” com comprimento de, no máximo, *comprimento máximo*. Isto significa que pontos “seguros” influenciam sua vizinhança (histerese).

Antes de aplicar este operador, é necessário utilizar um detector de bordas na imagem. No exemplo da Figura 27, o detector de borda utilizado foi o de Sobel de máscara  $3 \times 3$  e os parâmetros do operador de segmentação foram *baixo*=15, *alto*=30 e *comprimento máximo*=10.



Figura 27: Segmentação baseada nos contornos.

Percebe-se que o método resulta no contorno de uma região que precisará ser preenchido por algum outro método. A maior dificuldade de métodos deste tipo é o fechamento dos contornos e é provável que seja necessário outros processamentos para que seja realizado o devido preenchimento e finalizar a segmentação. Assim, o resultado ainda não é satisfatório para a análise que é realizada neste trabalho.

Na Figura 28 pode ser visto o resultado da segmentação da imagem através do método de crescimento de regiões. As regiões sementes foram iniciadas como retângulos de um pixel, a tolerância para a diferença entre os níveis de cinza foi 9 e o tamanho mínimo para uma região ser retornada foi 480. É visto que o resultado ainda não é totalmente satisfatório, porém, é muito melhor que os resultados dados pelos métodos anteriores. Por este motivo, ele foi o escolhido para os experimentos seguintes, sendo que na seção 4.4 pode ser visto que através de outras operações os resultados podem ser melhorados.

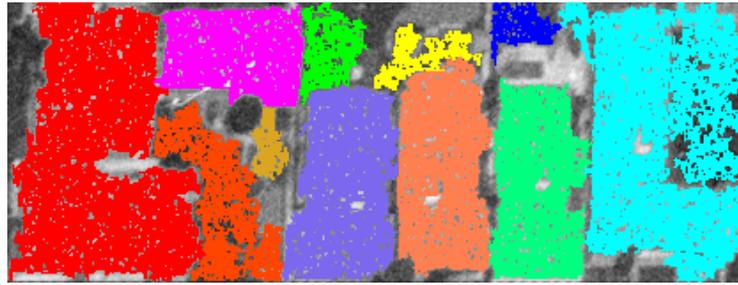


Figura 28: Segmentação através do método de crescimento de regiões.

É válido comentar que, além dos testes realizados com o recorte apresentado na Figura 23, foram feitos testes com outros recortes de imagem para os diferentes métodos de segmentação citados. Sendo que os resultados encontrados foram parecidos com os apresentados acima.

## 4.2 Verificação da Invariância por Rotação, Escala e Translação

Este experimento consiste em verificar a invariância dos momentos de Zernike por rotação, escala e translação. Para isso, foram desenhados manualmente esboços de contornos de três tipos de edifício (em forma de L, retangular e em forma de quadrado), construídos sem se preocupar com as proporções dos esboços encontrados no banco de dados. Então foram digitalizados, segmentados e posteriormente realizadas as transformações. Estes contornos estão apresentados na Figura 29.



(a) Esboço de edifício em forma de L. (b) Esboço de edifício em forma de quadrado. (c) Esboço de edifício em forma retangular.

Figura 29: Esboços de contornos de edifício desenhados manualmente.

Após a segmentação, cada esboço foi rotacionado no sentido horário pelos ângulos de  $45^\circ$ ,  $90^\circ$  e  $120^\circ$ , com exceção do edifício em forma de quadrado que, ao invés de  $90^\circ$  foi

rotacionado em  $60^\circ$ . Além disso, cada esboço de referência (Figuras 30a, 31a e 32a) teve suas dimensões aumentadas em 1,43 e reduzidas em 0,57, sendo que o de escala menor foi transladado para baixo e para cima da imagem. Desse modo, obteve-se 8 diferentes imagens para cada tipo de esboço de edifício, conforme mostram as Figuras 30, 31 e 32.

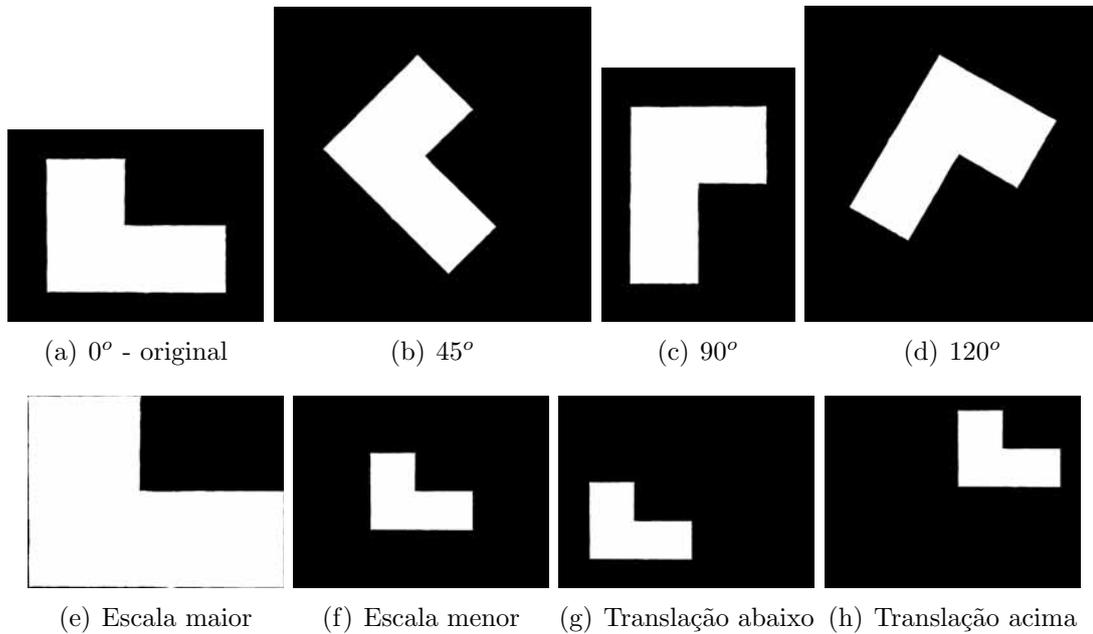


Figura 30: Esboços de contornos de edifício em forma de L.

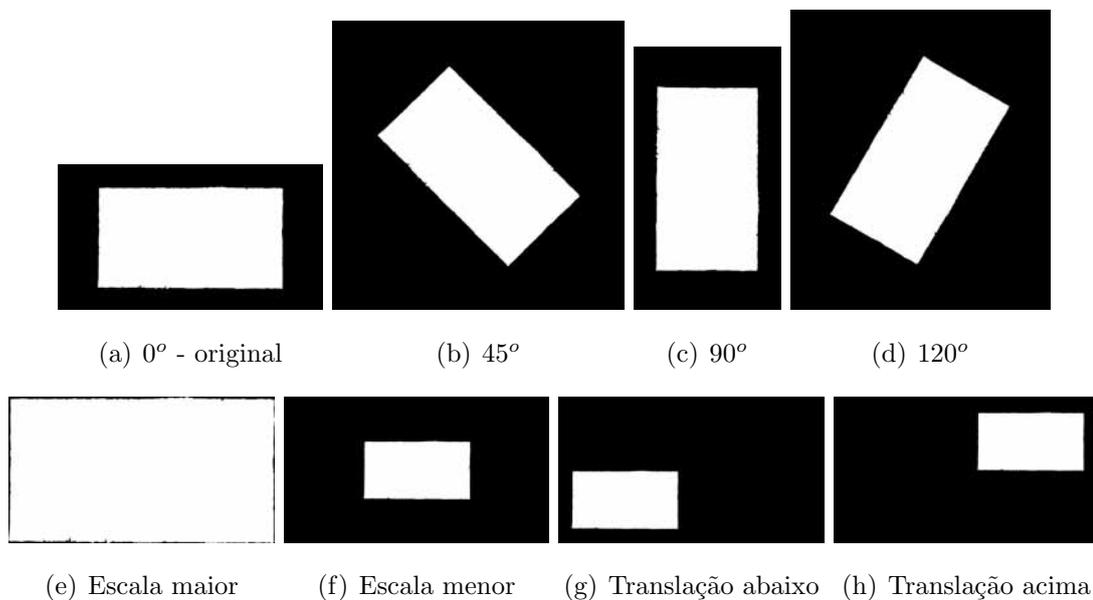


Figura 31: Esboços de contornos de edifício em forma retangular.

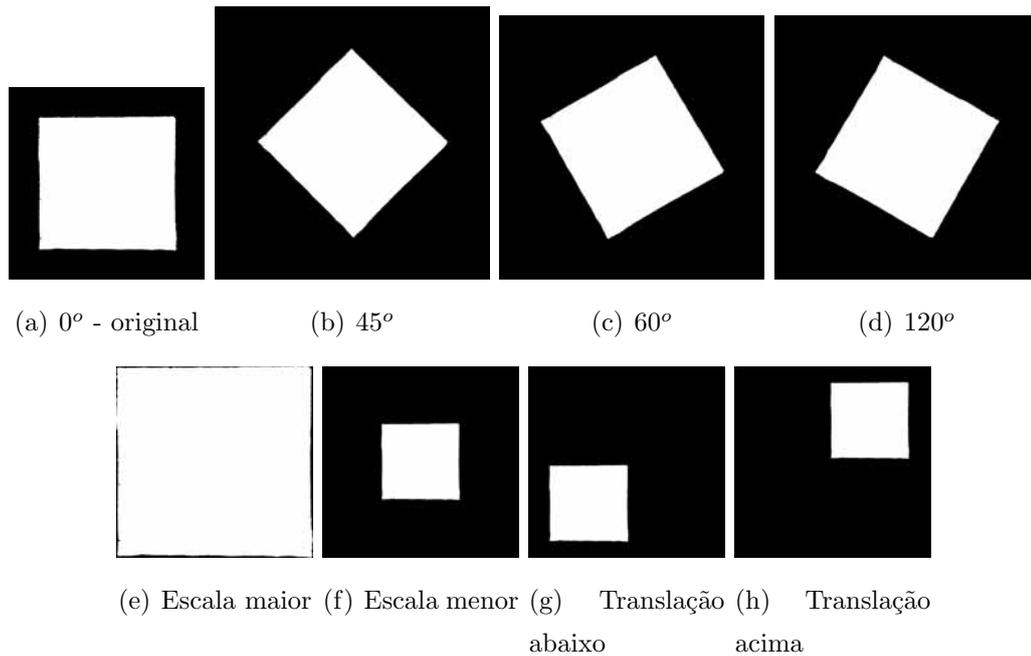


Figura 32: Esboços de contornos de edifício em forma de quadrado.

Os momentos de Zernike com ordem máxima 25 foram calculados para todas as imagens e a norma de cada um dos vetores de características de Zernike pode ser visto nas Tabelas 1, 2 e 3, onde também pode ser encontrada a distância euclidiana entre os vetores dos esboços de referência e suas transformações. Além disso, o limiar utilizado para determinar se a imagem corresponde a um contorno de edifício foi de 0,1 e o valor de  $\beta$  foi fixado como 25000.

Para melhor visualização dos dados, nas Figuras 33, 34 e 35 podem ser vistos gráficos com as normas dos vetores de características de Zernike de cada esboço de contorno de edifício.

Tabela 1: Normas e distâncias das transformações do edifício em forma de L

<b>Edifício em forma de L</b>	<b>Norma</b>	<b>Distância</b>
$0^\circ$	0,6751	-
$45^\circ$	0,6744	0,0164
$90^\circ$	0,6753	0,0142
$120^\circ$	0,6749	0,0227
Escala maior	0,6764	0,0099
Escala menor	0,6748	0,0067
Translação abaixo	0,6756	0,0109
Translação acima	0,6747	0,0098

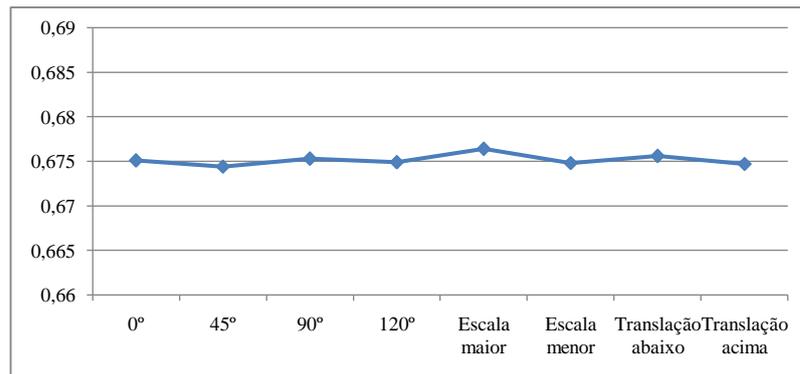


Figura 33: Gráfico das normas das transformações do edifício em forma de L.

Tabela 2: Normas e distâncias das transformações do edifício retangular

Edifício retangular	Norma	Distância
0°	0,6887	-
45°	0,6883	0,0130
90°	0,6886	0,0108
120°	0,6884	0,0199
Escala maior	0,6900	0,0104
Escala menor	0,6882	0,0158
Translação abaixo	0,6879	0,0125
Translação acima	0,6873	0,0152

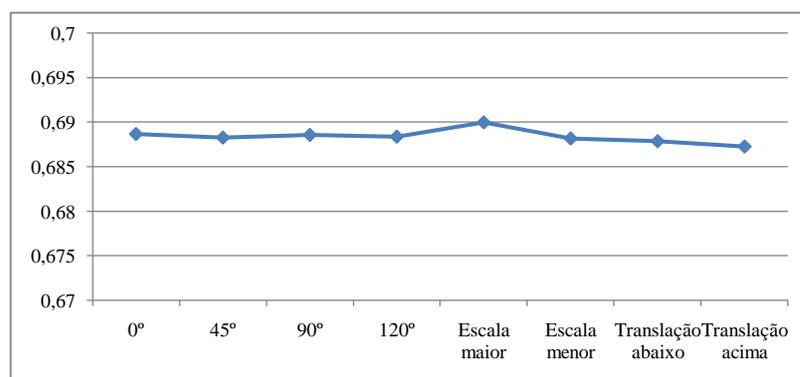


Figura 34: Gráfico das normas das transformações do edifício retangular.

Através das tabelas e gráficos pode-se perceber que a norma dos vetores de características de Zernike varia muito pouco (a partir da terceira casa decimal, com exceção apenas para o edifício de forma retangular com expansão da escala) com as transformações

Tabela 3: Normas e distâncias das transformações do edifício em forma de quadrado

<b>Edifício em forma de Quadrado</b>	<b>Norma</b>	<b>Distância</b>
0°	0,7332	-
45°	0,7328	0,0072
60°	0,7331	0,0094
120°	0,7331	0,0092
Escala maior	0,7335	0,0061
Escala menor	0,7328	0,0068
Translação abaixo	0,7315	0,0081
Translação acima	0,7327	0,0073

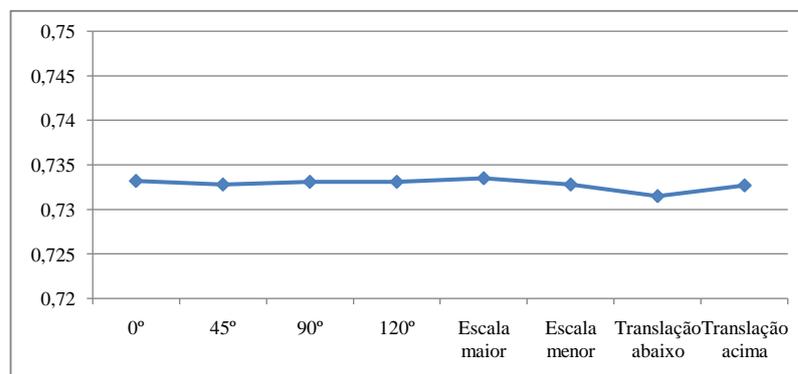


Figura 35: Gráfico das normas das transformações do edifício em forma de quadrado.

de rotação, escala e translação e, ainda, na Tabela 4 é possível observar quais são a média e o desvio-padrão das normas dos vetores de cada tipo de contorno de edifício.

Além disso, as distâncias entre os vetores dos edifícios de referência e suas transformações variaram, em média, apenas 0,012, sendo que o limiar usado para determinar se a região é correspondente a um edifício ou a outro objeto foi de 0,1.

Tabela 4: Média e Desvio-padrão das normas dos vetores de características

	<b>Média</b>	<b>Desvio-padrão</b>
<b>Edifício em forma de L</b>	0,6752	0,00059
<b>Edifício Retangular</b>	0,6884	0,00072
<b>Edifício em forma de Quadrado</b>	0,7328	0,00056

O desvio-padrão mostra o quanto de variação existe em relação à média (ou valor esperado), sendo que seu valor mínimo é 0 e indica que não há variabilidade. Desse modo, um baixo desvio-padrão indica que os dados tendem a estar próximos da média, o que acontece com os desvios-padrão apresentados na Tabela 4. Isto mostra que as normas dos vetores de características realmente variam muito pouco à medida que são feitas transformações nos edifícios.

Todas essas informações mostram que os vetores de características de cada tipo de esboço de contorno de telhado estão muito próximos um do outro, mesmo com as transformações de rotação, escala e translação. Mostrando, assim, a invariância dos momentos de Zernike com relação a essas transformações, alcançando-se o que era esperado.

Essas invariâncias, como dito anteriormente, são bastante atrativas para a utilização dos momentos de Zernike no reconhecimento dos contornos de edifícios. No entanto, vale lembrar que a invariância acontece apenas na transformação isogonal (similaridade), ou seja, o fator de escala é uniforme.

### 4.3 Cena Simulada

Este experimento consiste em simular uma cena com esboços de contornos de edifícios, distribuindo nela os esboços encontrados no banco de dados e os esboços desenhados manualmente, variando suas rotações e escalas. Também foram inseridas algumas formas que visam imitar a presença de vegetação na imagem. A cena produzida pode ser vista na Figura 36.



Figura 36: Cena simulada com esboços de edifícios.

Como a cena que foi produzida simula um caso ideal, onde não há influência das diferentes quedas d'água dos telhados e de outros objetos, como calçadas, árvores e ruas, adjacentes aos edifícios, o método de segmentação segmentou corretamente todas as regiões com o limiar 1 para a diferença entre os níveis de cinza que determina o crescimento das regiões sementes. A Figura 37 mostra a segmentação da imagem da cena simulada, em que as diferentes cores usadas servem para diferenciar uma região da outra e cada região foi numerada para facilitar a análise dos resultados obtidos.

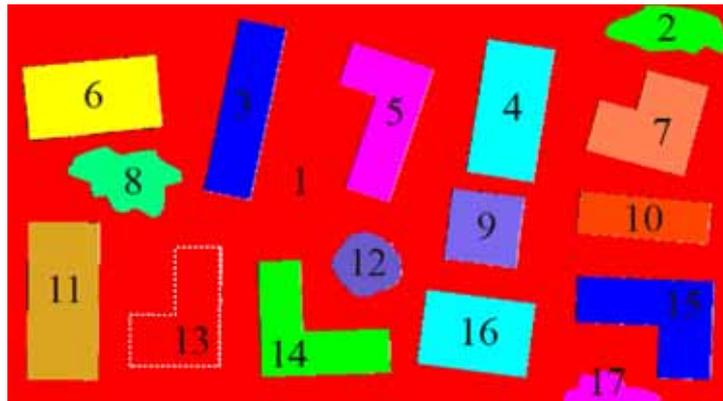


Figura 37: Segmentação da cena simulada.

As regiões segmentadas foram identificadas corretamente pelo método com a ordem máxima 25 para os momentos de Zernike,  $\beta = 25000$  e limiar de 0,09 para a distância entre os vetores de características das regiões encontradas pela segmentação e os vetores dos esboços contidos no banco de dados.

É possível visualizar na Tabela 5 quais regiões foram identificadas como forma de contorno de edifício, qual o tipo dessa forma e qual região foi identificada como não correspondente a um edifício. Isto possibilitou a obtenção do valor ótimo na classificação de todos os tipos de forma de contorno de edifício (REE=100%) e todos os esboços contidos na cena simulada foram identificados, isto é, a acurácia de detecção de edifícios também foi de 100%.

Tabela 5: Reconhecimento das regiões da cena simulada

	<b>Região</b>
<b>Edifício em forma de L</b>	5, 7, 13, 14, 15
<b>Edifício Retangular</b>	3, 4, 6, 10, 11, 16
<b>Edifício em forma de Quadrado</b>	9
<b>Não correspondente a edifício</b>	1, 2, 8, 12, 17

Portanto, num caso ideal, em que as regiões dos contornos de telhado de edifício são segmentadas corretamente, pois não há a influência de outros objetos que interfiram nos contornos de telhado, o método de extração baseado nos momentos de Zernike obteve 100% de sucesso no reconhecimento dos edifícios.

## 4.4 Cenas Reais

Foram feitos três experimentos com dados reais e assim como no experimento 2, as diferentes cores são utilizadas no processo de segmentação para diferenciar uma região da outra e cada região foi numerada.

É importante ressaltar que algumas partes da imagem não foram segmentadas, não fazendo parte de nenhuma região. Isso aconteceu por causa da restrição de área mínima definida como 480 pixels para que a região seja considerada como uma possível edificação. Desse modo, regiões pequenas, menores que 480 pixels, não foram retornadas no processo de segmentação.

A base de dados com esboços de possíveis contornos de telhado de edifício contém três tipos de forma: em forma de L, em forma de quadrado e retangular. Lembrando que os momentos de Zernike são invariantes a apenas um fator de escala (similaridade), mas não são invariantes com relação a variações com dois fatores de escala (afinidade). Desse modo, não foi possível atender a todas as variações possíveis nas dimensões dos esboços em forma de L e em forma retangular contidos na base de dados (Figura 22), visto que estas formas são inúmeras. No entanto, procurou-se construir as formas mais relevantes, que foram se destacando de acordo com a realização dos experimentos.

Além disso, nestes experimentos foram fixados os valores da ordem máxima para os momentos de Zernike em 25 e  $\beta = 25000$ , variando-se apenas o valor do limiar para a distância entre os vetores de característica de Zernike calculado a partir de uma região segmentada na imagem e um outro vetor de característica de Zernike do esboço de edifício armazenado na base de dados, que permite rotular a região como edifício e qual a sua forma ou em outro objeto.

### 4.4.1 Recorte 1

A Figura 38a mostra o Recorte 1 e sua transformação em níveis de cinza usando o NandA pode ser vista na Figura 38b. Antes da segmentação da imagem foi necessário aplicar o filtro da Média com máscara  $3 \times 3$ , com o intuito de suavizar ruídos ou algumas bordas provenientes das quedas internas do telhado, que podem fazer com que o mesmo telhado seja dividido em duas ou mais regiões. As regiões encontradas podem ser vistas na Figura 39. O limiar entre a diferença dos níveis de cinza dos pixels, que determina o crescimento das regiões, utilizado no método de segmentação foi 3.



Figura 38: Recorte 1 e sua transformação em escala de cinza.

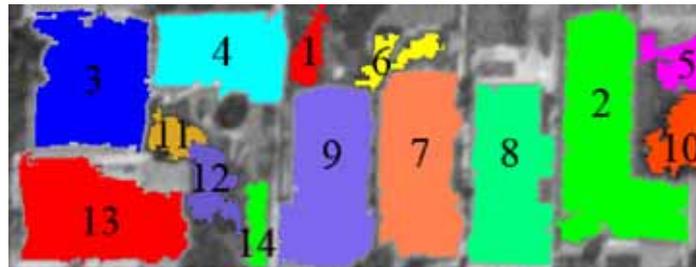


Figura 39: Segmentação do Recorte 1.

A Tabela 6 apresenta o resultado da classificação dos edifícios obtido pelo método baseado nos momentos de Zernike com o limiar para a distância entre os vetores da região segmentada e os vetores dos esboços do banco de dados definido como 0,2.

Tabela 6: Reconhecimento das regiões do Recorte 1

	<b>Região</b>
<b>Edifício em forma de L</b>	2, 10
<b>Edifício Retangular</b>	3, 4, 7, 8, 9
<b>Edifício em forma de Quadrado</b>	
<b>Não correspondente a edifício</b>	1, 5, 6, 11, 12, 13, 14

Analisando esses resultados, pode-se dizer que houve apenas um falso positivo, ou seja, dos 7 edifícios encontrados, tem-se que 6 estão corretos, implicando no valor  $REE=85,7\%$  da razão de extração de edifícios. Como existem 7 edifícios na cena, houve apenas 1 edifício não identificado (região 13), produzindo uma acurácia de detecção de edifícios  $ADE=85,7\%$ . A ocorrência deste falso negativo decorre, provavelmente, da má segmentação da região deste edifício.

Vale lembrar que pequenos detalhes no contorno das regiões não são “percebidos” pelo método devido à robustez dos momentos de Zernike, de maneira que o método reconhece a forma que mais se aproxima da região dada. Isto pode ser observado na região 10, que foi identificada erroneamente pelo método como uma região correspondente a edifício em forma de L.

#### 4.4.2 Recorte 2

Podem ser vistos o Recorte 2 na Figura 40a e sua respectiva transformação em escala de cinza na Figura 40b.

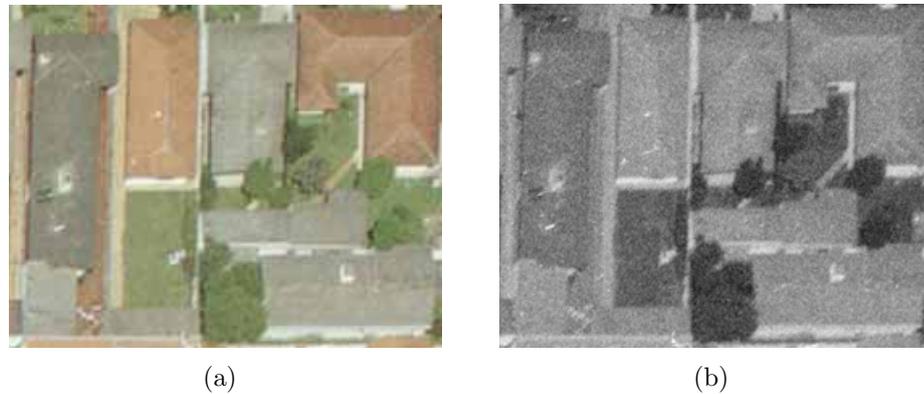


Figura 40: Recorte 2 e sua transformação em escala de cinza.

Neste caso, assim como no Recorte 1, antes da segmentação da imagem foi aplicado o filtro da Média com máscara  $3 \times 3$  e o limiar entre a diferença dos níveis de cinza dos pixels, utilizado no método de crescimento de regiões, também foi 3. O resultado da segmentação pode ser visto na Figura 41.

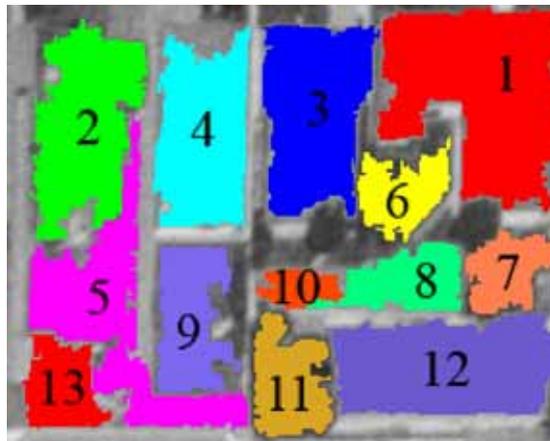


Figura 41: Segmentação do Recorte 2.

Tabela 7: Reconhecimento das regiões do Recorte 2

	Região
Edifício em forma de L	1, 6
Edifício Retangular	2, 3, 4, 12, 13
Edifício em forma de Quadrado	
Não correspondente a edifício	5, 7, 8, 9, 10, 11

Na Tabela 7 é possível observar o resultado da classificação dos edifícios obtido pelo método de extração proposto utilizando o limiar 0,21 para a distância entre os vetores da região segmentada e os vetores dos esboços do banco de dados. Com base nesta tabela, percebe-se que 6 edifícios encontrados estão corretos, havendo 1 falso positivo (região 6), obtendo-se a  $REE=85,7\%$ . Sendo que a região 2 foi identificada corretamente pelo método, porém, esta região não corresponde ao edifício por completo.

Além disso, existem 7 edifícios na cena, havendo apenas 1 falso negativo e produzindo  $ADE=85,7\%$ . O edifício não identificado corresponde à região 8, onde pode se dizer que o método falhou devido à errônea segmentação. Esta segmentação ruim decorre, provavelmente, do solo exposto ao redor do edifício, que foi segmentado na mesma região juntamente com o edifício.

#### 4.4.3 Recorte 3

Primeiramente, como nos casos anteriores, o Recorte 3 (Figura 42a) foi transformado em níveis de cinza e seu resultado pode ser visualizado na Figura 42b.

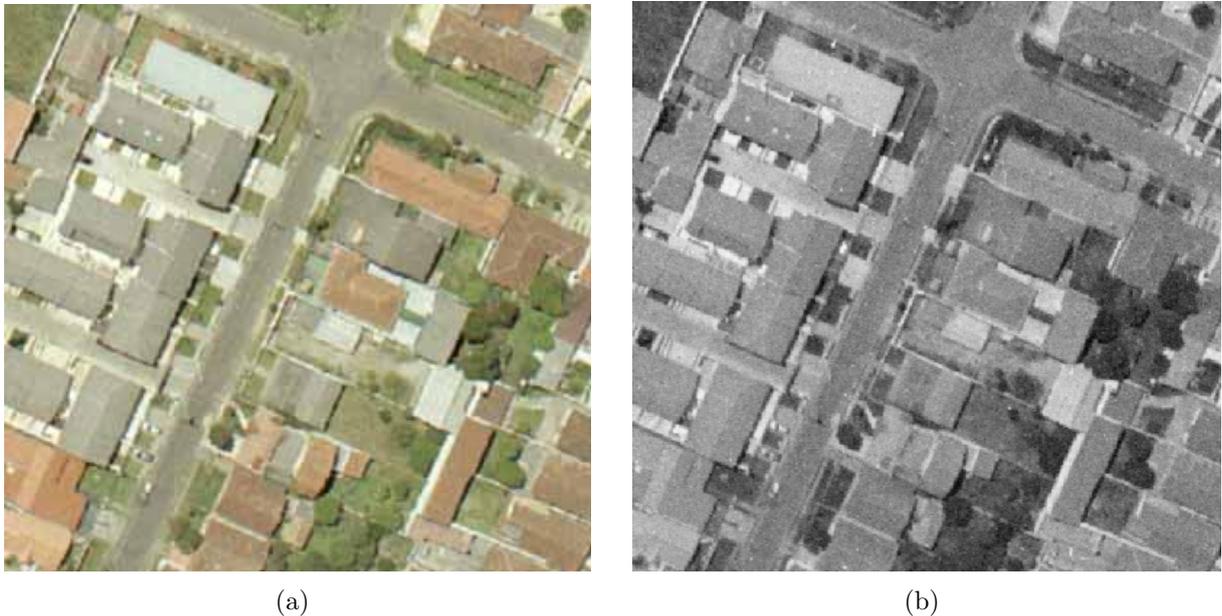


Figura 42: Recorte 3 e sua transformação em escala de cinza.

Assim como nos casos anteriores, novamente, foi aplicado o filtro da Média de máscara  $3 \times 3$  na imagem para realizar a segmentação desse recorte, que pode ser vista na Figura 43. O limiar 3 foi utilizado no método de crescimento de regiões para a segmentação.

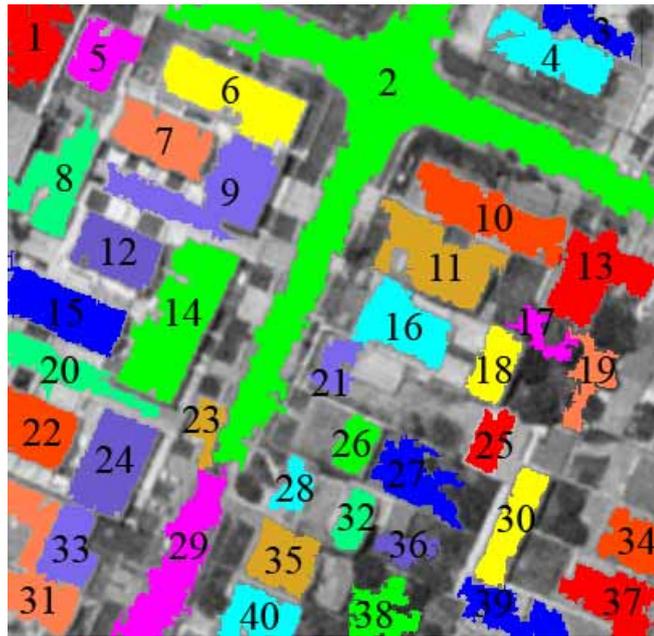


Figura 43: Segmentação do Recorte 3.

Para este experimento, o limiar utilizado para comparação entre os vetores de características de Zernike foi de 0,2. Na Tabela 8 estão os resultados encontrados pelo método de reconhecimento de edifícios, onde pode-se observar quais as regiões do recorte foram identificadas como forma de edifício, qual o tipo de forma e quais regiões não correspondem a edifício.

Tabela 8: Reconhecimento das regiões do Recorte 3 com segmentação automática

	Região
<b>Edifício em forma de L</b>	5, 13, 33
<b>Edifício Retangular</b>	4, 6, 7, 10, 12, 14, 15, 18, 24, 25, 26, 30, 32
<b>Edifício em forma de Quadrado</b>	35
<b>Não correspondente a edifício</b>	1, 2, 3, 8, 9, 11, 16, 17, 19, 20, 21, 22, 23, 27, 28, 29, 31, 34, 36, 37, 38, 39, 40

Para o recorte 3, a razão de extração de edifícios obteve valor ótimo 100%, ou seja, todos os 17 edifícios encontrados estão corretos. Por consequência, não houve falsos positivos. Sendo que as regiões 26 e 33 correspondem a edifícios, porém, eles não foram segmentados por inteiro por essas regiões.

Foram considerados 32 edifícios existentes no recorte e como 17 edifícios foram detectados corretamente, logo, a acurácia de detecção de edifícios foi de 53,1%.

Com o objetivo de ressaltar a eficiência do método, mostrando que algumas falhas são oriundas do processo automático de segmentação, este recorte foi analisado também

de outra maneira. Com o auxílio do aplicativo Photoshop, os contornos de edifício da imagem foram delimitados manualmente e preenchidos. O resultado disto pode ser visto na Figura 44a.

Para maior complexidade e maior relação com a realidade, o recorte com os contornos delimitados manualmente passou pelos mesmos procedimentos que o recorte original. A imagem foi filtrada com o filtro da Média de máscara  $3 \times 3$  e segmentada (com os mesmos parâmetros para o método de segmentação utilizado no caso anterior) e analisado novamente. Na Figura 44b pode-se observar o resultado dessa segmentação, onde as regiões com \* não foram consideradas como edifício, pois suas áreas eram menores que o valor mínimo estipulado.

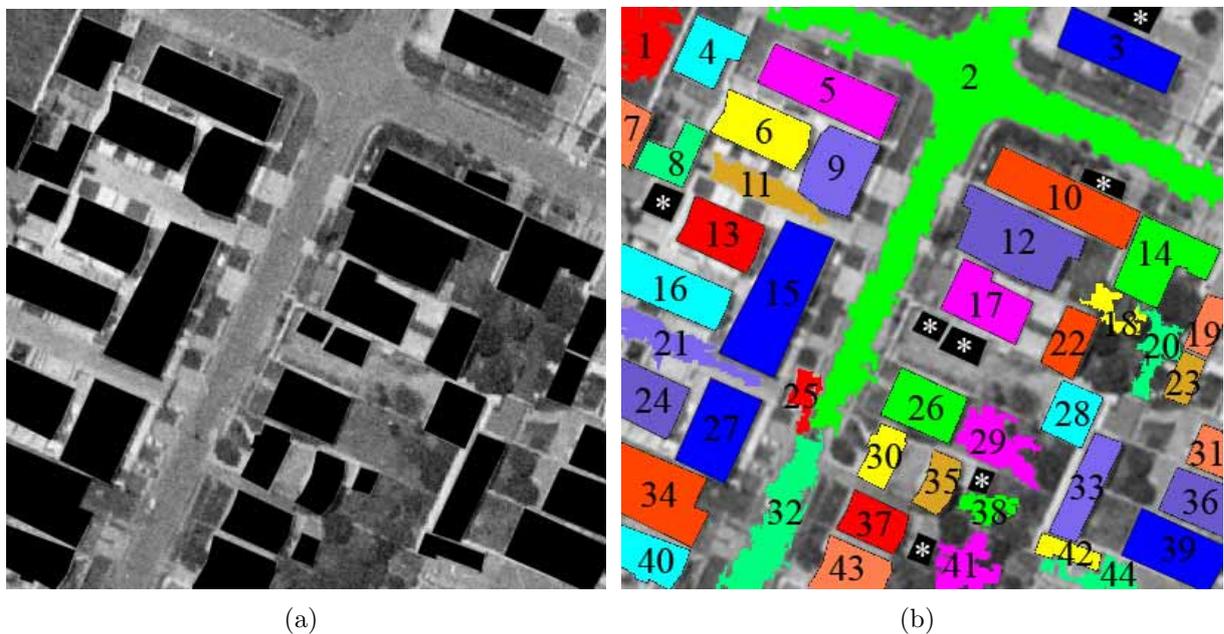


Figura 44: Recorte 3 com contornos de edifícios delimitados manualmente e sua segmentação.

Na Tabela 9 é possível visualizar quais regiões foram identificadas pelo método como correspondentes a edifício e qual sua forma e quais regiões não correspondem a edifícios.

Tabela 9: Reconhecimento das regiões do Recorte 3 com os contornos de edifícios delimitados manualmente

	Região
<b>Edifício em forma de L</b>	4, 8, 14, 31, 43
<b>Edifício Retangular</b>	3, 5, 6, 9, 10, 12, 13, 15, 16, 17, 19, 22, 23, 26, 27, 28, 30, 33, 34, 35, 36, 37, 39, 40, 42
<b>Edifício em forma de Quadrado</b>	
<b>Não correspondente a edifício</b>	1, 2, 7, 11, 18, 20, 21, 24, 25, 29, 32, 38, 41, 44

No recorte 3 com os contornos dos telhados de edifício delimitados manualmente, o limiar utilizado no método dos momentos de Zernike foi 0,2 e obteve-se 93,3% (28/30) de edifícios identificados corretamente (REE), um índice bem maior que no caso em que a segmentação foi feita automaticamente pelo aplicativo. Assim, houve apenas 2 falsos positivos (6,7%), que correspondem às regiões 31 e 43, sendo que elas correspondem a edifícios em forma retangular, mas foram identificadas como em forma de L, provavelmente por estarem na borda da imagem e estarem com sua forma incompleta, o que dificultou a correta classificação.

Além disso, também melhorando o índice anterior, os falsos negativos correspondem a apenas 12,5% (4/32), sendo que esses contornos correspondem às regiões 7, 24, 31 e 43, que se encontram incompletos nas bordas da imagem, o que dificulta seu reconhecimento, pois não estão com sua forma completa. Dessa maneira, a ADE também obteve um melhor percentual de 87,5%.

Desse modo, pode-se dizer que delimitando-se manualmente os contornos de edifícios foi possível alcançar um resultado bem mais satisfatório comparado com o caso de se segmentar a imagem automaticamente. Isto mostra a necessidade de melhorar o processo de segmentação, pois este é um passo decisivo no método proposto de classificação de regiões baseado em sua forma.

#### 4.4.4 Comparação com Outros Métodos

Mahi, Isabaten e Serief (2014) apresentam uma média de 83,5% de acurácia da detecção de edifícios (razão entre o número de edifícios classificados corretamente e o número total de edifícios) do método proposto que também utiliza momentos de Zernike. Os autores comparam este resultado com o desempenho da ferramenta de extração de características do *software* ENVI (*Environment for Visualizing Images*), que teve uma acurácia de classificação de 79%. Enquanto que em Pakizeh e Palhang (2010) os experimentos tiveram, em média, ADE=77%.

Comparando os resultados produzidos pela metodologia proposta com estes números, é possível notar a eficiência do descritor de forma baseado nos momentos de Zernike apresentado nesta pesquisa, pois conseguiu nos recortes 1 e 2 ADE=85,7%, um índice melhor que os apresentados nos trabalhos citados. Enquanto que no recorte 3, a ADE foi de apenas 53,1%, mas ao delimitar manualmente os contornos dos edifícios, este índice sobe para 87,5%, indicando a forte influência da segmentação da imagem para que o resultado seja mais satisfatório.

Com relação à razão de extração de edifícios, que representa a porcentagem dos edifícios identificados corretamente, pode-se comparar com os resultados produzidos por Rütther, Martine e Mitalo (2002). Neste trabalho é apresentado um método de extração de edifícios em áreas de assentamentos informais, através da estratégia de delinear edifícios pela técnica de otimização de programação dinâmica, baseada na posição dos contornos de edifício aproximados obtidos por um Modelo Digital de Superfície (MDS). E então, a extração do edifício é feita por meio do algoritmo de *snakes*.

Novamente, o método proposto mostrou índices superiores, pois nos recortes 1 e 2 obteve-se REE=85,7% e no recorte 3 com a segmentação automática a REE foi de 100% e REE=93,3% no caso da delimitação manual dos contornos de edifício antes da segmentação. Enquanto que em Rütther, Martine e Mitalo (2002) a média da razão de extração de edifícios de seus experimentos foi de 62,3% e em Pakizeh e Palhang (2010) a média foi de 79,8%.

Dessa maneira, pode-se notar o bom desempenho do método proposto de extração de edifícios baseado nos momentos de Zernike. No entanto, vale ressaltar que as comparações entre os diferentes métodos foram feitas de acordo com o resultado de cada trabalho, ou seja, as imagens utilizadas não foram as mesmas.

## 5 CONCLUSÃO

Neste trabalho foi apresentado um método baseado nos momentos de Zernike para detecção de edifícios em imagens de alta resolução e realizados alguns testes experimentais para avaliar os resultados obtidos. Para isso, foi necessário um estudo dos fundamentos de processamento digital de imagens, princípio básico essencial para compreensão dos estudos posteriores. E então foram realizados outros estudos sobre segmentação de imagens e momentos de Zernike, que são baseados nos polinômios ortogonais de Zernike.

O método consiste em criar um vetor de características de Zernike para cada região determinada num processo de segmentação da imagem e então compará-lo com vetores de características de esboços de contornos de edifícios armazenados num banco de dados e através de um limiar determinar se essa região corresponde ou não à forma de um edifício.

A utilização deste método é atrativa pelo fato dos momentos de Zernike serem invariantes às transformações de rotação, translação e escala (similaridade), cuja efetividade foi vista nos experimentos. Os testes serviram também para demonstrar que o método mostrou-se ótimo numa cena ideal e conseguiu atingir bons índices de qualidade para o reconhecimento de edifícios em cenas reais.

Pode-se observar também, no experimento 3, que em cenas com poucos edifícios (recortes 1 e 2), onde foi possível um melhor resultado no processo de segmentação, o desempenho do método proposto foi mais satisfatório do que no caso do recorte 3, onde existe uma quantidade maior de edifícios presentes na cena, como também de outros objetos e, por consequência, tudo isso dificulta uma boa segmentação das regiões.

Todos estes resultados podem ser vistos e analisados na Tabela 10.

Assim, é possível dizer que falhas foram geradas devido, principalmente, ao processo de segmentação da imagem, visto que regiões correspondentes a edifícios foram segmentadas juntamente com árvores, calçadas ou até mesmo outro edifício que estavam ao seu redor. Desse modo, as formas dessas regiões não são compatíveis com os modelos de edifícios presentes na base de dados, o que também acontece com edifícios que estão nas bordas

Tabela 10: Resultados dos experimentos

	Cena simulada	Cenas reais		
		Recortes 1 e 2	Recorte 3 (segm. automática)	Recorte 3 (contornos del. manual.)
<b>REE</b>	100%	85,7%	100%	93,3%
<b>Falsos Positivos</b>	-	14,3%	-	6,7%
<b>ADE</b>	100%	85,7%	53,1%	87,5%
<b>Falsos Negativos</b>	-	14,3%	46,9%	12,5%

da imagem. Além disso, algumas regiões que não correspondem a edifícios, como ruas, calçadas e árvores, foram segmentadas de maneira que sua forma fosse semelhante a de um edifício, ou então devido às inclinações do telhado, o mesmo edifício foi segmentado em mais de uma região.

Neste sentido, pode-se afirmar que a segmentação de imagens em métodos voltados para a rotulação de regiões, que é o caso desta pesquisa, é uma deficiência difícil de ser resolvida. Uma alternativa é utilizar o Índice de Vegetação por Diferença Normalizada (NDVI - *Normalized Difference Vegetation Index*), quando houver a existência de informações do infravermelho das imagens, onde é possível eliminar mais facilmente a vegetação presente na imagem. E ainda, o estudo de outras técnicas de segmentação incentivam possíveis trabalhos futuros. Para isso, pode-se utilizar também a representação HSI da imagem, com eliminação da componente I para minimizar a influência da diferença de iluminação nas diferentes faces de um dado telhado.

Outro ponto notado pelos experimentos realizados foi o fato de que o edifício em forma de L foi o que proporcionou maior quantidade de erros, provavelmente devido à sua forma não ser tão regular como a retangular e a quadrada, além de ter mais possibilidades nas variações em suas dimensões. A busca de uma maneira para resolver este problema estimula futuros trabalhos, que podem inclusive levar em consideração outros tipos de forma de edifício.

Para isso, pode-se modelar as formas de telhado com a transformação afim, em que existem dois fatores de escala (um para o eixo x e outro para o eixo y). Dessa maneira, é possível considerar apenas a forma elementar retangular no banco de dados e realizar a sub-segmentação de regiões na imagem, a fim de se obter regiões em forma de arranjos de retângulos. Isto se mostra um desafio interessante para trabalhos futuros.

Além disso, também pode se realizar um estudo para a regularização dos contornos das regiões segmentadas para que não hajam tantas irregularidades. Desse modo, as formas das regiões segmentadas seriam mais semelhantes aos esboços do banco de dados e o limiar para a distância entre seus vetores de características poderia ser menor, podendo eliminar alguns falsos positivos. Além disso, uma menor ordem máxima para os momentos de Zernike ainda conseguiria identificar corretamente os objetos, diminuindo o esforço computacional.

Futuramente, também é possível estudar a adaptação automática da ordem ótima dos momentos de Zernike e do limiar da distância entre os vetores de características da região de interesse e dos esboços contidos no banco de dados, que permite discriminar as diferentes formas de edifício e não edifício.

## *Referências*

- BHADAURIA, A.; BHADAURIA, H.; KUMAR, A. Building extraction from satellite images. *IOSR Journal of Computer Engineering*, v. 12, n. 2, p. 76–81, 2013.
- BISNETO, O. S. et al. Aberrações de alta ordem: associação com a idade e erros de refração. *Arquivos Brasileiros de Oftalmologia*, v. 70, n. 2, p. 290–297, 2007.
- BOUZIANI, M.; GOITA, K.; HE, D.-C. Rule-based classification of a very high resolution image in an urban environment using multispectral segmentation guided by cartographic data. *IEEE Trans. Geosci. Remote Sens.*, v. 48, n. 8, p. 3198–3211, 2010.
- BRUZZONE, L.; CARLIN, L. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Trans. Geosci. Remote Sens.*, v. 44, n. 9, p. 2587–2600, 2006.
- CELEBI, M. E.; ASLANDOGAN, Y. A. A comparative study of three moment-based shape descriptors. *Proceedings of the International Conference on Information Technology: Coding and Computing*, v. 1, p. 788–793, 2005.
- CHEN, L. et al. Building detection in an urban area using lidar data and QuickBird imagery. *International Journal of Remote Sensing*, v. 33, n. 16, p. 5135–5148, 2012.
- CHONG, C.-W.; RAVEENDRAN, P.; MUKUNDAN, R. A comparative analysis of algorithms for fast computation of Zernike moments. *Pattern Recognition*, v. 36, p. 731–742, 2003.
- CRÓSTA, A. P. *Processamento Digital de Imagens de Sensoriamento Remoto*. Campinas: UNICAMP, 1992.
- DUNKL, C. F.; XU, Y. *Orthogonal Polynomials of Several Variables*. Cambridge: Cambridge University Press, 2001.
- FELIPE, J. C.; OLIOTI, J. B.; TRAINA, A. J. M. Discriminação de aspectos malignos em massas tumorais de mamografias usando características de forma das imagens. *in V Workshop de Informática Médica - Porto Alegre, RS*, 2005.
- FLUSSER, J.; SUK, T.; ZITOVÁ, B. *Moments and Moment Invariants in Pattern Recognition*. Torquay, UK: John Wiley & Sons Inc., 2009.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. New Jersey: Pearson Prentice Hall, 2008.
- HOFMANN, P. Detecting urban features from IKONOS data using an object-oriented approach. *In: GRSS/ISPRS Joint Workshop Urban*, 2003.

- HWANG, S.-K.; KI, W.-Y. A novel approach to the fast computation of Zernike moments. *Pattern Recognition*, v. 39, p. 2065–2076, 2006.
- JAIN, R.; KASTURI, R.; SCHUNCK, B. G. *Machine vision*. New York: McGraw-Hill, 1995.
- JANKOV, M. et al. Frentes de ondas (wavefronts) e limites da visão humana. parte 1 - fundamentos. *Arquivos Brasileiros de Oftalmologia*, v. 65, n. 6, p. 679–684, 2002.
- JÄHNE, B. *Digital Image Processing*. Germany: Springer, 2002.
- KHOTANZAD, A.; HONG, Y. H. Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 12, n. 5, p. 489–497, 1990.
- KIM, H. S.; LEE, H.-K. Invariant image watermark using Zernike moments. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 13, n. 8, p. 766–775, 2003.
- KIM, W.-Y.; KIM, Y.-S. A region-based shape descriptor using Zernike moments. *Signal Processing: Image Communication. Elsevier Science B.V.*, v. 16, p. 95–102, 2000.
- LI, S.; LEE, M.-C.; PUN, C.-M. Complex Zernike moments features for shape-based image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, v. 39, n. 1, p. 227–237, 2009.
- LIAO, S. X. *Image Analysis by Moments*. Tese — Department of Electrical and Computer Engineering, University of Manitoba, Canadá, 1993.
- MAHI, H.; ISABATEN, H.; SERIEF, C. Zernike moments and SVM for shape classification in very high resolution satellite images. *The International Arab Journal of Information Technology*, v. 11, n. 1, p. 43–51, 2014.
- MARCATO, V. J. *Reconstrução de contornos e cumeeiras de telhados de edifícios a partir de imagens aéreas de alta-resolução e poliedros representativos de edifícios extraídos de dados LASER*. Dissertação (Mestrado em Ciências Cartográficas) — Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista, Presidente Prudente, 2013.
- NIXON, M. S.; AGUADO, A. S. *Feature Extraction and Image Processing*. Second. UK: Elsevier Ltd., 2008.
- PAKIZEH, E.; PALHANG, M. Building detection from aerial images using Hough transform and intensity information. *Proceedings of 18th Iranian Conference on Electrical Engineering (ICEE)*, p. 532–537, 2010.
- PEDRINI, H.; SCHWARTZ, W. R. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. São Paulo: Thomson Learning, 2008.
- PRATA, A. J.; RUSCH, W. V. T. Algorithm for computation of Zernike polynomials expansion coefficients. *Journal of the Optical Society of America*, v. 28, n. 7, p. 749–754, 1989.
- RÜTHER, H.; MARTINE, H. M.; MTALO, E. Application of snakes and dynamic programming optimisation technique in modeling of buildings in informal settlement areas. *ISPRS Journal of Photogrammetry & Remote Sensing*, v. 56, p. 269–282, 2002.

SHACKELFORD, A. K.; DAVIS, C. H. A combined fuzzy pixel-based and object-based approach for classification of high resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sens.*, v. 41, n. 10, p. 2354–2363, 2003.

SINGH, D. et al. Building extraction from very high resolution multispectral images using NDVI based segmentation and morphological operators. *Proceedings of Students Conference on Engineering and Systems (SCES)*, p. 1–5, 2012.

SOWMYA, A.; TRINDER, J. Modelling and representation issues in automated feature extraction from aerial and satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 55, n. 1, p. 34–47, 2000.

TAHMASBI, A. et al. A novel breast mass diagnosis system based on Zernike moments as shape and density descriptors. in *Proc. IEEE, 18th Iranian Conference on Biomedical Engineering*, p. 100–104, 2011.

TAHMASBI, A.; SAKI, F.; SHOKOUHI, S. B. Classification of benign and malignant masses based on Zernike moments. *J. Computers in Biology and Medicine*, v. 41, n. 8, p. 726–735, 2011.

TEAGUE, M. R. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, v. 70, n. 8, p. 920–930, 1980.

THOMAZ, D. V. *Extração de contornos de telhados usando princípios de snake balloon*. Dissertação (Mestrado em Matemática Aplicada e Computacional) — Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista, Presidente Prudente, 2012.

VALE, G. M. do. *Uso de técnicas de otimização e agrupamento hierárquico para a complementação automática da malha viária em regiões rurais a partir de imagens aéreas de média e alta resolução*. Tese (Doutorado em Ciências Cartográficas) — Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista, Presidente Prudente, 2007.