

Paulo César Cortez

Reconhecimento de Formas 2D Usando uma  
Técnica Seqüencial Integrada e Modelos  
Poligonais

Tese submetida ao corpo docente da Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal da Paraíba - Campus II como parte dos requisitos necessários para obtenção do grau de Doutor em Engenharia Elétrica.

João Marques de Carvalho - Ph. D.  
(Orientador)

Campina Grande, Paraíba, Brasil

©Paulo César Cortez, 1996



C827r

Cortez, Paulo Cesar

Reconhecimento de formas 2D usando uma tecnica sequencial integrada e modelos poligonais / Pulo Cesar Cortez. - Campina Grande, 1996.

122 f.

Tese (Doutorado em Engenharia Eletrica) - Universidade Federal da Paraiba, Centro de Ciencias e Tecnologia.


1. Formas 2D - 2. Formas 2D - 3. Reconhecimento de Formas em 2D - 4. Tese I. Carvalho, Joao Marques de, Dr. II. Universidade Federal da Paraiba - Campina Grande (PB) III. Título


CDU 621.3(043)


**RECONHECIMENTO DE FORMAS 2D USANDO UMA TÉCNICA SEQUENCIAL  
INTEGRADA E MODELOS POLIGONAIS**

**PAULO CÉSAR CORTEZ**


Tese Aprovada em 15.07.1996


  
**PROF. JOÃO MARQUES DE CARVALHO, Ph.D., UFPB**  
Orientador

  
**PROF. ARNALDO DE ALBUQUERQUE ARAÚJO, D.Sc., UFMG**  
Componente da Banca

  
**PROF. JACQUES FACON, Dr.Ing., CEFET-PR**  
Componente da Banca

**PROF. GUILHERME VILAR, Dr.Ing., UFPB**  
Componente da Banca

  
**PROF. MISAEL ELIAS DE MORAIS, Dr.-Ing.**  
Componente da Banca

  
**PROF. LUIZ PEREIRA CALÔBA, Dr.-Ing., COPPE/UFRJ**  
Componente da Banca

CAMPINA GRANDE - PB

Julho - 1996

## Agradecimentos

A execução desta tese somente foi possível graças ao apoio, à dedicação, ao esforço, ao estímulo e ao profissionalismo do Prof. João Marques de Carvalho que me acompanhou em todos os momentos desta. Agradeço-o por tudo isto.

Agradeço o apoio e o estímulo oferecidos por minha companheira Sueli, em muitos momentos difíceis por que passei. Também agradeço a sua mãe Guilhermina que muito me apoiou.

Meus agradecimentos aos colegas do DEE/UFC, em particular: João César, Oscar Brasil, Pimentel, Fátima, Elano, Mário Fialhos e Alexandre Moreira.

Devo sinceros agradecimentos a todos os colegas do LAPS, em particular, aos colegas Joseana, Rosângela, Robson, Mônica, Moreira, Washington, Silvana Cunha, Madeiro, Aldenor, Fátima, Gustavo, Eliane, entre outros.

Agradeço a todos aqueles que contribuíram direta ou indiretamente para a elaboração desta tese, particularmente aos colegas: Rinaldo, Wallington, Paulo Márcio, Joabson, Bruno, Natanael, Albos, José Ferreira, José Roberto, Ângela, Lucimar, Luis Carlos. Agradeço o carinho a mim oferecido por todas estas e outras pessoas, cujos nomes não constam desta minúscula lista.

Agradeço ao amigo Giovanni pelo irrestrito apoio deste o início deste trabalho até o seu final. Devo também meus sinceros agradecimentos ao amigo Macário pelo incentivo dado em todos os momentos.

O apoio material oferecido pela CAPES/PICD e COPELE para a realização deste trabalho foram indispensáveis, portanto, não poderia deixar de agradecê-los.

Finalmente, agradeço o carinho, a amizade e a hospitalidade oferecidos pelos campinenses e pela cidade de Campina Grande que, com certeza, permanecerão comigo.

Dedico esta tese ao meu pai Miguel Cortez, a minha mãe Iolanda Cortez, aos meus filhos Aresqui e Bartira e a todos os trabalhadores conscientes do meu país que lutam por uma sociedade justa e igualitária.

“Não basta que seja pura e justa a nossa causa.  
É necessário que a pureza e a justiça existam dentro de nós”.

Agostinho Neto

## Resumo

Esta tese apresenta uma nova metodologia para reconhecimento de formas em 2D utilizando modelos poligonais e seus atributos geométricos, bem como a implementação desta metodologia em um sistema de visão artificial. Esta metodologia possibilita tratar os processos de modelagem e reconhecimento de formas como parte de um mesmo problema, através do uso de funções discriminantes. O sistema desenvolvido contém significantes inovações, tanto no processo de modelagem poligonal como no de casamento de modelos, através da geração e verificação de hipóteses. Dentre estas destaca-se a execução das tarefas de maneira seqüencial e integrada, e o uso de funções discriminantes com valor de limiar angular variável, para o agrupamento e verificação de hipóteses. São testadas cinco técnicas diferentes de avaliação de hipóteses, utilizando cinco cenas distintas. Conclui-se que o sistema resultante é capaz de reconhecer e localizar formas 2D isoladas ou parcialmente visíveis, posicionadas aleatoriamente.

## **Abstract**

This thesis presents a new methodology for polygonal model based 2D shape recognition. This methodology enables a unified treatment to the model building and recognition problems, through the use of discriminant functions. An artificial vision system implementing the proposed methodology is also described. Among the relevant innovations presented by the system, one can mention the integration of the sequential procedures required and the use of discriminant functions with variable angular thresholds, for hypothesis grouping and validation. Five different hypothesis evaluation techniques have been tested, for five distinct test scenes. The resulting system is able to deal with 2D shapes, isolated or superposed (partially visible) and randomly positioned.

# Sumário

<b>1</b>	<b>Caracterização do problema</b>	<b>1</b>
1.1	Introdução . . . . .	1
1.2	Aspectos gerais de um sistema de visão . . . . .	4
1.3	O problema . . . . .	6
1.4	Objetivo . . . . .	8
<b>2</b>	<b>Reconhecimento de formas 2D: revisão bibliográfica e proposta de nova metodologia</b>	<b>9</b>
2.1	Introdução . . . . .	9
2.2	Classificação de sistemas para reconhecimento de objetos . . . . .	10
2.2.1	Classificação segundo Suetens et al. . . . .	11
2.2.2	Classificação segundo Chin e Dyer . . . . .	12
2.3	Sistemas baseados em modelos . . . . .	12
2.3.1	Métodos dos atributos globais . . . . .	13
2.3.2	Métodos de atributos estruturais . . . . .	17
2.3.3	Métodos dos grafos relacionais . . . . .	21
2.4	O método proposto . . . . .	26



<b>3</b>	<b>Modelagem poligonal de contornos bidimensionais</b>	<b>29</b>
3.1	Introdução . . . . .	29
3.2	Pré-processamento . . . . .	31
3.3	Segmentação . . . . .	34
3.4	Modelagem poligonal . . . . .	41
3.5	Algoritmos para a modelagem poligonal de contornos 2D . . . . .	43
3.5.1	Algoritmo de Ramer . . . . .	44
3.5.2	Algoritmo Split-Merge . . . . .	44
3.5.3	Algoritmo de Cortez . . . . .	45
3.5.4	Algoritmo de Ray-Ray . . . . .	50
3.5.5	Algoritmo de Cheng-Hsu . . . . .	51
3.5.6	Algoritmo que utiliza a transformada de Hough . . . . .	52
3.5.7	Algoritmo de Kung-Kaswand . . . . .	52
3.6	Resultados experimentais . . . . .	53
3.7	Análise dos resultados . . . . .	57
3.7.1	Quanto ao tempo de processamento . . . . .	62
3.7.2	Quanto ao número de lados dos modelos . . . . .	62
3.7.3	Quanto à posição dos vértices nos modelos . . . . .	63
3.8	Conclusões . . . . .	63
<b>4</b>	<b>Reconhecimento de formas 2D usando modelos poligonais</b>	<b>64</b>
4.1	Metodologia proposta . . . . .	65
4.2	Geração das hipóteses . . . . .	67
4.2.1	Controle do número de hipóteses . . . . .	69
4.3	Avaliação das hipóteses . . . . .	73

4.3.1	Primeiro método de avaliação de hipóteses . . . . .	77
4.3.2	Segundo método de avaliação de hipóteses . . . . .	79
4.4	Localização de formas reconhecidas . . . . .	83
<b>5</b>	<b>Resultados experimentais</b>	<b>84</b>
5.1	Introdução . . . . .	84
5.2	Parâmetros para avaliação de desempenho . . . . .	85
5.3	Cenas a serem analisadas . . . . .	87
5.4	Resultados obtidos pelas primeira, segunda e terceira técnicas . . . . .	90
5.5	Resultados obtidos pelas quarta e quinta técnicas . . . . .	100
5.6	Análise geral dos resultados e conclusões . . . . .	107
<b>6</b>	<b>Conclusões, contribuições e trabalhos futuros</b>	<b>110</b>

# Lista de Tabelas

3.1	Resultados do tempo médio consumido por cada um dos processos que constituem a etapa de pré-processamento. . . . .	33
3.2	Resultados dos algoritmos BTT e BTVE para os contornos das figuras 3.7 e 3.8 . . . . .	39
3.3	Resultados dos algoritmos para o objeto Arco com 126 pontos. . . . .	55
3.4	Resultados dos Algoritmos para o objeto Fenda com 387 pontos. . . . .	56
4.1	Resumo dos dois métodos e suas técnicas. . . . .	82
5.1	Resultados obtidos pela primeira, segunda e terceira técnicas para as cinco cenas. . . . .	99
5.2	Resultados obtidos pela quarta e quinta técnicas para as cinco cenas. . . . .	106

# Lista de Figuras

1.1	Estrutura geral de um sistema de reconhecimento. . . . .	4
1.2	Situação típica de um ambiente industrial. . . . .	7
2.1	Classificação dos sistemas em função das complexidades dos dados e dos modelos. . . . .	11
2.2	Diagrama em blocos de um sistema genérico segundo Chin e Dyer [1] .	14
2.3	Diagrama em blocos do sistema proposto. . . . .	27
3.1	Histograma da imagem original da figura 3.2.a . . . . .	32
3.2	Seqüências de imagens obtidas, a partir da imagem original, durante o pré-processamento: a) original, b) limiarizada, c) filtrada e d) gradiente.	33
3.3	Exemplos de vizinhança expandida para rastreamento. . . . .	35
3.4	A janela mostrada indica uma região crítica para aplicação do algoritmo <i>boundary tracking</i> tradicional. . . . .	36
3.5	Diferentes caminhos percorridos pelos algoritmos para a janela sobre a imagem gradiente mostrada na figura 3.4 . . . . .	37
3.6	Imagem gradiente contaminada por ruído com RSR = 3.0 db. . . . .	39
3.7	Contorno obtido pelo algoritmo BTVE para o objeto Arco (ampliada 2 vezes). . . . .	39
3.8	Contorno obtido pelo algoritmo BTVE para o objeto Fenda (ampliada 1,5 vezes). . . . .	40

3.9	Aproximação poligonal da curva C pelo polígono P. . . . .	42
3.10	Esquema para entendimento do algoritmo. . . . .	46
3.11	Fluxograma do algoritmo Cortez. . . . .	48
3.12	Modelagem de um octógono regular. . . . .	49
3.13	Atributos extraídos dos modelos. . . . .	50
3.14	Modelo obtido pelo algoritmo Cortez para o objeto Arco. . . . .	57
3.15	Modelo obtido pelo algoritmo Split-Merge para o objeto Arco. . . . .	58
3.16	Modelo obtido pelo algoritmo Ramer para o objeto Arco. . . . .	59
3.17	Modelo obtido pelo algoritmo Cortez para o objeto Fenda. . . . .	60
3.18	Modelo obtido pelo algoritmo Split-Merge para o objeto Fenda. . . . .	60
3.19	Modelo obtido pelo algoritmo Ramer para o objeto Fenda. . . . .	61
4.1	Orientação dos segmentos: a) na cena e b) no modelo. . . . .	69
4.2	Ilustração geométrica dos parâmetros necessários ao cálculo dos valores dos limiares. . . . .	72
4.3	Posicionamento do modelo casado na cena: a) usando o grupo completo de hipóteses e b) usando apenas as três primeiras hipóteses. . . . .	81
5.1	Conjunto de objetos que compõem as cinco cenas. . . . .	88
5.2	Contornos e modelos poligonais das cenas: a) cena 1 e b) cena 2. . . . .	88
5.3	Contorno e modelo poligonal da cena 3, contendo cinco objetos presentes. . . . .	89
5.4	Modelo poligonal da cena 4. . . . .	90
5.5	Modelo poligonal da cena 5. . . . .	91
5.6	Modelos poligonais das formas reconhecidas superpostos aos modelos das cenas: a) cena 1 pela aplicação da primeira técnica, b) cena 1 pela aplicação da segunda e terceira técnicas, c) cena 2 pela aplicação da segunda técnica e d) cena 2 pela aplicação da terceira técnica (incluindo fragmentação). . . . .	92

5.7	Resultado obtido pela aplicação da primeira técnica à cena 3. . . . .	93
5.8	Resultado obtido pela aplicação da segunda técnica à cena 3. . . . .	94
5.9	Resultado obtido pela aplicação da terceira técnica à cena 3. . . . .	95
5.10	Resultado obtido pela aplicação das técnicas à cena 4: a) primeira, b) segunda e c) terceira. . . . .	96
5.11	Resultado obtido pela aplicação da primeira técnica à cena 5. . . . .	97
5.12	Resultado obtido pela aplicação da segunda técnica à cena 5. . . . .	97
5.13	Resultado obtido pela aplicação da terceira técnica à cena 5. . . . .	98
5.14	Modelos poligonais das duas cenas contendo os objetos reconhecidos resultante da aplicação das duas técnicas: a) quarta técnica à cena 1, b) quinta técnica à cena 1 e c) quinta técnica à cena 2. . . . .	101
5.15	Resultados obtidos pela quarta técnica aplicada à cena 3. . . . .	102
5.16	Resultados obtidos pela quinta técnica aplicada à cena 3. . . . .	103
5.17	Resultados obtidos pelas duas técnicas aplicadas à cena 4: a) quarta técnica e b) quinta técnica. . . . .	104
5.18	Resultados obtidos pela quarta técnica aplicada à cena 5. . . . .	105
5.19	Resultados obtidos pela quinta técnica aplicada à cena 5. . . . .	105

# Lista de símbolos

- n.c. : níveis de cinza.
- $I(x, y)$ : imagem digital.
- $P_a$ : ponto anterior.
- $P_p$ : ponto presente.
- $P_{ni}, i = 1, 2, 3$ : candidatos a próximo ponto.
- $S_i, i = 1, 2, 3$ : a soma dos níveis de cinza do conjunto  $N_i$ .
- $\mathcal{N}^*$ : é o conjunto dos números naturais estritamente positivos.
- $L_d$ : a distância máxima permitida entre um ponto do contorno e o segmento anterior.
- $N$ : o número de pixels que deverá ser percorrido por vez no processo de rastreamento.
- $P_i$ : último ponto selecionado como possível vértice do polígono.
- $P_{i-1}$  e  $P_{i-2}$ : os pontos que definem o último segmento anteriormente detectado.
- $V_J$ :  $J$ -ésimo vértice do modelo.
- $\mathcal{M}$ : conjunto de modelos.
- $L_{k_n}$ :  $n$ -ésimo segmento do  $k$ -ésimo modelo.
- $l_{n'}$ :  $n'$ -ésimo segmento da cena.
- $\epsilon_i$ : limiar máximo para o casamento de segmentos.
- $\theta(l_i)$ : a orientação do  $i$ -ésimo segmento da cena.
- $\theta(L_I)$ : a orientação do  $I$ -ésimo segmento do modelo.
- $\theta^{(h)}$ : variação de orientação entre o  $i$ -ésimo segmento da cena e o  $I$ -ésimo segmento do modelo.

- $\vec{V}_k^{(h)}$ : vetor que caracteriza a hipótese  $h$ .  
 $f e^{(h)}$ : fator de escala da hipótese  $h$ ,  
 dado pela razão  $\frac{l_i}{L_I}$ .  
 $p_1$  e  $p_2$ : pesos.  
 $\Delta el$ : variação entre os erros  $el^{(h)}$  e  $el^{(t)}$ .  
 $el_{max}$ : variação máxima permitida para o erro  $el^{(\cdot)}$ .  
 $\Delta\theta$ : variação de orientação dos segmentos.  
 $ea_{max}$ : variação máxima permitida para a orientação dos segmentos.  
 $t_x^{(*)}$  e  $t_y^{(*)}$ : os parâmetros relativos à translação nas direções  $x$  e  $y$ ,  
 respectivamente.  
 $\Delta D_{pm}$ : distância entre o ponto médio da cena e a projeção  
 do segmento do modelo na cena.  
 $t_{cl}$ : tempo empregado para a avaliação das hipóteses (classificação) e  
 localização do modelo na cena.  
 $\Delta \bar{D}_{pm}$ : variação média de distância.  
 $\Delta \bar{\theta}$ : variação média de orientação.  
 $f_g$ : valor associado ao grupo de hipóteses.  
 $p_{oc}$ : percentual de oclusão.



# Capítulo 1

## Caracterização do problema

### 1.1 Introdução

O homem tem procurado, ao longo do tempo, construir máquinas capazes de substituí-lo na realização de suas tarefas manuais. Um passo decisivo nesta direção foi dado pela revolução industrial, exigida pelo desenvolvimento social de então.

O surgimento da sociedade moderna possibilitou a criação de máquinas mais complexas para substituir o trabalho humano em diversas tarefas, reduzindo o esforço do homem e melhorando suas condições de vida. Além disso, a maioria das atividades desenvolvidas pelo homem são realizadas com o auxílio da visão, particularmente aquelas atividades que envolvem o reconhecimento de caracteres (por exemplo, leitura deste texto) e objetos (os objetos que nos cercam). Nestes processos estão envolvidos diversos aspectos tais como a determinação de cor, textura, formas e medidas de dimensões, entre outros. Neste contexto, o cientista russo Tyurin, em 1900, introduziu o reconhecimento de caracteres para auxiliar os deficientes visuais [2].

Um grande impulso no processo de desenvolvimento tecnológico da sociedade moderna foi dado pelo aparecimento dos dispositivos semicondutores e posteriormente dos computadores digitais. O desenvolvimento da tecnologia de computadores digitais, particularmente a de integração de circuitos e de transdutores e/ou sensores nas últimas

duas décadas, permitiu aumentar enormemente o campo das suas aplicações. Simultaneamente, progressos importantes foram obtidos na área de inteligência artificial, tornando possível construir máquinas com capacidade cada vez maior de se comunicar com o meio externo e processar inteligentemente as informações recebidas através de sensores para realizar determinadas tarefas [3].

Inicialmente estas tarefas eram simples como, por exemplo, leitura de fitas magnéticas e cartões perfurados. Aos poucos, porém, foram se tornando mais complexas, como ler caracteres manuscritos, reconhecer a voz de um locutor, detectar objetos numa fotografia, guiar veículos e mísseis, interpretar imagens médicas e analisar imagens aéreas e sísmicas [4]. A execução de tais tarefas somente foi possível em decorrência do enorme desenvolvimento dos computadores digitais ocorrido nas últimas duas décadas, que permitiu a manipulação de um grande volume de dados.

Algumas das tarefas listadas acima são realizadas pela utilização da visão computacional, ou visão artificial. O conjunto de técnicas assim denominado, possui como objetivo básico extrair informações sobre uma cena, analisando imagens da mesma [5].

Para que uma imagem analógica possa ser processada por um computador digital, primeiro é necessário convertê-la em uma imagem digital, ou seja, em uma matriz de pontos. O valor de cada elemento desta matriz está associado ao valor do nível de cinza ou cor do ponto correspondente na grade discreta de amostragem da imagem. Estes elementos são usualmente conhecidos como pixels ("picture elements"). Em muitas aplicações, um único valor é medido em cada ponto amostrado, representando o valor do brilho daquele ponto na imagem. Outras aplicações requerem informações sobre as cores e portanto vários valores são medidos para cada ponto amostrado, representando a intensidade do brilho do ponto em um conjunto de  $k$  bandas espectrais. Deste modo, cada pixel será representado por uma  $k$ -upla desses valores [5].

As imagens usadas em visão computacional são obtidas na faixa de energia visível do espectro eletromagnético. Assume-se que a cena é iluminada por uma fonte de luz natural ou artificial e que o sensor óptico incorporado ao sistema de visão transforma qualquer imagem da cena em uma imagem plana, perpendicular ao campo de visão do sensor óptico. São ignorados os objetos transparentes [5].

Visão computacional tem como principal objetivo permitir ao computador entender os aspectos do seu meio, utilizando as informações provenientes dos sensores visuais. Por isto, a maioria das técnicas utilizam imagens obtidas na faixa visível do espectro eletromagnético (monocromática ou colorida), embora uma variedade de outras técnicas tenham sido desenvolvidas usando imagens de raios-X, luz infravermelha e ultravioleta, imagens térmicas, sonar e radar [6].

Durante muitos anos, visão computacional se desenvolveu como um ramo da inteligência artificial, tendo posteriormente se afirmado como uma área específica. Visão computacional lida com uma grande quantidade de disciplinas, como processamento de sinais, psicologia e psicofísica, muitos ramos da matemática e estatística, computação gráfica e muitos ramos da ciência da computação [6]. Aplicações de visão computacional envolvem mais diretamente, em geral, técnicas e conceitos das áreas de Processamento Digital de Imagens, Reconhecimento de Padrões e Inteligência Artificial [7].

Processamento Digital de Imagens, aplicado à visão computacional, trata do conjunto das transformações aplicadas sobre uma ou várias imagens de entrada e de sua descrição. Como saída deste processo, tem-se um conjunto de medidas e propriedades dos objetos ou regiões, contidas na cena representada pela imagem. O processo derivado da descrição de uma cena é algumas vezes chamado de análise da imagem.

Reconhecimento de Padrões trata de classificar os dados resultantes da descrição da imagem. Em geral, a função principal de um sistema dotado de visão artificial é reconhecer objetos, formas ou padrões, de vários tipos que possam estar presentes em uma determinada cena, ou seja, classificá-los [5, 8].

Inteligência Artificial procura dotar os computadores de procedimentos semelhantes aos dos seres humanos na solução de problemas análogos.

Os diversos sistemas de visão existentes apresentam características gerais que são abordadas na seção seguinte.

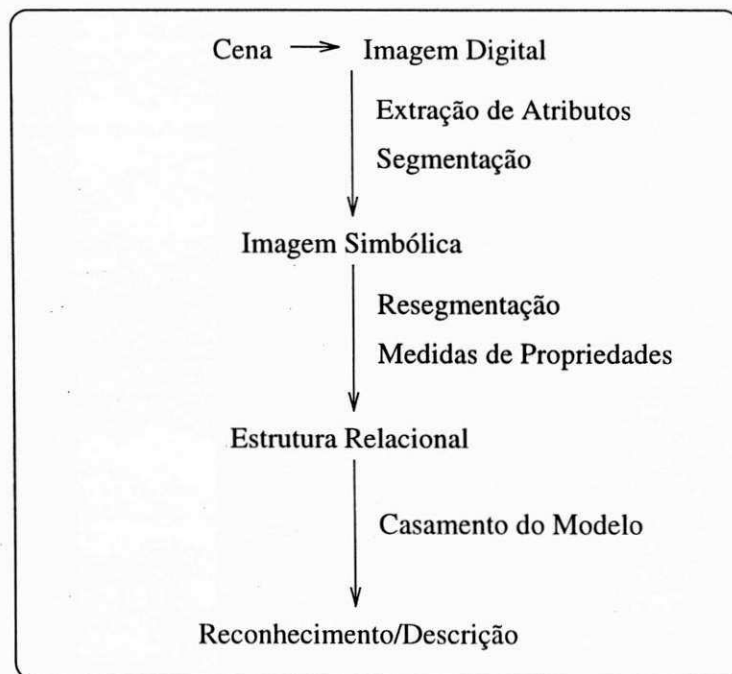


Figura 1.1: Estrutura geral de um sistema de reconhecimento.

## 1.2 Aspectos gerais de um sistema de visão

A estrutura genérica de um sistema para reconhecimento de objetos em cenas bidimensionais a partir de uma imagem da cena, é ilustrado na figura 1.1, proposto por Rosenfeld [5].

Um objeto é definido como uma disposição de partes, cujas propriedades (níveis de cinza, textura, formas, comprimentos, etc.) e relações (posições relativas, comprimentos relativos, etc.) satisfazem determinadas condições. Deste modo, para reconhecer um objeto numa imagem, verifica-se se uma coleção de partes da imagem corresponde às partes do objeto e se satisfazem às condições apropriadas [5].

Os processos usados para detectar partes numa imagem são conhecidos como segmentação e/ou técnicas de deteção de atributos. Deteção de atributos identifica os tipos especiais de padrões locais na imagem, como por exemplo linhas ou curvas, manchas, cantos, etc.. Tais padrões podem ser detectados pela comparação da imagem com modelos (em cada posição e orientação) ou pela comparação de várias combinações das

primeiras e segundas derivadas em cada ponto da imagem. A segmentação identifica regiões distintas de uma imagem (subpopulações de pixels) ou particiona a imagem em regiões conectadas, cada uma delas “homogênea” com base em algum critério. Este processo produz como saída uma imagem “simbólica”, na qual o “valor” de cada pixel é um rótulo, indicando a qual região o mesmo pertence, antes que um nível de cinza [5] (figura 1.1).

Regiões identificadas durante o processo de segmentação, podem ser reagrupadas em novas regiões com base em vários critérios, sendo este processo denominado de resegmentação. Por exemplo, pode-se dividir um conjunto de pixels em regiões conectadas, identificando partes de uma região como côncava, convexa, alongada, etc.. Da mesma forma, diferentes curvas ou arcos podem ser resegmentados com base em critérios geométricos [5] (figura 1.1).

De posse das regiões obtidas pelo processo de (re)segmentação de uma imagem, pode-se agora medir várias propriedades destas regiões e as relações entre elas. Propriedades geométricas dependem unicamente da forma e do tamanho do conjunto de pixels contido nestas regiões (por exemplo área e alongamento), enquanto que outras propriedades dependem dos valores de níveis de cinza destes pixels (por exemplo, valor médio e textura). Este processo é chamado de descrição [5] (figura 1.1).

As informações resultantes do processo de descrição são, em geral, adequadas à representação por grafo rotulado. Os nós do “grafo objeto” representam as partes ou regiões da imagem de um objeto e os arcos, ligando os nós, as relações entre estas regiões ou partes. Os nós e arcos, que representam valores ideais de propriedade ou um conjunto constante destes valores, são “rotulados” com atributos invariantes, deste modo, a primeira aproximação para o reconhecimento de objetos em uma imagem é feita extraindo-se regiões ou partes apropriadas da imagem (detecção de atributos ou técnicas de (re)segmentação), computando-se suas propriedades e relações (descrição) e representando essas informações em grafo rotulado (“grafo cena”), onde os rótulos são os valores de propriedades e relações extraídos da cena. Portanto, o reconhecimento pode ser feito encontrando os subgrafos do grafo cena que aproximem o grafo objeto, ou que satisfaçam as condições definidas pelo grafo objeto. Esta aproximação se aplica

a cenas bidimensionais mas pode também ser aplicada a cenas tridimensionais se a orientação do objeto for conhecida. Neste último caso, dado um modelo 3D do objeto, primeiro faz-se uma predição de como o objeto apareceria sob uma dada vista para depois criar um grafo objeto para a correspondente predição [5].

### 1.3 O problema

Diversas aplicações de visão computacional envolvem cenas essencialmente bidimensionais, destacando-se aquelas que utilizam superfícies planas na manufatura. Neste contexto, o reconhecimento de objetos em 2D é a tarefa de encontrar e identificar partes de uma imagem 2D de uma cena que correspondam aos objetos desta mesma cena. Esta tarefa se inicia estabelecendo modelos ou descrições gerais de cada objeto a ser reconhecido [9].

As principais aplicações de visão artificial na indústria se concentram na inspeção automática e manipulação robótica de partes. Dotar os robôs de visão, significa torná-los flexíveis o suficiente para a realização de tarefas complexas, podendo seu comportamento ser adaptado às tarefas, ou seja, transformá-los em polivalentes [4].

Visão computacional, quando aplicada à manipulação automática, requer como tarefa inicial reconhecer e localizar precisamente os objetos presentes na cena. A complexidade desta tarefa depende da disposição espacial dos objetos na cena [4].

Uma situação típica de um ambiente industrial para manipulação automática é aquela de uma linha de montagem em que os objetos se encontram num plano horizontal fixo, ou em uma esteira rolante, com condições de iluminação que, em geral, podem ser controladas, como ilustrado na figura 1.2. A solução para este tipo de situação é na maior parte das vezes relativamente simples e consiste em colocar uma ou mais câmeras em posições fixas, de modo que a cena de interesse se encontre dentro dos respectivos campos de visão. Nestas condições, o reconhecimento e a localização dos objetos presentes na cena podem ser feitos através de uma análise e classificação das cenas em 2D apenas.

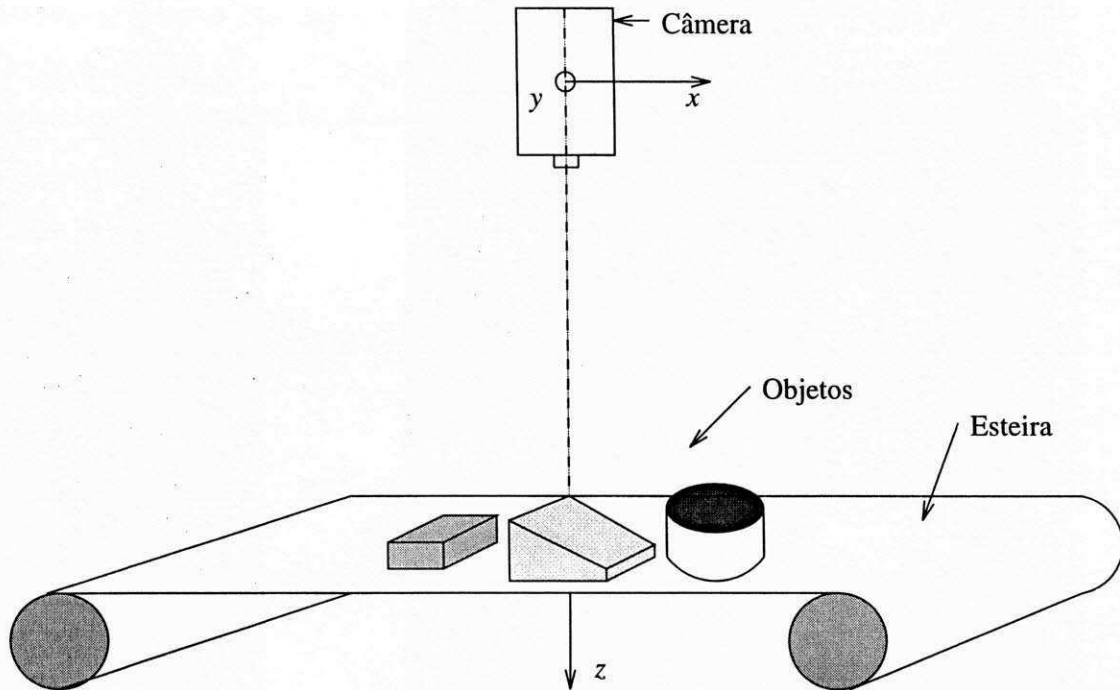


Figura 1.2: Situação típica de um ambiente industrial.

O grau de complexidade da situação recém descrita pode aumentar consideravelmente quando se admite as mais diversas disposições espaciais dos objetos presentes na cena e também pela variação das condições de iluminação ou até mesmo pela presença de ruídos ambientais. Pode-se resumir os fatores que aumentam a complexidade desta situação como:

- baixo contraste entre objeto e fundo na cena,
- condições de iluminação degradadas e/ou presença de ruídos,
- objetos em contato,
- objetos parcialmente visíveis,
- objetos com formas geométricas variadas e
- alto grau de similaridade entre os objetos.

## 1.4 Objetivo

A tese descrita neste texto tem como principal objetivo contribuir para o estado da arte no campo de reconhecimento de formas bidimensionais em imagens. Com este fim, uma nova metodologia foi desenvolvida que utiliza modelos poligonais e seus atributos geométricos para realizar a tarefa de identificar objetos. Esta metodologia é composta por um conjunto de novas técnicas, tanto para modelagem quanto para reconhecimento, e apresenta quatro importantes e inovadores aspectos: a) utiliza um algoritmo de segmentação por rastreamento de contornos (*boundary tracking*) que emprega uma vizinhança expandida, chamado BTVE (*boundary tracking* com vizinhança expandida), b) integra as várias fases do processo de reconhecimento (figura 1.1) em uma única etapa, c) utiliza um algoritmo seqüencial para a modelagem de objetos e d) realiza a geração e agrupamento de hipóteses ainda durante o processo de segmentação.

As técnicas desenvolvidas foram implementadas num sistema de visão computacional composto por um conjunto de programas, utilizando a linguagem C, com a capacidade de reconhecer e localizar objetos em cenas 2D. Estes objetos podem aparecer isolados, em contato ou mesmo parcialmente visíveis. A inclusão no problema de cenas nas quais os objetos possam estar em contato ou parcialmente oclusos, implica em estabelecer um sério compromisso entre a flexibilidade do sistema, relativa a sua capacidade de reconhecer objetos em contato ou parcialmente visíveis, de um lado e a sua complexidade do outro.

Esta tese é descrita da seguinte forma: o capítulo 2 faz uma revisão dos sistemas baseados em modelos usados para o reconhecimento de objetos 2D, o capítulo 3 trata do processo de modelagem poligonal e avalia os algoritmos desenvolvidos para este fim, o capítulo 4 apresenta a metodologia desenvolvida para o processo de reconhecimento, o capítulo 5 trata dos resultados experimentais obtidos pelas cinco técnicas desenvolvidas e o capítulo 6 descreve as conclusões e principais contribuições desta tese e aponta os trabalhos futuros a serem realizados.



## Capítulo 2

# Reconhecimento de formas 2D: revisão bibliográfica e proposta de nova metodologia

### 2.1 Introdução

O desenvolvimento das pesquisas e da tecnologia em Visão Artificial associado ao fato de que o homem utiliza a visão em quase todas as suas atividades, possibilitou um aumento considerável de sua aplicação nas mais diversas áreas nas últimas décadas. Dentre estas, destacam-se as áreas de medicina [10], sensoriamento remoto [11], automação bancária [12] e automação industrial [13]. Esta última, utiliza a Visão Artificial principalmente para inspeção e controle de qualidade.

Dependendo da aplicação específica, um sistema de visão sofisticado pode ser necessário, como os sistemas desenvolvidos por Wong [14], Chen e Ho [15] e por Tsang e To [16] que trabalham em três dimensões (3D). Com o intuito de tornar estes sistemas mais robusto, o sistema desenvolvido por Zhang et al. [17] utiliza múltiplas fontes de dados, obtidas por diferentes sensores, para a inspeção automática industrial em 3D. Porém, alguns destes sistemas utilizam modelos e estratégias de reconhecimento

já empregadas anteriormente em sistemas de reconhecimento que operam apenas em duas dimensões [1].

Diversos problemas podem ser resolvidos com o uso de sistemas de visão bidimensional. Em alguns destes problemas, a natureza bidimensional está intrinsicamente associada ao problema. Este é o caso, por exemplo, do reconhecimento de caracteres, da restauração de documentos escritos e de obras de arte, etc. . Uma grande parte destes sistemas são sistemas que tentam casar uma cena (imagem de entrada) com algum elemento pertencente a um conjunto de modelos construídos a priori durante a fase de treinamento do sistema. Estes sistemas são conhecidos como sistemas baseados em modelos (*model-based*). A principal vantagem desta classe de sistemas é a grande compressão de dados obtida através da representação dos contornos por modelos ou por um pequeno número de atributos.

Neste capítulo, é apresentada uma revisão bibliográfica dos sistemas bidimensionais para reconhecimento de objetos que utilizam métodos baseados em modelos, alguns deles considerados clássicos na literatura especializada. Inicialmente são caracterizados os sistemas bidimensionais usados para reconhecimento de objetos. Em seguida é feita uma classificação dos sistemas baseados em modelos, seguida de suas descrições, bem como uma caracterização geral de cada uma das classes com suas principais vantagens e desvantagens. Finalmente, o problema a ser resolvido por esta pesquisa é apresentado e sua solução discutida.

## **2.2 Classificação de sistemas para reconhecimento de objetos**

A classificação de sistemas para reconhecimento de objetos é feita levando-se em conta suas principais características, dependendo do ponto de vista da abordagem. Duas abordagens bastante citadas na literatura foram consideradas no presente estudo. Uma mais recente que classifica os sistemas pela estratégia usada para o reconhecimento [9] e outra que classifica os sistemas conforme a dimensão espacial da descrição,

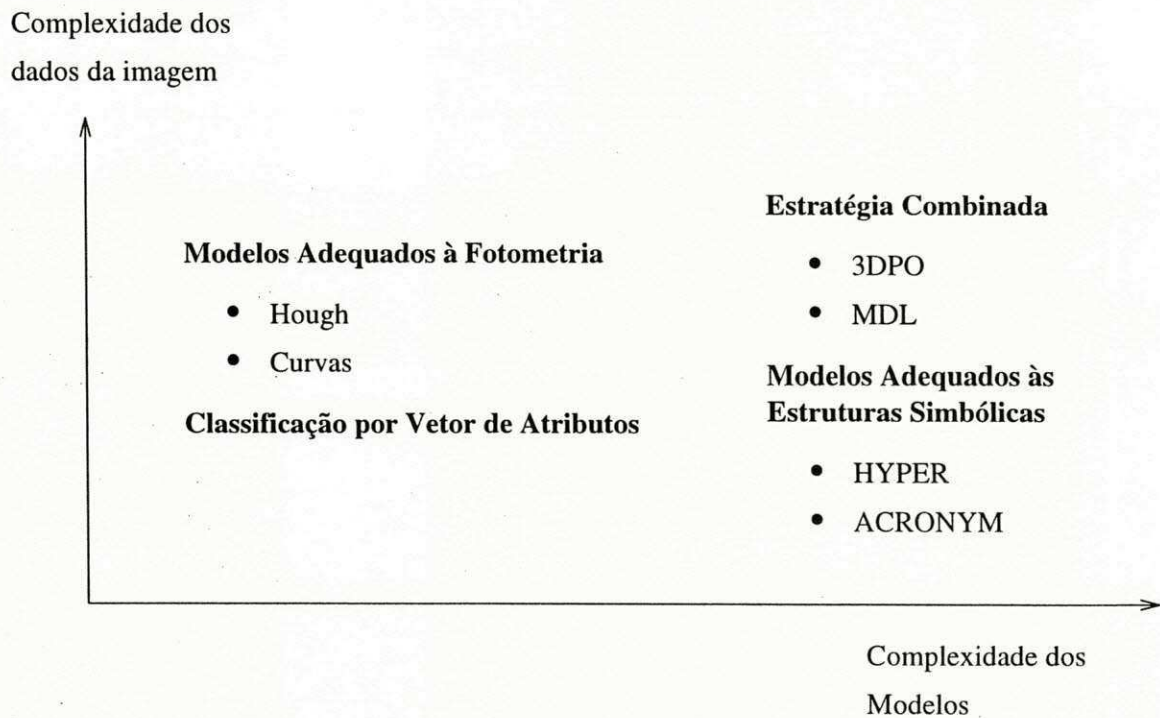


Figura 2.1: Classificação dos sistemas em função das complexidades dos dados e dos modelos.

subdividindo-os também quanto à natureza dos atributos utilizados [1].

### 2.2.1 Classificação segundo Suetens et al.

A abordagem proposta por Suetens et al. [9] é genérica, servindo para sistemas de reconhecimento de objetos em duas ou três dimensões. Entretanto, não possui uma forma específica para sistemas em 2D.

Esta classificação se baseia nas estratégias usadas para o processo de reconhecimento que por sua vez dependem da complexidade dos dados e dos modelos. A figura 2.1 mostra esta classificação, na qual aparecem quatro estratégias como função da complexidade dos dados e dos modelos, indicando para cada uma das estratégias um sistema conhecido ou um tipo de técnica mais usada.

Suetens et al. definem a complexidade dos dados pela medida da quantidade de

ruído presente na imagem e pela ambigüidade semântica das imagens [9]. Portanto, a complexidade dos dados aumenta com o aumento da presença de ruído e/ou a ambigüidade contida na imagem.

A complexidade dos modelos é medida pelo nível de detalhes da estrutura de dados requerida para o armazenamento dos modelos e pelas técnicas necessárias à organização dos dados dos modelos [9].

Esta classificação apresenta como desvantagem o fato de não considerar que a complexidade dos dados, conforme definida pelos autores, possa vir a ser reduzida por uma etapa de pré-processamento mais sofisticada [18, 19, 20, 21, 22, 23]. Além disso, por ser geral, não trata das aplicações industriais especificamente.

### **2.2.2 Classificação segundo Chin e Dyer**

A classificação, proposta por Chin e Dyer, baseia-se na dimensão espacial da descrição e representação dos objetos, agrupando os sistemas em três classes iniciais 2D,  $2\frac{1}{2}$ D e 3D.

Esta classificação é a mais adequada para o propósito desta pesquisa porque é específica para sistemas baseado em modelos para uso em aplicações industriais, principalmente em visão para robôs. Subdivide a classe dos sistemas com representação 2D, de forma bastante clara, conforme as características dos atributos utilizados, facilitando ainda mais sua compreensão.

Com base no exposto acima, adota-se, a classificação de Chin e Dyer para os sistemas baseados em modelos descritos na seção seguinte. Entretanto, sempre que necessário serão introduzidos comentários com relação à adequação desta, principalmente quando se tratar de novas técnicas propostas.

## **2.3 Sistemas baseados em modelos**

Boa parte dos sistemas de visão desenvolvidos até o momento, usam modelos 2D para a descrição e reconhecimento de peças industriais em cenas complexas [1]. Alguns

têm como base as propriedades radiométricas (como por exemplo, os níveis de cinza, cor, textura, etc.) para a descrição e o reconhecimento de cenas. Este é o caso dos sistemas propostos por Asano et al. [24], Mital et al. [25], Sepeda e Stemmer [26] e por Fukumi et al. [12]. Outros utilizam propriedades geométricas que são muito importantes em aplicações industriais [1]. Dentre esses sistemas podem ser citados os sistemas desenvolvidos por Rillo [27], Turney et al. [28], Bhanu e Faugeras [29], Ayache e Faugeras [30, 13], Fukuda et al. [31], Maes [32], Boussofiane e Bertrand [33], Wen e Lozzi [34], Stein e Medioni [35], Sethi e Ramesh [36], Tsang et al. [37], Nasrabadi e Li [38], Liu e Srinath [39], Basak e Pal [40], Cortez e Carvalho [41] e muitos outros.

Conforme a classificação realizada por Chin e Dyer, os métodos utilizados no reconhecimento de objetos 2D, dependendo da natureza dos atributos usados, podem se dividir em três classes: métodos dos atributos globais, métodos dos atributos estruturais e métodos dos grafos relacionais. A decisão na escolha dos atributos é subjetiva e depende do modelo e da aplicação específica. O procedimento em cada uma destas classes é formado por três partes básicas: extração de atributos, modelagem das cenas e casamento. A figura 2.2 mostra, em diagrama de blocos, um sistema genérico de reconhecimento de objetos baseado em modelos bidimensionais que opera de acordo com este procedimento [1].

A partir das partes básicas que compõem um sistema baseado em modelos, são descritas as classes de métodos: métodos dos atributos globais, métodos dos atributos estruturais e métodos dos grafos relacionais. Esta descrição é iniciada pelos métodos dos atributos globais.

### 2.3.1 Métodos dos atributos globais

Esta classe usa, como o próprio nome sugere, atributos globais dos modelos, tais como perímetro, área, momentos de inércia, centróide, níveis de cinza dos objetos, cor, contorno ou amostras do contorno do objeto<sup>1</sup>. Geralmente utiliza uma combinação destes atributos.

---

<sup>1</sup>Introduzido aqui por entender-se que este atributo está relacionado com a forma do objeto como um todo, ou seja, é um atributo global.

às mudanças de escala, sendo portanto usadas para representar os objetos durante o treinamento e para realizar a classificação na etapa de reconhecimento.

Outro sistema, cuja representação dos objetos é mais compacta do que aquela usada pelo sistema descrito anteriormente, também baseado em morfologia matemática, foi proposto por Trahanias [53]. Neste sistema os objetos são representados por conjuntos de pontos (esqueletos), também descritos por Pitas e Venetsanopoulos [54], obtidos através de operações morfológicas. Esta representação denominada de MST (*Morphological Skeleton Transform*) deve ser normalizada quanto à translação, rotação e escala para que possa ser utilizada pelo sistema. As informações contidas nestes conjuntos sobre a forma dos objetos são compactadas numa função SKF (*Skeleton Function*). O casamento é feito pelo cálculo da distância entre as MSTs, através de uma função custo, computada pixel por pixel, cujos pesos dependem das SKFs.

O processo de reconhecimento, realizado pelos métodos dos atributos globais, é feito associando-se, em geral, um vetor de atributos  $n$ -dimensional ao modelo do objeto, onde  $n$  é o número de atributos utilizados. Normalmente, métodos estatísticos de reconhecimento de padrões são empregados, em que vetores de atributos extraídos da cena são comparados com cada um dos vetores de atributos dos modelos armazenados. Os modelos armazenados são formados durante a fase de treinamento ou aprendizagem do sistema, a partir de imagens representativas de cada um dos objetos possíveis de serem reconhecidos. Aos poucos esta técnica está sendo lentamente substituída por técnicas baseadas em redes neuronais, em modelos escondidos de Markov [45, 46, 55], transformadas de Hough [44] e da distância [39], bem como por técnicas baseadas em morfologia matemática [52, 53], embora continue sendo aplicada em trabalhos recentes como o sistema desenvolvido por Das et al. [48].

Em geral, métodos baseados em atributos globais consomem a maior parte do tempo na tarefa de extração de atributos. Estes atributos são fáceis e rápidos de serem obtidos, podendo ser computados em tempo real. Além disso, são invariantes às transformações de translação e rotação, o que significa que os objetos podem se encontrar em qualquer posição e orientação. Por outro lado, o esforço computacional exigido

na classificação aumenta rapidamente com o número de atributos mas, em geral, produz resultados aceitáveis [1]. Outra importante vantagem destes métodos é a grande sensibilidade à auto-occlusão, ou seja, objetos defeituosos são sempre rejeitados, sendo adequados portanto às aplicações de controle de qualidade. Sua principal desvantagem está em não permitir o reconhecimento de objetos em contato, parcialmente oclusos ou ainda com algum defeito. Por este motivo são inadequados às aplicações em 3D, onde cada vista de um objeto pode ser descrita por um modelo diferente. Porém, podem ser muito úteis como ponto inicial para métodos mais sofisticados [9].

Algumas aplicações são muito adequadas ao uso dos métodos dos atributos globais, dentre elas se destacam o reconhecimento de caracteres e inspeção visual para controle de qualidade em linhas de produção. Entretanto, uma adicional desvantagem destes métodos é a incapacidade de distinguir entre modelos com atributos globais semelhantes. Uma forma de melhorar o desempenho é usar, na classificação, a decisão hierárquica associada ao classificador de vizinho mais próximo. Deste modo, se  $n$  atributos são usados na descrição de todos os modelos, então, cada modelo é representado por um ponto no espaço  $n$ -dimensional de atributos. Por conseguinte, uma dada lista de atributos de uma imagem de entrada, será reconhecida como sendo aquela da qual se encontra mais próxima, no espaço de atributos, dentre as listas de atributos dos modelos treinados.

Os sistemas baseados em morfologia matemática possuem uma desvantagem peculiar que é a representação de objetos por conjuntos de pontos, acarretando em um enorme esforço computacional durante o processo de reconhecimento. Além disto, dependem de uma escolha adequada para os elementos estruturantes usados pelas operações morfológicas. Porém, toleram que os objetos sofram pequenas distorções, ou mesmo defeitos, o que os torna adequados à muitas aplicações.

As novas técnicas surgidas a partir do final da década passada, principalmente aquelas que empregam redes neuronais, morfologia matemática, modelos escondidos de Markov, transformadas de Hough e da distância, em geral, toleram a auto-occlusão. Dentre estas técnicas, aquelas baseadas em redes neuronais são mais rápidas do que as baseadas em morfologia matemática.

Um dos primeiros métodos pertencentes à classe dos métodos de atributos globais é o protótipo desenvolvido por Gleason e Agin, divulgado em 1979, denominado de SRI Vision Module para aplicações industriais [1].

A classe de métodos descrita nesta subseção não satisfaz as exigências de diversas aplicações que necessitam de sistemas mais complexos e com melhor desempenho. Os métodos descritos a seguir se enquadram nesta categoria.

### 2.3.2 Métodos de atributos estruturais

Os métodos de atributos estruturais usam atributos locais para a construção dos modelos e descrevem propriedades mais complexas dos objetos, em comparação com os métodos de atributos globais. Estes atributos incluem segmentos retos, arcos de curvatura constante e curvas, dentre outros, para definir as partes de contornos dos modelos dos objetos. Os atributos são representados abstratamente e organizados de forma estruturada em listas ou em uma seqüência de equações. O ordenamento dos atributos é feito de tal forma que seguir a lista de atributos seqüencialmente é equivalente a percorrer a borda do objeto.

Os modelos baseados em atributos locais estão associados aos métodos de casamento sintático, nos quais os atributos locais são transformados em primitivas que são organizadas em sentenças por regras gramaticais bem estruturadas. O reconhecimento usa os procedimentos de verificação de hipóteses. A estruturação dos atributos locais nos modelos é usada para predizer a localização dos objetos na cena. Desta forma, atributos de objetos hipotéticos são medidos, na base da predição hipotética do modelo, para a verificação ordenada, finalizando com o casamento.

Neste sentido, Maes [32] propõe um método para reconhecimento de formas, representadas por polígonos, baseado no casamento de sentenças (*strings*) circulares. Este método resolve, segundo o autor, o principal problema do casamento de sentenças lineares que é a dificuldade de encontrar a orientação do objeto para determinar o próprio símbolo inicial na sentença correspondente. No sistema de Maes, os lados de



um polígono são representados por uma seqüência de símbolos, constituindo uma sentença. Esta seqüência reflete as posições relativas entre os lados para um dos sentidos de marcação, horário ou anti-horário. Este método é sensível às condições de iluminação, ruído e à orientação dos objetos, que produzem inconsistências no processo de segmentação e conseqüentemente variações na aproximação poligonal.

Pesquisando em outra direção, Tsang et al. [37] desenvolveram uma aproximação para o reconhecimento de objetos oclusos empregando uma rede neuronal tipo *Perceptron*. Neste método, as bordas de um objeto são extraídas pelo cruzamento por zero do operador Laplaciano e transformadas numa curva 1D chamada de  $\theta(s)$ . O parâmetro  $s$  é o comprimento transversal do contorno a partir de um ponto inicial, escolhido aleatoriamente. Nesta curva são introduzidos um fator de compensação e uma filtragem gaussiana para eliminar descontinuidades artificiais. Da curva resultante, por um processo analítico, são extraídos pontos de controle pela detecção das descontinuidades de ordem zero e de primeira ordem. Por análise de regressão sobre um conjunto de pares  $(\theta, s)$  de amostras da curva  $\theta(s)$  é calculado o coeficiente de correlação e com base neste são obtidos os atributos: retas e curvas. Durante a fase de treinamento, cada um dos objetos é apresentado sob 10 orientações diferentes, com variação entre elas de 18° graus. O problema principal deste método é o sofisticado processo para extrair os atributos, embora este fato não seja abordado pelos autores.

Partindo do princípio de que a aproximação poligonal não é única e manipulando um grande número de objetos, Stein e Medioni [35] propuseram um método com indexação estrutural. Neste, a representação dos objetos é feita por supersegmentos, extraídos dos vários modelos poligonais de cada objeto para diferentes valores de precisão, segundo o autor para aumentar a robustez da representação quanto à variações dos modelos. Os supersegmentos são formados por cadeias de segmentos adjacentes do modelo poligonal. Estes supersegmentos e as suas propriedades geométricas são codificados e armazenados em uma tabela, na qual o código é usado como chave para buscas feitas durante o processo de geração de hipóteses. O casamento se realiza com base um supersegmento com 6 segmentos, 2 supersegmentos com 5 segmentos, 3 supersegmentos com 4 segmentos e 4 supersegmentos compostos por 3 segmentos cada.

Todos os supersegmentos são obtidos a partir dos mesmos 6 segmentos. A principal vantagem deste método é o uso da indexação no armazenamento da estrutura de dados, a qual reduz o tempo de busca durante o processo de geração de hipóteses. Porém, o uso de um grande volume de atributos e modelos para representar cada objeto compromete o tempo gasto para o casamento como pode ser comprovado pelos resultados apresentados: em torno de 10 segundos para o reconhecimento de formas de animais e 70 segundos para nove formas de aviões. Isto reflete o forte compromisso, válido para qualquer sistema de visão, entre o grande volume de modelos e o tempo consumido para os reconhecer.

Na mesma direção do trabalho de Stein e Medioni, mas usando uma função para dividir os dados dos modelos e os armazenar numa estrutura chamada de árvore  $k - d$ , Sethi e Ramesh [36] desenvolveram um sistema no qual os atributos de cada modelo são guardados em hipercélulas. Estas unidades são obtidas pela divisão do espaço  $k$ -dimensional, de forma a manter o número de pontos por célula constante. Na etapa de treinamento, o sistema estabelece as associações entre os atributos e os modelos para produzir as tabelas, ponteiros, e outras estruturas necessárias, dependendo da escolha para representação dos objetos. Durante a fase de reconhecimento, as mesmas superposições de atributos locais são usadas para compor a pesquisa associativa. Cada acesso aos dados do modelo, produz um voto para um dos modelos armazenados. O modelo reconhecido é aquele com maior número de votos. A principal vantagem deste método é a eficiente estrutura para armazenar os modelos (árvores  $k - d$ ), a qual particiona recursivamente um espaço multi-dimensional com a finalidade de reduzir o tempo gasto na busca. Entretanto, podem ocorrer colisões de dados, dependendo da função responsável pela divisão e, principalmente, do conjunto de formas a serem reconhecidas.

Um sistema muito citado na literatura foi implementado por Ayache [4] e se baseia na representação dos objetos por modelos poligonais. Atributos geométricos dos modelos são extraídos e armazenados ordenadamente numa árvore. O casamento é feito pela técnica chamada HYPER (*Hypotheses Predict and Evaluated Recursively*) para a geração e verificação das hipóteses através de uma estimação recursiva da posição

dos modelos hipotéticos na cena. Esta é uma técnica iterativa, consumindo portanto muito tempo de processamento para completar a avaliação, embora o autor afirme que a mesma é muito rápida. Alguns resultados ultrapassam os cem segundos e ainda não convergem para uma solução. Uma forma de reduzir o número de hipóteses geradas e conseqüentemente o tempo gasto nas avaliações é pelo uso apenas dos 5 (cinco) segmentos mais longos (privilegiados). Este método é preciso e flexível no que diz respeito ao reconhecimento de objetos parcialmente oclusos. Além disso, utiliza um algoritmo paralelo associado a um hardware dedicado para a geração e verificação de hipóteses.

Boussofiane e Bertrand [33] num trabalho mais recente propõem uma técnica também baseada em modelos poligonais. Os segmentos dos modelos dos objetos e da cena são armazenados numa tabela. Cada caminho nesta tabela representa o casamento entre segmentos dos modelos e da cena. O problema de fragmentação de segmentos é representado por caminhos verticais na tabela, expressando o fato de que sucessivos segmentos na cena correspondem a um único segmento do modelo. O maior caminho percorrido nesta tabela corresponde ao melhor casamento. Este método possui duas vantagens, a primeira é a rapidez da pesquisa na tabela realizada durante a geração das hipóteses e a segunda é incorporar o problema de segmentos fragmentados no processo de reconhecimento.

Basak e Pal [40] desenvolveram um sistema (PsyCOP) que integra a transformada generalizada de Hough com o fato de que os processos de identificação e localização de objetos ocorrem em regiões distintas do cérebro humano. Este sistema é implementado em duas redes neuronais separadas, uma responsável pela identificação e a outra pela localização dos objetos na cena, controladas por um módulo denominado ACN (*attention control network*). Além disso, incorpora as informações estruturais dos objetos contidas nos seus modelos poligonais. A estratégia de reconhecimento emprega a técnica de verificação iterativa de hipóteses, considerando-se 300 iterações para os resultados apresentados. As maiores desvantagens deste sistema são: o grande número de neurônios usados ( $\approx 5 \times 10^5$ ), é limitado a objetos cujos modelos possuem menos que 64 segmentos, não pode ser empregado com objetos de forma qualquer e finalmente o uso da técnica de verificação iterativa de hipóteses no processo de classificação que,

em geral, consome muito tempo.

Uma das desvantagens dos métodos de atributos estruturais é que os atributos de bordas (locais) não são, em geral, invariantes às translações e às rotações. Por conseguinte, o reconhecimento consiste de uma seqüência de tentativas de casamento de alguns atributos locais e de uma rigorosa análise gramatical das sentenças. O resultado de cada tentativa é usado para restringir a pesquisa de outros atributos, visando reduzir o tempo consumido neste processo. Os procedimentos para a verificação de hipóteses consomem muito tempo se o modelo não encontrar uma aproximação desejada. Muitas soluções foram propostas em trabalhos já relatados no sentido de melhorar o desempenho destes métodos.

Os métodos de atributos estruturais representam um avanço em relação aos métodos de atributos globais, em termos de capacidade e de robustez. Entretanto, requerem mais complexos e sofisticados processos de treinamento e de reconhecimento, necessitando por isso, de maior esforço computacional. Essa classe utiliza muitos atributos locais para modelar os objetos. Contudo, no processo de pesquisa e casamento dos atributos de uma imagem, pode, em geral, ser evitado o teste de todos os atributos, trabalhando-se apenas com os mais significativos para o problema em questão.

Algumas aplicações exigem sistemas mais complexos do que aqueles apresentados nesta subseção. Uma alternativa é utilizar métodos que empregam grafos relacionais, os quais são descritos em seguida.

### **2.3.3 Métodos dos grafos relacionais**

Os métodos dos grafos relacionais utilizam uma combinação de atributos locais e das relações entre estes atributos. Dentre estas, as relações geométricas entre os atributos locais (cantos e retas) têm um particular interesse.

A estrutura relacional é representada por um grafo no qual cada nó está associado a um atributo local e é rotulado pela lista de propriedades daquele atributo. Os arcos representam as relações entre os atributos dos pares de nós que se encontram ligados e são rotulados com uma lista de valores relacionais (distâncias e adjacências).

O reconhecimento do modelo do objeto, em geral, é feito através do casamento dos grafos, um grafo-cena da imagem de entrada e um grafo-modelo para cada um dos modelos, construídos durante o processo de treinamento. Portanto, em geral, são utilizadas técnicas de casamento de grafos, as quais tentam encontrar isomorfismos entre grafos. O casamento torna-se evidente se houver um mapeamento bijetor entre nós e arcos dos grafos. Este mapeamento é de utilização limitada no casamento de estruturas relacionais. Esta limitação pode ser removida pela atribuição de pesos aos nós e arcos para expressar sua importância no casamento entre as duas estruturas. Portanto, podem ser encontrados nós e arcos com valores nulos, conduzindo ao problema de casamento inexato, no qual a correspondência entre arcos e nós é relaxada [11].

O problema de determinar o melhor casamento entre duas estruturas relacionais pode ser resolvido por grafos de associação, formados pela associação de nós e arcos destas estruturas. Cada nó do grafo de associação consiste de uma dupla de nós oriundos das duas estruturas relacionais a serem casadas. Os arcos entre os nós deste grafo exprimem uma relação de compatibilidade entre as estruturas. O problema do casamento entre duas estruturas pode ser resolvido encontrando-se um conjunto de associações mutuamente compatíveis e o melhor casamento será encontrado pelo maior conjunto destas associações. Em grafos de associação, este conjunto é conhecido como um clique, ou seja, um conjunto de nós totalmente conectados. Se nenhum nó puder ser adicionado sem destruir suas propriedades, este clique é denominado de clique máximo [11].

A detecção de clique máximo exige um enorme esforço computacional. Uma alternativa para este problema é eliminar o maior número possível das associações pouco compatíveis, através de um algoritmo paralelo. Tal procedimento pode ser realizado por esquemas de relaxação em grafos de associação [11].

Os métodos dos grafos relacionais podem ser usados para manipular cenas em que os objetos se encontram em contato ou parcialmente oclusos. Neste caso as partes visíveis dos objetos correspondem a subgrafos e o reconhecimento se reduz ao casamento dos subgrafos.

Algumas vantagens adicionais dessa classe de métodos podem ser citadas: a extração dos atributos locais consome pouco tempo porquê estes são simples e podem ser extraídos seqüencialmente; os modelos são pouco sensíveis à pequenas variações para um dado tipo de objeto; a ausência de alguns atributos locais (por causa de ruídos ou oclusão), não impede o reconhecimento de um dado objeto, com base nos atributos restantes associados ao modelo.

Os métodos de grafos relacionais diminuem os requisitos com relação a como os objetos são apresentados ao sistema para que o reconhecimento seja bem sucedido. Desde que o modelo possua os atributos locais e suas relações na forma de grafo, o casamento não depende unicamente da presença exata dos atributos de borda. Também outros atributos e propriedades de suas interconexões (distâncias, adjacências, etc.) influem neste processo. Para o reconhecimento ter sucesso, o único requisito é que um conjunto chave de atributos locais se encontrem visíveis e corretamente posicionados. Isto pode constituir uma restrição, pois, um conjunto de objetos que possuam formas muito semelhantes (silhuetas) pode produzir um conjunto chave de atributos locais que apresentem suas propriedades muito próximas nos grafos, portanto, comprometendo o reconhecimento.

A principal desvantagem dos métodos dos grafos relacionais é o fato de que muitos atributos são extraídos e agrupados para o reconhecimento de um objeto. Portanto, os algoritmos de casamento usados por esses métodos devem ser mais complexos, do ponto de vista computacional, e conseqüentemente mais lentos do que os algoritmos utilizados pelas classes anteriores. O tempo necessário para resolver o casamento cresce de forma exponencial com os números de nós e arcos. A técnica analítica desenvolvida por Umeyama [56] contorna este problema, produzindo um casamento ótimo, mas apenas para o caso no qual os grafos estão próximos do isomorfismo. Entretanto, uma pesquisa recente realizada por Wen e Lozzi [34], busca reduzir o número de nós e arcos do grafo para reduzir a complexidade computacional na busca do clique máximo. Isto é feito pela divisão sucessiva de um objeto e do modelo em pares de subpolígonos, atentando-se para o casamento entre eles. Utilizam-se somente aqueles subpolígonos cuja seqüência de lados não sofre qualquer falha para a formação de um grafo de associação, onde a

consistência entre os nós é medida por momentos invariantes. Além disso, apenas os maiores lados são examinados durante o processo de casamento entre um objeto e um modelo.

Outra desvantagem dos métodos de grafos relacionais ocorre quando o grafo-modelo é complexo, resultando num grande esforço computacional para a detecção de cliques [29]. Uma tentativa de melhorar o desempenho nestes casos é o uso de técnicas de relaxação como relatado por Bins et al. [11, 57] e Ayache e Faugeras [30]. Estas técnicas permitem o casamento inexato, ou seja, um arco ou nó pertencente a um grafo-teste, não necessita ter correspondência com nenhum outro arco ou nó do outro grafo-modelo. Segundo Li [58], técnicas de relaxação não garantem um casamento ótimo global, ou seja, podem convergir para um ótimo local. Este autor propõe um método cuja descrição dos objetos é baseada em ARS (*attributed relational structure*), derivada da ARG (*attributed relational graphs*). O casamento é feito pelo mapeamento ótimo entre as ARSs da cena e do modelo. Um ganho funcional é formulado a partir das relações invariantes para medir o melhor casamento, maximizado pelo uso do método de rotulação por relaxação. Este método foi empregado pelo autor para casar linhas retas e pontos.

Christmas et al. [59] desenvolveram recentemente os fundamentos teóricos para o processo de relaxação probabilística. Para tanto, o problema do casamento é formulado dentro da estrutura Bayesiana para a rotulação contextual. Esta formulação do problema conduz a uma única e evidente fórmula que prescreve de maneira unificada e consistente como se associam os atributos unários à entidades simples, os atributos binários aos pares de objetos e um resumo do conhecimento anterior do mundo às interações entre objetos. Estas relações são conduzidas como parte do problema de rotulação de objetos. A fundamentação teórica desenvolvida, além de oferecer um melhor entendimento do processo de rotulação por relaxação probabilística, exclui as componentes heurísticas deste processo. Do ponto de vista computacional, oferece uma metodologia clara para a estruturação de tais processos. As principais desvantagens são a natureza iterativa e a escolha dos valores de probabilidades condicionais iniciais, para que o processo seja conduzido. Esta metodologia foi testada em dois diferentes

problemas: no casamento de uma rede de estradas extraídas de uma imagem de um mapa e no casamento de pares de imagens estéreo.

Dois trabalhos recentes realizados por Nasrabadi e Li [38] e por Suganthan et al. [60] apontam para um significativo avanço dos métodos de grafos relacionais, explorando o alto grau de paralelismo das redes neuronais de Hopfield.

O sistema proposto por Nasrabadi e Li produz para cada objeto um grafo-objeto e os integra em um único grafo-modelo global. Durante o reconhecimento procura-se casar um grafo-cena, construído nesta etapa, contra o grafo-modelo global para identificar e localizar a presença de modelos na cena. O problema do casamento de grafos é mapeado sobre uma rede de Hopfield com uma função energia apropriada, representando os nós nos dois grafos a serem casados, para encontrar o melhor casamento. Esta rede é implementada para minimizar a função energia. O estado de cada neurônio representa a possibilidade de casamento entre um nó do grafo-modelo e um nó do grafo-cena (isomorfismo). Os casos considerados para a análise são aqueles em que os objetos se encontram isolados, oclusos e em contato uns com os outros. Segundo os autores, esta técnica pode tolerar uma grande quantidade de ruído, distorção e oclusão, além de incorporar facilmente as restrições de casamento explicitamente na equação de energia, fato que não ocorre com as técnicas de relaxação.

Suganthan et al. [60] propõem um mapeamento ótimo em uma rede de Hopfield para realizar o casamento entre grafos homomórficos ARG (*attributed relational graph*) para criar cenas através do casamento de linhas retas. Ao contrário do método proposto por Nasrabadi e Li, neste caso, quando o problema é mapeado usando pesos estimados para cada modelo, a rede converge rapidamente para um baixo nível de energia.

Os trabalhos relatados indicam que a classe dos métodos dos grafos relacionais ainda continua a se desenvolver, principalmente pelo uso de redes neuronais. Entretanto, um trabalho recente, realizado por Wong [14], ainda emprega a técnica de casamento por subgrafos isomórficos no reconhecimento de objetos em 3D.

As técnicas de pesquisa hierárquica em grafos usadas por Barrow e Tenenbaum [61] podem reduzir o tempo computacional no processo de casamento pela decomposição do modelo em componentes independentes, introduzindo um certo grau de paralelismo.



Comparada com as classes anteriores, os métodos dos grafos relacionais apresentam uma forma de modelagem e de casamento mais complexos. O procedimento de casamento envolve técnicas de pesquisas e casamento de grafos. Estas técnicas são, em geral, computacionalmente mais lentas e ainda sem uma proposta de hardware especial que possibilite seu uso em aplicações industriais, exceção feita para os métodos que utilizam redes neurais.

A partir desta revisão não exaustiva das classes de métodos baseados em modelos para reconhecimento de objetos em imagens 2D, pode-se propor um novo método descrito em seguida.

## **2.4 O método proposto**

Esta tese relata o desenvolvimento de um sistema de visão baseado em modelos poligonais para reconhecimento de formas 2D. Estas formas podem se encontrar isoladas, parcialmente oclusas ou em contato. Este sistema pode ser classificado (subseção 2.3.2) como pertencente à categoria dos métodos de atributos estruturais.

Os modelos poligonais dos contornos são obtidos através de um algoritmo seqüencial descrito com detalhes no capítulo 3. Este algoritmo produz os modelos, durante a fase de treinamento, simultaneamente aos processos de segmentação e de extração de atributos. Isto significa que estas três etapas podem ser integradas em uma única etapa.

Ao decidir pelo esquema de geração e verificação de hipóteses, são adotadas estratégias no sentido de reduzir o número de hipóteses geradas, para reduzir o tempo de avaliação das mesmas. Além disso, criou-se uma estrutura indexada para armazenar os modelos com o intuito de tornar a busca a mais rápida possível, durante o processo de geração de hipóteses.

O número de hipóteses geradas é limitado com base no comprimento médio e no desvio padrão dos segmentos de cada modelo, calculados durante o treinamento. Durante o reconhecimento, este número também é reduzido pelo agrupamento de hipóteses

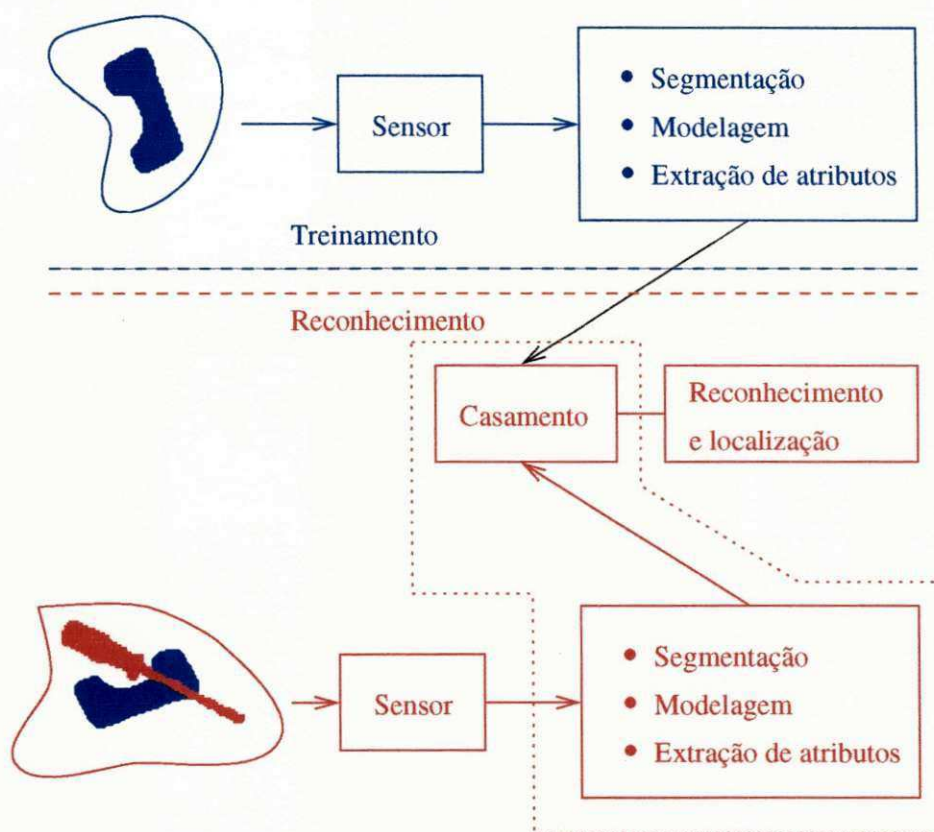


Figura 2.3: Diagrama em blocos do sistema proposto.

semelhantes, segundo uma dada função de similaridade. Isto ocorre durante o processo de geração de hipóteses.

As hipóteses geradas são agrupadas pelo grau de similaridade entre elas e ordenadas conforme o erro relativo ao comprimento dos segmentos casados. Desta forma, hipóteses isoladas serão abandonadas antes de serem verificadas. Isto contribui ainda mais para a eliminação de hipóteses erradas e para diminuir o tempo de verificação de hipóteses.

Durante o reconhecimento, a etapa de geração de hipóteses também é incorporada às demais. Todos os procedimentos são realizados sequencialmente e integrados. A figura 2.3 mostra, em diagrama de blocos os diversos procedimentos integrados pelo sistema.

A estratégia de realizar os diversos procedimentos de forma simultânea, permite

a utilização de uma arquitetura paralela em hardware (funcional) para melhorar o desempenho do sistema como um todo.

O sistema desenvolvido é descrito com mais detalhes nos capítulos seguintes, iniciando-se pelo processo de modelagem e segmentação dos contornos 2D, descritos no capítulo que se segue.

## Capítulo 3

# Modelagem poligonal de contornos bidimensionais

### 3.1 Introdução

Os contornos bidimensionais (2D) de qualquer objeto contém grande quantidade de informação sobre a forma deste objeto, podendo portanto serem utilizados para identificação do mesmo. Conseqüentemente, modelos destes contornos também podem ser usados com a mesma finalidade, ou seja, em processos de reconhecimento de formas de objetos.

Segundo Marshall [6], as principais propriedades, desejáveis em qualquer modelo para uso em sistemas de reconhecimento de formas ou sistemas de visão, são:

1. oferecer uma paridade direta entre as características do modelo e as características dos dados observados,
2. permitir a avaliação direta da posição e orientação do objeto e
3. possibilitar a predição do aparecimento do objeto em qualquer posição.

Além disso, a aproximação de um contorno 2D por modelos, em geral, é feita com o intuito de reduzir a quantidade de dados a serem processados para facilitar sua

manipulação e diminuir o tempo de processamento, podendo até mesmo ser usada em aplicações em tempo real. Em alguns casos, uma taxa de redução de mais de 20 vezes pode ser obtida, ou seja, o número de pontos que compõem o modelo é mais de 20 vezes menor do que o número de pontos do contorno [4, 62, 63].

Muitos métodos para construção de modelos foram propostos durante as últimas décadas, principalmente a partir do final da década passada. Dentre estes podem ser citados os que realizam a transformação do contorno 2D em uma curva 1D, através do uso de alguma propriedade geométrica [42, 43, 44, 45, 46, 47, 39]. Em outros, modelos são formados por pontos críticos e por arcos, detectados de uma curva 1D obtida a partir do contorno 2D, como por exemplo o modelo usado por Tsang et al. [37]. Este, dentre os citados, é o único método que possui flexibilidade para ser usado em um sistema de visão mais complexo, ou seja, aquele que permite o reconhecimento de objetos parcialmente oclusos ou em contato. Nos demais casos, embora os modelos gerados satisfaçam as características enumeradas antes para um modelo desejado, estes não são locais (exceto [37]), ou seja, não permitem descrever os objetos como um conjunto de partes independentes. Esta propriedade é essencial para a obtenção da flexibilidade necessária para utilização em sistemas de visão mais complexos.

Dentre todos os tipos existentes, modelos poligonais (aproximação polinomial de primeira ordem) são os mais utilizados, sobretudo em problemas de reconhecimento de formas, dos quais diversos exemplos existem na literatura [4, 29, 13, 38, 36, 35, 64, 41, 34, 32, 33, 40]. Estes modelos apresentam todas as propriedades desejáveis listadas anteriormente, além de serem:

1. locais, permitindo descrever diferentes partes do contorno do objeto ou cena de forma independente;
2. compactos, em geral a maioria dos objetos pode ser modelado com menos de 100 lados;
3. gerais, podendo ser aplicados à qualquer forma geométrica;
4. sensíveis às variações de posição e orientação;

5. facilmente representados por um conjunto de retas.

Estas propriedades fazem com que o uso de modelos poligonais permita uma fácil manipulação dos dados e uma maior rapidez no processamento e que, segundo Arman and Aggarwal [65], a escolha da estratégia de casamento depende também da representação. Além disso, a representação por modelos poligonais é adequada às estratégias de geração e verificação de hipóteses.

Na seção 3.2 é descrita a etapa de pré-processamento da imagem contendo o objeto a ser modelado. Esta etapa é responsável pela preparação da imagem original para aplicação do algoritmo de segmentação por rastreamento de bordas (*boundary tracking segmentation*), descrito na seção 3.3, o qual extrai da imagem gradiente (resultante do pré-processamento) a lista de pontos que definem o contorno do objeto. Nas seções 3.4 e 3.5 são descritos os aspectos gerais de modelagem poligonal e as classes dos algoritmos, respectivamente. Na seção 3.5.3 é descrito o algoritmo utilizado, sendo que os resultados experimentais são apresentados na seção 3.6. Finalmente a análise destes resultados é feita na seção 3.7 e algumas conclusões tiradas são apresentadas na seção 3.8.

## 3.2 Pré-processamento

A obtenção da lista dos pontos de contorno, através do método de rastreamento de bordas, exige que a imagem original passe por uma etapa de pré-processamento, cujo objetivo é garantir que as bordas do objeto a ser modelado sejam contínuas e bem definidas. Isto significa que a imagem resultante do pré-processamento não poderá possuir falhas de bordas, caso contrário a segmentação será comprometida.

Para obtenção dos resultados apresentados nesta pesquisa, o pré-processamento consistiu em submeter a imagem original sucessivamente aos processos de limiarização, filtragem passa-baixas (média) e cálculo do gradiente.

A técnica de limiarização foi empregada para isolar os objetos, removendo o fundo original da imagem, já que as imagens utilizadas apresentam um histograma bimodal.

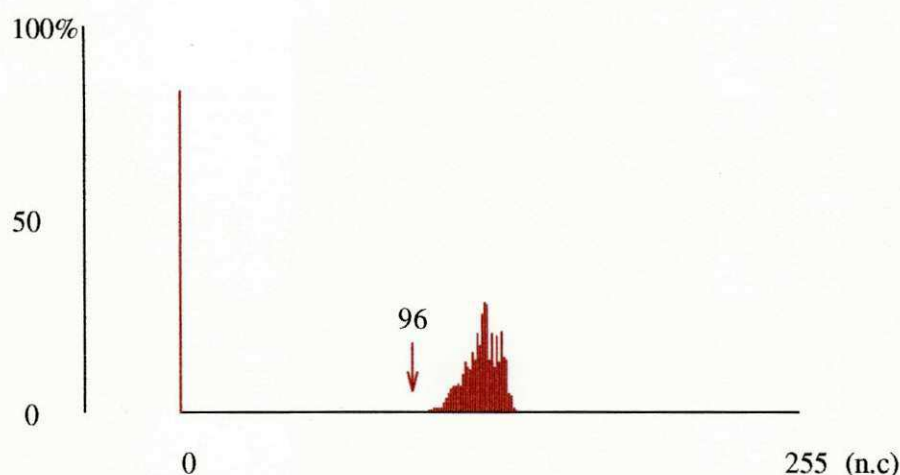


Figura 3.1: Histograma da imagem original da figura 3.2.a.

Para situações mais complexas, técnicas mais sofisticadas estão disponíveis [66, 22]. A figura 3.1 exibe um histograma de níveis de cinza da imagem original (figura 3.2.a) e o valor de limiar empregado pela operação de limiarização desta. Observa-se que o histograma é realmente bimodal, justificando desta forma a escolha desta técnica.

A filtragem passa-baixas foi realizada por convolução no domínio espacial, com a finalidade de suavizar a imagem limiarizada e facilitar a tarefa do detetor de bordas pelo espalhamento destas [8, 67, 68]. Empregou-se a máscara  $M$  ( $3 \times 3$ ), que implementa o filtro da média, dada pela equação que segue,

$$M = \frac{1}{9} \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} \quad (3.1)$$

Em seguida, usou-se um filtro passa-altas implementando o operador gradiente discreto de Roberts para fazer a detecção das bordas do objeto [8, 67]. O valor deste gradiente  $GR(x, y)$  num ponto  $P$  qualquer de coordenadas  $(x, y)$  de uma imagem digital  $I(x, y)$  é obtido como:

$$GR(x, y) = \{[I(x, y) - I(x - 1, y - 1)]^2 + [I(x, y - 1) - I(x - 1, y)]^2\}^{\frac{1}{2}}, \quad (3.2)$$

podendo ser aproximado por:  $GR(x, y) = |I(x, y) - I(x - 1, y - 1)| + |I(x, y - 1) - I(x - 1, y)|$ .

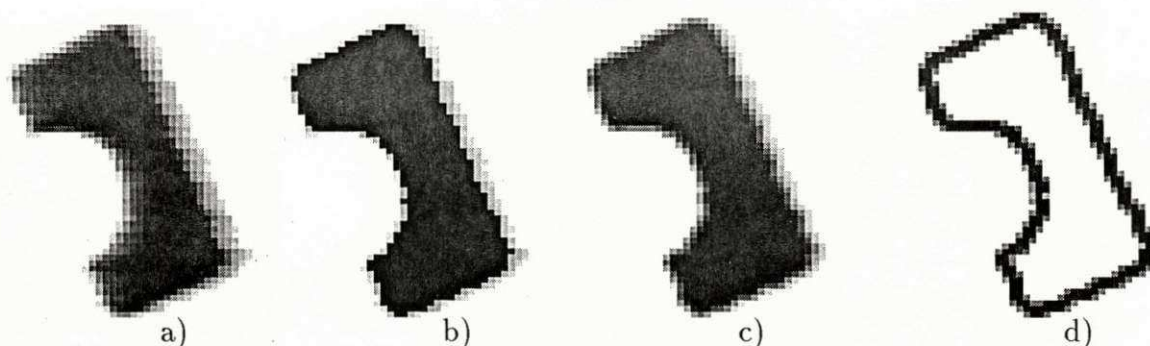


Figura 3.2: Seqüências de imagens obtidas, a partir da imagem original, durante o pré-processamento: a) original, b) limiarizada, c) filtrada e d) gradiente.

A figura 3.2 ilustra um exemplo da aplicação do pré-processamento recém descrito. Figura 3.2.a é a imagem de entrada deste processo, ou seja, a imagem original, figura 3.2.b é a imagem obtida pela limiarização da imagem original, figura 3.2.c é a imagem resultante da aplicação da filtragem passa-baixas e figura 3.2.d é a imagem de saída, ou seja, a imagem gradiente.

Os valores contidos na tabela 3.1 são dados em milisegundos e correspondem ao tempo consumido por cada um dos processos envolvidos pela etapa de pré-processamento, ou seja, limiarização, filtragem passa-baixa e gradiente. Estes valores de tempo foram obtidos tomando-se a média de dez execuções distintas para cada um dos processos. Utilizou-se as imagens de dois objetos: Arco (figura 3.2) e a Fenda.

Objetos	Limiarização (ms)	Filtragem (ms)	Gradiente (ms)
Arco	86.40	132.60	960.40
Fenda	88.10	133.20	1104.60

Tabela 3.1: Resultados do tempo médio consumido por cada um dos processos que constituem a etapa de pré-processamento.

Com a utilização de algoritmos paralelos que operam segundo a técnica de fluxo de dados e a possibilidade de integração dos mesmos em circuitos dedicados, o tempo gasto



pelos processos que compõem a etapa de pré-processamento perdeu sua relevância. Graças ao uso do paralelismo, à medida que a imagem está sendo gerada e a partir de uma certa quantidade de linhas iniciais, os processos mencionados antes já podem ser executados simultaneamente. Desta forma, a cada pulso de relógio um pixel de saída é gerado, restando apenas um atraso inicial no processamento.

A imagem resultante do pré-processamento é submetida ao processo de segmentação descrito a seguir.

### 3.3 Segmentação

Uma vez obtida, a imagem gradiente é submetida ao algoritmo de segmentação, que tem como objetivo extrair a lista dos pontos que compõem o contorno do objeto. A complexidade dos objetos e o tipo de modelo escolhido determinam a natureza da segmentação [6], embora já existam sistemas automáticos para segmentação de imagens complexas, como imagens do cérebro [69]. No presente trabalho, a etapa de segmentação foi implementada por um método do tipo *boundary tracking*, o qual rastreia o contorno do objeto pixel a pixel de forma contínua e seqüenciada [70].

Em relação à forma clássica (tradicional) de segmentação por rastreamento de bordas [70, 6], o algoritmo proposto apresenta três inovações. Estas inovações visam aumentar a flexibilidade, quanto a sua utilização em bordas com variações abruptas de direção e a confiabilidade, quanto à garantia de formar contornos fechados. A primeira destas originalidades se refere à obtenção do ponto inicial de rastreamento, a qual é feita pela varredura das linhas da imagem gradiente, a partir do topo, até encontrar o primeiro ponto de borda (gradiente não nulo). Isto melhora o desempenho do algoritmo, pois elimina a necessidade de varrer toda a imagem em busca do valor de maior gradiente, forma tradicionalmente utilizada, reduzindo portanto o tempo de processamento [64].

A segunda inovação é com relação à escolha do próximo ponto de contorno durante o processo de rastreamento. O algoritmo de rastreamento tradicional trabalha com uma vizinhança  $3 \times 3$  centrada no último pixel rastreado (ponto presente). Para o

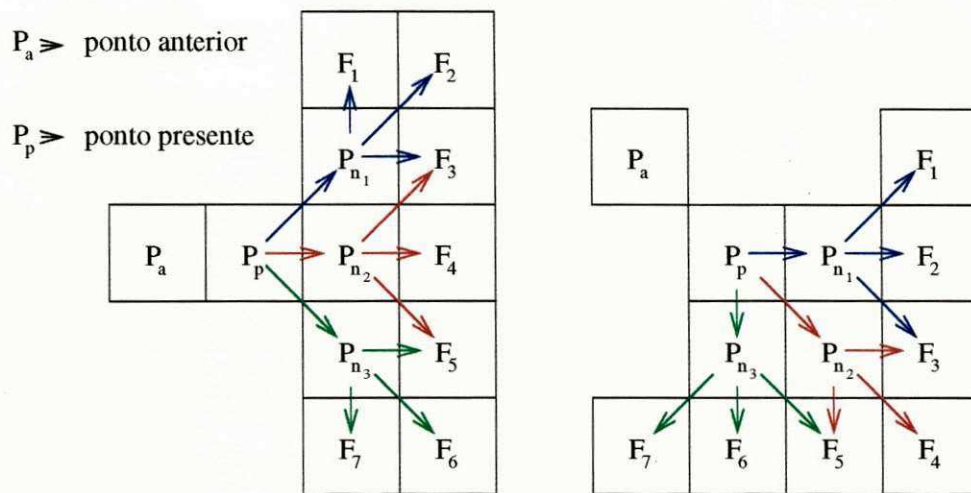


Figura 3.3: Exemplos de vizinhança expandida para rastreamento.

presente trabalho esta vizinhança revelou-se insatisfatória, causando interrupções no rastreamento. Isto ocorre principalmente em regiões da curva de contorno apresentando mudanças abruptas de direção. O problema foi resolvido com a utilização de uma vizinhança expandida, exemplificada na figura 3.3. Nesta figura,  $P_a$  e  $P_p$  são os pontos anterior e presente que definem a direção atual de rastreamento. Os pontos  $P_{n1}$ ,  $P_{n2}$  e  $P_{n3}$  são os candidatos a próximo ponto, simetricamente distribuídos em relação à direção atual  $(\overline{P_a P_p})$ . Esta direção será atualizada para uma das três direções  $\overline{P_p P_{ni}}$ ,  $i = 1, 2, 3$ , dependendo dos pontos simetricamente distribuídos em relação a cada uma delas, ou seja, dos conjuntos  $N_1 = \{F_1, F_2, F_3\}$  para  $P_{n1}$ ,  $N_2 = \{F_3, F_4, F_5\}$  para  $P_{n2}$  e  $N_3 = \{F_5, F_6, F_7\}$  para  $P_{n3}$ . Os pontos  $F_k, k = 1, \dots, 7$  são representados pelos valores de seus níveis de cinza na imagem gradiente. Seja  $S_i, i = 1, 2, 3$  a soma dos níveis de cinza do conjunto  $N_i$  e  $S_M$  a maior dessas somas, isto é,

$$\begin{cases} S_1 = F_1 + F_2 + F_3 \\ S_2 = F_3 + F_4 + F_5 \\ S_3 = F_5 + F_6 + F_7 \\ S_M = \max(S_i). \end{cases} \quad (3.3)$$

Dentre os três candidatos, o próximo ponto de contorno é aquele  $P_{ni}, i = 1, 2, 3$  com maior nível de cinza que satisfaz também a condição de que o respectivo  $S_i$  é maior ou

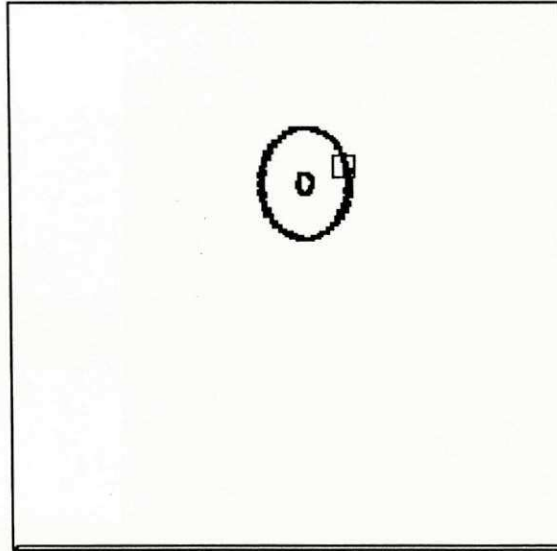


Figura 3.4: A janela mostrada indica uma região crítica para aplicação do algoritmo *boundary tracking* tradicional.

igual a 60% de  $S_M$ .

A terceira inovação do método está na escolha do sentido de rastreamento. No algoritmo *boundary tracking* tradicional este sentido é aleatório, sendo determinado pelos valores dos níveis de cinza dos pixels vizinhos ao pixel inicial. Para o algoritmo aqui proposto, escolheu-se o sentido horário para o rastreamento, a partir do ponto inicial. Para o presente sistema de visão, isto constitui uma vantagem, principalmente pela facilidade de obtenção da orientação dos lados do modelo poligonal em relação a uma referência e, conseqüentemente, do ângulo entre eles. Esta característica também é vantajosa para sistemas de visão nos quais a seqüência de lados do modelo poligonal é de fundamental importância, como os sistemas desenvolvidos por Stein e Medioni [35] e por Maes [32].

O método de segmentação acima descrito, revelou-se muito confiável e eficiente, sendo capaz de seguir o contorno de objetos mesmo em regiões com mudanças abruptas de direção. Para ilustrar esta propriedade do algoritmo *boundary tracking* usando a vizinhança expandida (BTVE) [71] bem como caracterizar suas vantagens em relação ao método *boundary tracking* tradicional (BTT), utiliza-se a figura 3.4, imagem gradiente

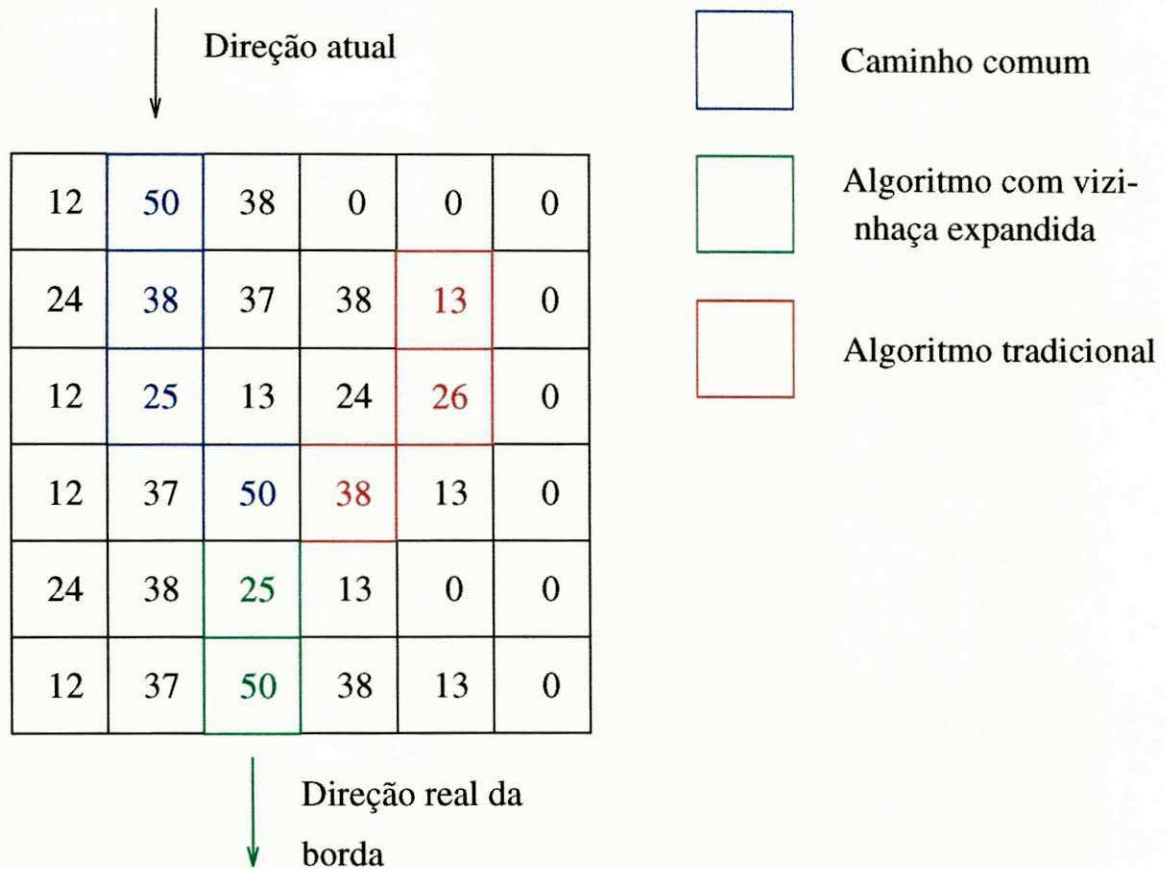


Figura 3.5: Diferentes caminhos percorridos pelos algoritmos para a janela sobre a imagem gradiente mostrada na figura 3.4.

de uma roda dentada. Esta imagem possui uma região crítica para a operação do algoritmo BTT, identificada pela janela retangular mostrada na figura. Para esta janela, a figura 3.5 exibe com detalhes os caminhos percorridos pelos algoritmos BTT e BTVE. Observa-se que o algoritmo tradicional se desvia para uma região que não contém a borda, enquanto o método com vizinhança expandida continua na direção real da borda.

Outra desvantagem do algoritmo BTT, é o seu comportamento quando submetido a uma imagem gradiente contaminada por ruído. Segundo Castleman [70], este algoritmo pode suportar apenas imagens com baixo nível de ruído. Conhecendo o funcionamento do método, torna-se claro sua fragilidade, principalmente se o ponto inicial detectado

(maior valor de gradiente) for um pixel ruidoso. A imagem gradiente contaminada por ruído pode acarretar em um maior número de situações não suportadas pelo método, além daquela descrita acima.

Já o algoritmo BTVE opera satisfatoriamente com imagens gradientes contaminadas por ruído. Para tanto, este método utiliza a vizinhança ampliada para testar a conectividade do ponto inicial, caracterizando se este ponto é ou não ruidoso para que se inicie o processo de rastreamento e impõe a condição de que o valor  $S_M$  da maior soma (equação 3.3) deve ser não nulo, ou seja,  $S_M \neq 0$ . Para esta figura, o contorno obtido pelo algoritmo BTVE permanece quase inalterado quando comparado com o contorno formado a partir da mesma imagem gradiente sem ruído.

A figura 3.6 mostra uma imagem gradiente contaminada por ruído com uma relação sinal/ruído de 3.0 db. Esta imagem é obtida pela equação,

$$\begin{cases} y(i, j) = x(i, j) + f.n(i, j), \\ f = \frac{\tau_x}{\tau_n \cdot \sqrt{10^{0.1RSR_d}}} \end{cases} \quad (3.4)$$

onde  $y(i, j)$  é a imagem contaminada pelo ruído,  $x(i, j)$  é a imagem de entrada com variância  $\tau_x^2$ ,  $n(i, j)$  é uma imagem ruído qualquer com uma variância  $\tau_n^2$  e  $f$  é o fator de degradação que depende da relação sinal ruído desejada ( $RSR_d$ ) [72].

A obtenção do ponto inicial pelo algoritmo BTVE, quando a imagem gradiente se encontra contaminada por ruído, pode aumentar o tempo de processamento deste algoritmo. Isto decorre do teste de conectividade de cada ponto encontrado para sua correta identificação, ou seja, se é ou não um ponto de borda (inicial). Por esta razão, é possível que, neste caso específico, a solução adotada pelo método BTT para a busca do ponto inicial possa vir a ser mais vantajosa.

Dois exemplos de contorno, obtidos pela aplicação do algoritmo de rastreamento de bordas com vizinhança expandida para os objetos Arco e Fenda são mostrados nas figuras 3.7 e 3.8, respectivamente.

Os tempos consumidos, em milisegundos, pelos algoritmos BTT ( $t_{btt}$ ) e BTVE ( $t_{btve}$ ) na obtenção dos contornos mostrados nas figuras 3.7(Arco) e 3.8(Fenda) e a diferença percentual entre os tempos de processamento dos dois algoritmos,  $\Delta_r t\% = \frac{(t_{btve} - t_{btt}) \cdot 100}{t_{btt}}$

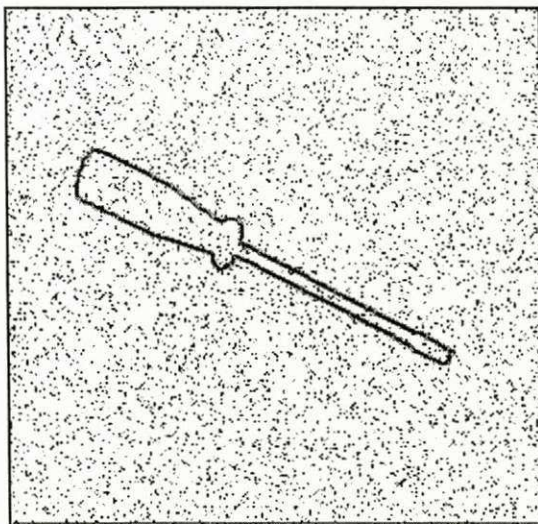


Figura 3.6: Imagem gradiente contaminada por ruído com RSR = 3.0 db.

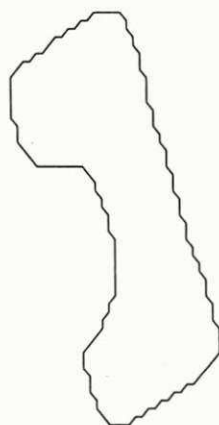


Figura 3.7: Contorno obtido pelo algoritmo BTVE para o objeto Arco (ampliada 2 vezes).

Objetos	$t_{btt}$ (ms)	$t_{btve}$ (ms)	$\Delta_r t\%$
Arco	63.6	69.8	9.75
Fenda	128.5	145.2	13.00

Tabela 3.2: Resultados dos algoritmos BTT e BTVE para os contornos das figuras 3.7 e 3.8.

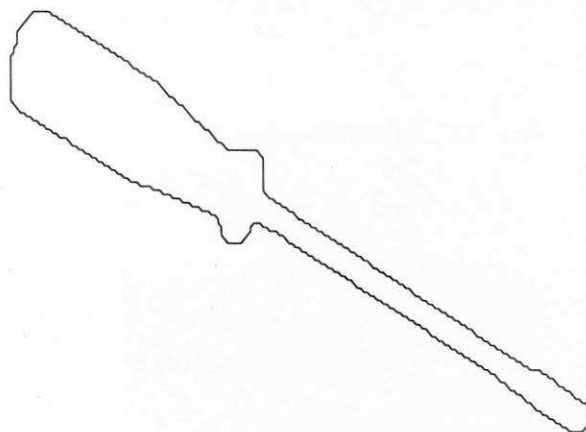


Figura 3.8: Contorno obtido pelo algoritmo BTVE para o objeto Fenda (ampliada 1,5 vezes).

são mostrados na tabela 3.2. Os resultados apresentados indicam que o algoritmo BTVE consome, em média, 11.50% mais tempo do que o BTT para os dois contornos obtidos. Este comportamento era esperado, uma vez que o algoritmo BTVE utiliza cálculos sobre uma vizinhança maior, aumentando com isto o número de operações realizadas.

Outro parâmetro importante que pode ser extraído dos dados mostrados na tabela 3.2 é a sensibilidade dos algoritmos em relação ao aumento do número de pontos de contorno. Esta sensibilidade está relacionada com a complexidade dos algoritmos [73]. O presente caso, o aumento do número de pontos de contorno do objeto da figura 3.8, com 126 pontos, em relação ao objeto da figura 3.7, com 387 pontos, é um pouco mais de 200%. Os acréscimos correspondentes no tempo de processamento dos algoritmos BTT e BTVE são 102% e 108%, respectivamente. Isto resulta numa diferença de 6% a favor do algoritmo BTT, comprovando a maior complexidade do algoritmo BTVE.

Resta analisar como os métodos, BTT e BTVE resolvem o problema de convergência, de modo a garantir a formação de um contorno fechado. Para o primeiro método, isto é assegurado pelo fato de que o ponto inicial é o pixel de maior valor de nível de cinza encontrado na imagem gradiente. Deste modo, fica garantido o retorno obrigatório a este ponto, uma vez que o método seleciona, durante sua execução, sempre o próximo ponto com maior valor de gradiente. Para o algoritmo BTVE, quando

realizado simultaneamente com o processo de modelagem (a ser descrito na subseção 3.5.3), a convergência é assegurada fazendo com que o vértice inicial passe a constar como o ponto inicial de rastreamento. Como este vértice foi visitado antes pelo método, para ser considerado como tal, necessariamente fará parte do caminho traçado. Quando o algoritmo BTVE é utilizado isoladamente, faz-se necessário percorrer pelo menos 4 pixels (comprimento da vizinhança expandida), a partir do primeiro ponto de borda detetado para se escolher o ponto inicial, assegurando desta forma a construção de um contorno fechado.

Os resultados apresentados nesta seção atestam a superioridade do método BTVE em relação ao método BTT e mostram que o mesmo permite obter a lista dos pontos de contorno sob as mais variadas condições.

À medida que o processo de segmentação está sendo executado, os pontos de contorno obtidos podem ser submetidos imediatamente ao processo de modelagem poligonal. Portanto, os processos de obtenção da lista de pontos do contorno e de modelagem poligonal deste contorno podem ser realizados numa única etapa.

### 3.4 Modelagem poligonal

Um contorno bidimensional pode ser representado por uma curva fechada, constituída pelos pontos de contorno. Transformar esta curva em um polígono simples<sup>1</sup>, significa dividir o conjunto de pontos do contorno em vários subconjuntos, correspondentes aos segmentos de reta (ou lados) do polígono. Cada um destes subconjuntos pode ser representado apenas pelos dois vértices necessários à formação do segmento de reta correspondente (referido a partir deste ponto do texto apenas como *segmento*).

Considere a curva fechada  $\mathcal{C}$  representando uma lista de pontos de contorno, mostrada na figura 3.9. Considere que estes pontos são os elementos do conjunto  $C = \{a_1, a_2, \dots, a_n\}$ , onde  $a_1 = a_n$  e  $n \in \mathcal{N}^*$  ( $\mathcal{N}^*$  é o conjunto dos números naturais estritamente positivos). Este conjunto é obtido durante o processo de segmentação descrito

---

<sup>1</sup>Polígonos simples são aqueles cujos lados não se cruzam, a não ser nos vértices [73].



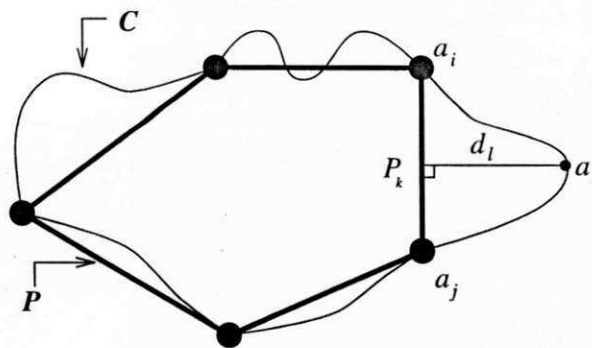


Figura 3.9: Aproximação poligonal da curva  $C$  pelo polígono  $P$ .

na seção 3.2. Deseja-se aproximar a curva  $C$  por um polígono  $\mathcal{P}$  (figura 3.9) cujos vértices pertencem ao conjunto  $V = \{a'_1, a'_2, \dots, a'_{n'}\}$ , onde  $V \subseteq C$  e  $n' \leq n$ ,  $n' \in \mathcal{N}^*$ . Esta aproximação deve obedecer a um determinado critério e é limitada por um erro máximo  $\varepsilon$ .

Seja  $S_k$  um subconjunto de  $C$  cujos elementos pertencem ao intervalo  $I_k = [a_i, a_j]$  com  $a_i = a'_k$ ,  $a_j = a'_{k+1}$  e  $k = 1, \dots, n' - 1$ . Desta forma, o critério de aproximação pode ser representado pela função,

$$f(S_k) \leq \varepsilon, \quad (3.5)$$

ou seja, todo elemento  $a_m \in S_k$ , com  $i < m < j$ , deve satisfazer a equação 3.5.

A definição de uma aproximação poligonal para a curva  $C$ , implica em que o subconjunto  $S_k$  é aproximado por um segmento  $P_k$ , definido pelos pontos  $a_i$  e  $a_j$ , ou seja,  $P_k = \overline{a_i a_j}$ . A figura 3.9 exhibe este processo para um determinado ponto  $a_l \in C$ , utilizando-se a distância euclidiana  $d_l$ , entre o ponto  $a_l$  e o segmento  $P_k$ , como critério de aproximação, na equação 3.5.

A curva  $C$  da figura 3.9 pode representar um contorno 2D de um determinado objeto. Sendo assim, a aproximação poligonal de contornos 2D é, na verdade, a modelagem da forma do objeto representada por aquele contorno [6]. Para resolver o problema de modelagem de contornos 2D foram desenvolvidos diversos tipos de algoritmos, alguns deles descritos na seção seguinte.

### 3.5 Algoritmos para a modelagem poligonal de contornos 2D

Do ponto de vista operacional, algoritmos para modelagem poligonal podem, em geral, ser agrupados em três classes: iterativos, seqüenciais e paralelos.

Algoritmos iterativos usam os dados de uma forma global, necessitam de uma dada solução inicial e de algum critério de convergência e são menos sujeitos a problemas decorrentes da digitalização e do pré-processamento da imagem. Entretanto, o desempenho destes algoritmos é dependente da qualidade da aproximação inicial fornecida e, portanto, podem necessitar de um grande número de iterações para convergir. Como exemplo, podem ser citados os algoritmos clássicos de Ramer [74] e Split-Merge [75].

Os algoritmos seqüenciais constroem o modelo ao longo da curva de contorno do objeto, sendo, em geral, mais simples, rápidos e quase todos usam uma forma de código direcional como base do processo de modelagem. Entretanto, algumas regras heurísticas utilizadas por estes algoritmos são, muitas vezes, de difícil compreensão. Para ajustar a aproximação necessita-se, em geral, ajustar também estas regras, o que os tornam menos flexíveis. Um exemplo é o algoritmo de Kung-Kaswand [76]. Outros, embora fáceis de entender, não possuem mecanismos que permitam controlar a precisão dos modelos formados, como, por exemplo, o método de Ray-Ray [77]. Adicionalmente, a eficiência destes algoritmos depende da qualidade da imagem resultante do pré-processamento e da segmentação, o que os torna mais sensíveis à ruídos ambientais e a problemas de digitalização. Além dos algoritmos já citados, outros exemplos de algoritmos seqüenciais são os métodos desenvolvidos por Cortez [78, 79], como parte do trabalho descrito neste documento, e um mais recente por Zhu e Chirlian [80].

Algoritmos paralelos são aqueles que necessitam apenas de uma quantidade mínima de linhas da imagem que contém a curva a ser modelada para iniciar sua operação. Estes algoritmos podem operar à medida que esta imagem está sendo gerada ou mesmo adquirida por varredura. Esta classe de algoritmos é apropriada para aplicações em tempo real por permitirem uma maximização da velocidade de operação. Um exemplo deste tipo de algoritmo é o método desenvolvido por Cheng-Hsu [62].

A seguir são descritos, de forma resumida, os algoritmos estudados neste trabalho, ou seja, os algoritmos iterativos de Ramer [74] e Split-Merge [75], os métodos seqüenciais desenvolvidos por Cortez [64], por Ray-Ray [77], um método utilizando a transformada de Hough [79], o algoritmo de Kung-Kaswand [76] e o método paralelo de Cheng-Hsu [62].

### 3.5.1 Algoritmo de Ramer

Este algoritmo adota como solução inicial a própria curva a ser modelada, ou seja, fazendo  $S_1 = C$ . A seguir verifica se  $f(S_1) > \varepsilon$ , se for divide  $S_1$  em dois subconjuntos  $S_1$  e  $S_2$ , onde  $S_1 \cap S_2 = a_i$ , com  $a_i \in C$ , tal que  $f(a_i)$  é máximo. Este processo continua até que o critério de aproximação  $f(S_k) \leq \varepsilon$  seja satisfeito para todo  $k = 1, \dots, n'$ .

### 3.5.2 Algoritmo Split-Merge

O algoritmo Split and Merge é muito citado na literatura pela sua eficiente precisão na construção de modelos [75]. Este método adota duas estratégias, uma para dividir os intervalos (*Split*) e a outra para uma possível união de intervalos consecutivos (*Merge*). O procedimento Split é idêntico ao de Ramer, descrito acima, sugerindo a descrição apenas do procedimento Merge.

A estratégia de união de dois subconjuntos  $S_k$  e  $S_{k+1}$  é verificar se todos os elementos pertencentes ao subconjunto  $S_{k'} = S_k \cup S_{k+1}$  também satisfazem a um dado critério de aproximação. Na prática, verifica-se esta condição apenas para o elemento resultante da interseção dos dois subconjuntos, pois, este é o elemento mais distante do segmento de reta formado pelos elementos inicial e final do subconjunto  $S_{k'}$ . Outra razão é que qualquer outro elemento pertencente a qualquer um dos dois intervalos que não satisfaça a um determinado critério de aproximação será encontrado pelo procedimento Split. Desta forma a união de dois subconjuntos pode ser controlada apenas pelos elementos inicial, intermediário (interseção dos subconjuntos) e final.

### 3.5.3 Algoritmo de Cortez

O algoritmo de modelagem poligonal proposto a seguir é um algoritmo seqüencial que possui apenas uma regra simples para a escolha dos pontos correspondentes aos vértices do polígono. Esta regra permite o controle da precisão (ou fidelidade) do modelo poligonal através de uma série de parâmetros operacionais, não sendo modificada quando estes parâmetros são alterados. Adicionalmente, o método descrito permite que as operações de segmentação, construção do modelo e extração de características geométricas sejam realizadas simultaneamente, o que o torna especialmente adequado para implementações em sistemas de processamento paralelo.

Inicialmente, faz-se necessário definir parâmetros e critérios para uso do algoritmo que são:

- a) a distância máxima permitida  $L_d$  entre um ponto do contorno e o segmento anterior, que será usada para determinar os vértices do modelo. A escolha de  $L_d$  é uma função da precisão desejada. Para este trabalho foi considerado  $L_d$  igual a 2 pixels;
- b) o critério para seleção dos pontos que serão os possíveis vértices do modelo poligonal. Um exemplo de critério é a variação no código direcional entre pixels consecutivos, o qual foi utilizado para obter os resultados experimentais aqui mostrados;
- c) o número  $N$  de pixels que deverá ser percorrido por vez no processo de rastreamento, antes de ser aplicado o critério definido no item b. Este parâmetro influi na precisão do modelo. Para este trabalho utilizou-se  $N$  igual a 2 pixels;
- d) um critério para eliminar vértices comuns a dois segmentos quase alinhados. Este critério também influi na precisão do modelo e pode ser definido em função da forma geral dos objetos utilizados.

Considerando o processo em andamento,  $P_i$  é definido como o último ponto selecionado como possível vértice do polígono, ou seja, o ponto atual. Os pontos que definem

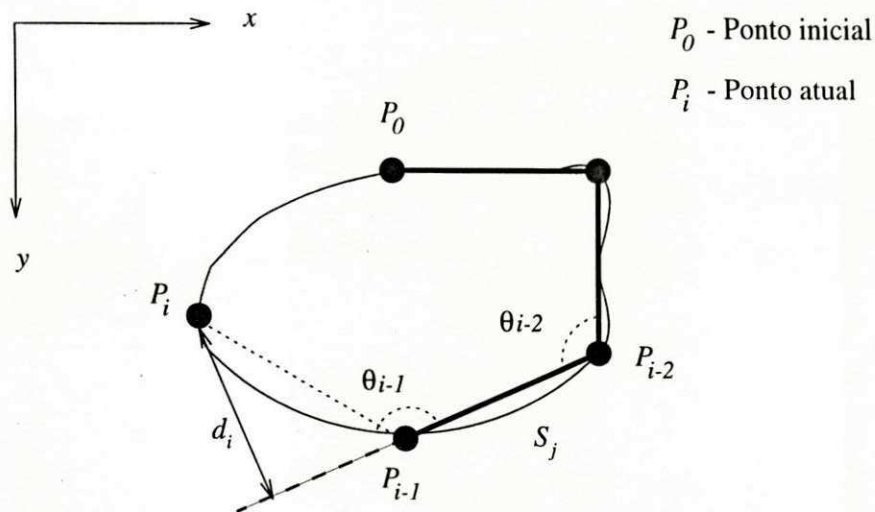


Figura 3.10: Esquema para entendimento do algoritmo.

o último segmento anteriormente detectado são  $P_{i-1}$  e  $P_{i-2}$ . Portanto,  $P_{i-2}$  é o último vértice detectado. O algoritmo utilizado se baseia na distância entre  $P_i$ , o ponto atual, e o segmento formado pelos pontos  $P_{i-2}$  e  $P_{i-1}$ , como ilustra a figura 3.10.

Descrição do algoritmo:

1. **início:** considera-se temporariamente o ponto  $P_0$ , como o vértice inicial, a partir do qual se inicia o processo de rastreamento e modelagem. Onde  $P_0$  é o primeiro ponto de borda obtido pela varredura da imagem gradiente para iniciar o processo de segmentação (seção 3.2). Rastrea-se a borda do objeto até eleger o ponto  $P_1$ , conforme os critérios já definidos no item b) acima, que juntamente com  $P_0$  comporá o segmento inicial  $S_0$ ;
2. a partir do ponto atual, rastrea-se  $N$  ou mais pontos, de acordo com o critério definido no item (b), até eleger o próximo ponto candidato a vértice  $P_i$ ;
3. calcula-se  $d_i$ , a distância do segmento  $\overline{P_{i-2}P_{i-1}}$  ao ponto atual  $P_i$ . Se  $d_i$  for maior do que  $L_d$  (figura 3.10), então,
  - (a)  $P_{i-1}$  será um vértice,  $V_j$ , o segmento  $S_j$  é atualizado para  $S_j = \overline{P_iP_{i-1}}$ , retorna ao passo 2;

senão,  
(b) retorna ao passo 2.

4. A partir de três vértices formados, pode-se tentar otimizar o modelo, eliminando o vértice central, dependendo do critério fixado no ítem (d);
5. **fim**: o processo de modelagem termina juntamente com a segmentação da imagem gradiente quando for encontrado o vértice inicial.

A figura 3.11 mostra o fluxograma do algoritmo de modelagem poligonal quando este opera junto com o processo de rastreamento.

Para unir segmentos consecutivos quase alinhados (ítem d desta subseção), adotou-se como critério a distância euclidiana entre o vértice comum aos dois segmentos e a semi-reta unindo os dois vértices extremos. O valor do limiar usado neste processo foi igual ao valor de  $L_d$  já definido. Este critério é usado para o método Cortez 2, variante do método Cortez 1, cujo critério empregado para a eliminação dos vértices é a variação angular entre os segmentos, com um limiar  $L_\theta = 10^\circ$ .

A figura 3.12 ilustra a execução do algoritmo descrito acima e a influência do valor do limiar  $L_d$  na construção de modelos, através da modelagem de um octógono regular. Figura 3.12.a mostra a seqüência de pontos do contorno do octógono e os correspondentes valores de distância calculados pelo algoritmo. Como resultado, podem ser obtidos como modelos um quadrado (linha pontilhada) para  $L_d = 2$  ou um octógono (linha contínua) para  $L_d = 1$ . Estas situações são mostradas na figura 3.12.b.

Além de produzir os modelos, o algoritmo descrito permite o cálculo dos atributos necessários ao reconhecimento dos objetos que possam estar presentes em uma determinada cena. Deste modo, simultaneamente com a construção dos modelos, são obtidos os seguintes atributos destes modelos: pontos médios entre os vértices, comprimento dos lados do polígono, orientação de cada lado (dada pelo ângulo entre o lado e o eixo- $x$  positivo), ângulo interno nos vértices e centróide do modelo. A figura 3.13 mostra estes atributos para dois lados de um modelo.

Os atributos são escolhidos em função do conjunto de objetos a serem reconhecidos e da estratégia empregada no processo de avaliação de hipóteses. Dependendo da

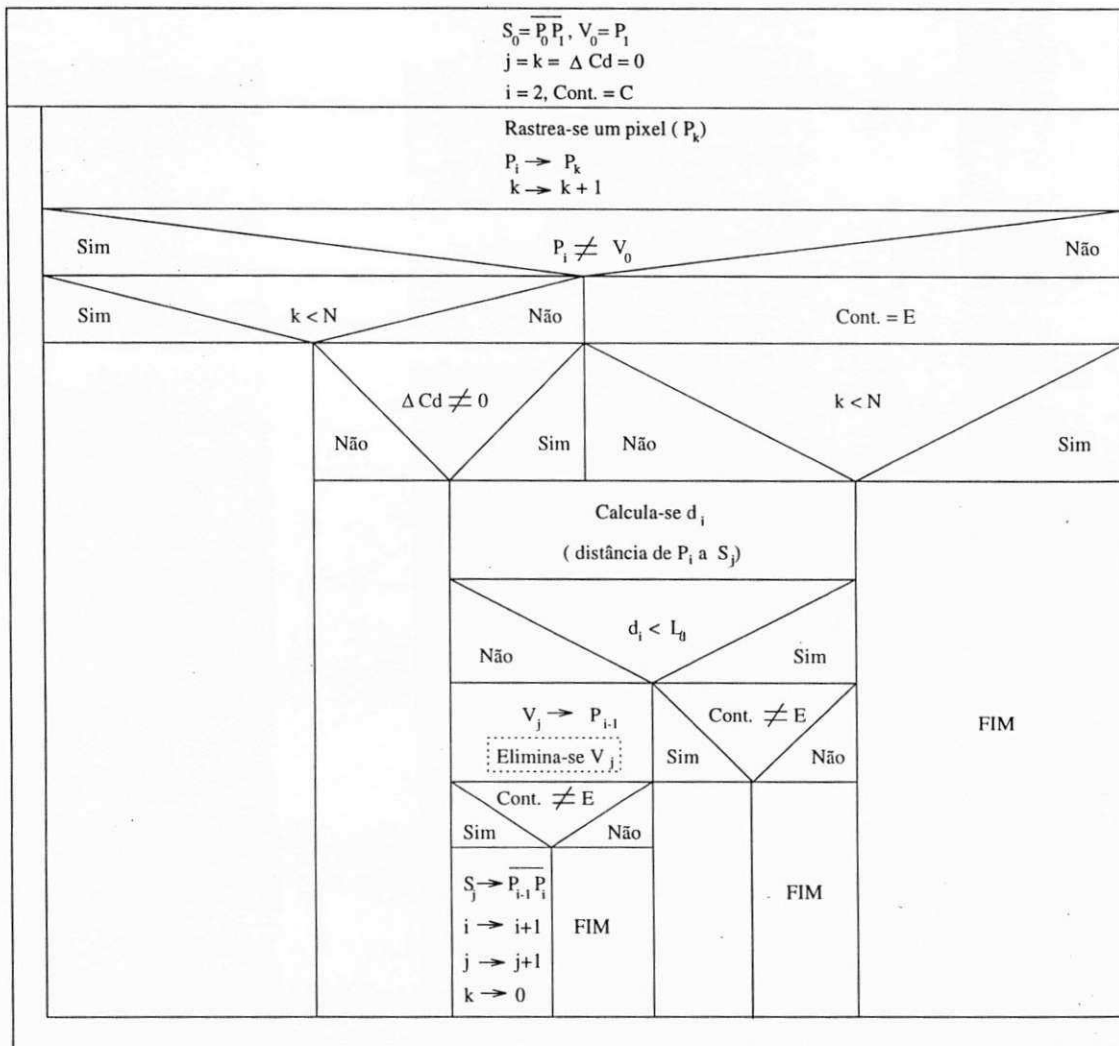
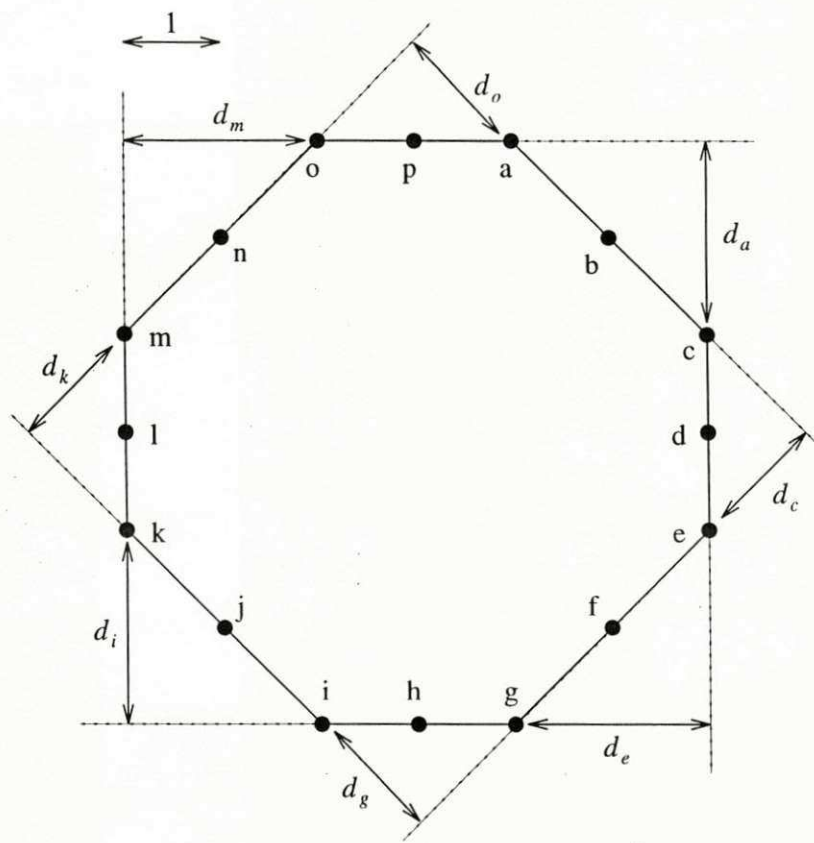
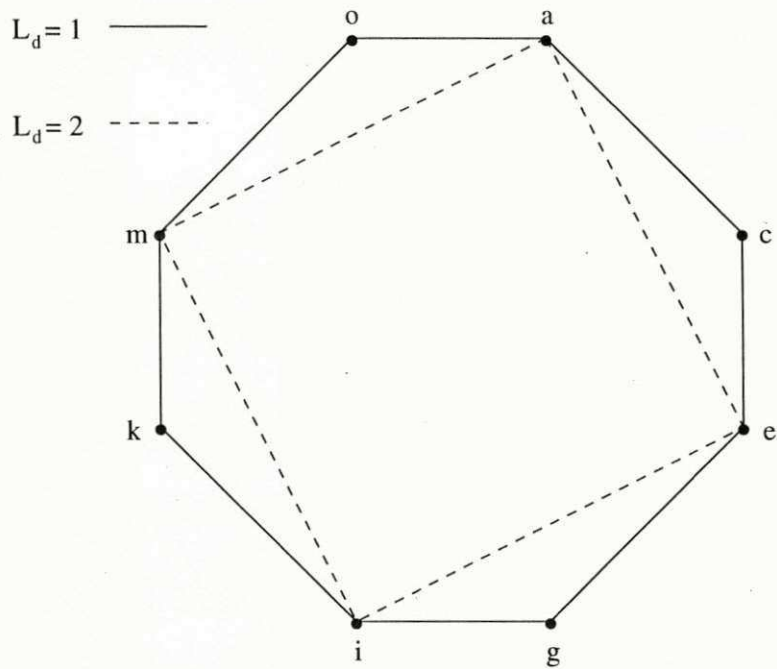


Figura 3.11: Fluxograma do algoritmo Cortez.



a) Testando possíveis vértices.



b) Modelos poligonais obtidos para  $L_d = 1$  e  $L_d = 2$ .

Figura 3.12: Modelagem de um octógono regular.



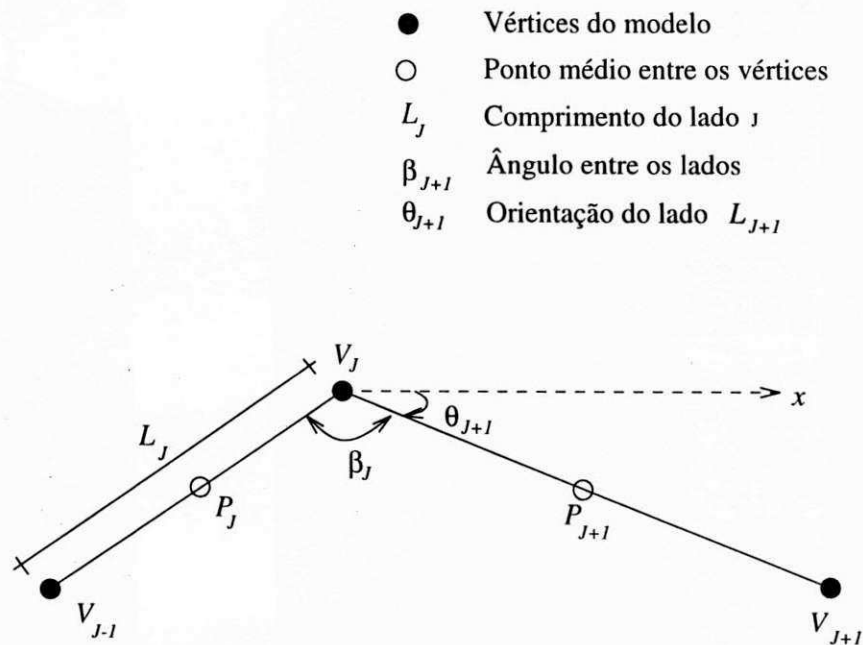


Figura 3.13: Atributos extraídos dos modelos.

complexidade dos objetos (furos de várias formas) e das características discriminantes entre eles, podem ser usados outros atributos ou incorporar no processo de modelagem os furos contidos no interior destes objetos.

### 3.5.4 Algoritmo de Ray-Ray

O algoritmo descrito nesta subseção procura encontrar um polígono ótimo, maximizando uma função objetiva dada pela diferença entre o comprimento dos segmentos retos e a soma absoluta dos erros entre estes segmentos e os respectivos pontos do contorno fechado. Em outras palavras, este algoritmo determina os maiores segmentos com a minimização da soma dos erros absolutos ao longo da curva, caracterizando-se desta forma como um algoritmo seqüencial [77].

### 3.5.5 Algoritmo de Cheng-Hsu

O funcionamento deste algoritmo tem como princípio a varredura linha a linha de uma imagem binária. Esta imagem binária é composta por pontos de borda com níveis de cinza iguais a 1 e os demais pontos com níveis de cinza 0 [62]. Para este trabalho esta imagem binária foi construída a partir da lista dos pontos de borda dos objetos.

Muitas propriedades geométricas podem ser calculadas através de uma máscara  $3 \times 3$  de uma vizinhança-8. Porém esta não possui informação suficiente sobre a curvatura local de um determinado ponto para classificá-lo como um vértice, pois sua resolução é múltipla de  $45^\circ$ . Por conseguinte, faz-se necessário utilizar máscaras maiores como  $5 \times 5$  ou  $7 \times 7$ . Entretanto, o aumento das dimensões da máscara produz uma explosão combinatorial, inviabilizando-as computacionalmente. Por isto, o método proposto por Cheng-Hsu utiliza uma máscara  $3 \times 3$  estendida, que pelo seu deslocamento, para frente e para trás, resulta num ganho considerável de informação sobre a curvatura local, possibilitando dessa forma a localização dos vértices [62].

O algoritmo de Cheng-Hsu se baseia no código direcional de bordas, podendo ser dividido em duas fases. A primeira é responsável pela pré-classificação dos pontos em pontos internos ou não ao contorno. A segunda fase, utiliza o deslocamento de uma máscara estendida, em relação a um ponto do contorno, para a localização dos vértices.

A pré-classificação dos pontos, em pontos internos durante a primeira fase, é feita através do número de pixels com valores diferentes de zero vizinhos do pixel a ser classificado e do código direcional destes pixels.

A segunda fase, responsável pela localização dos vértices, usa o deslocamento ou passo de uma máscara  $3 \times 3$ , para frente (*forward*) e para trás (*backward*) sobre os pontos de borda. Este deslocamento pode ser feito uma, duas ou mais vezes conforme desejado, caracterizando-se como 1-passo, 2-passos, 3-passos ou  $k$ -passos. Tomando-se como base o deslocamento desta máscara estendida ( $k$ -passos)  $3 \times 3$ , os autores (Cheng e Hsu) definem as seqüências direcionais para frente e para trás,  $\{d_i^f\}_{i=1}^k$  e  $\{d_i^b\}_{i=1}^k$  respectivamente. As seqüências direcionais assumem os valores do código direcional de borda de cada ponto. A partir destas, são definidas também as anti-direções  $\tilde{d}_1^f$  e  $\tilde{d}_1^b$ .

como função daquelas. Finalmente, o cálculo da curvatura é realizado por operações sobre estas seqüências e as anti-direções [62].

No presente trabalho, foi desenvolvida também uma versão seqüencial para este algoritmo que será referenciada por Cheng-Hsu-S para diferenciar do método original chamado de Cheng-Hsu-O [81].

### 3.5.6 Algoritmo que utiliza a transformada de Hough

O método aqui descrito para modelagem poligonal de contornos 2D se baseia na idéia dos passos para frente e para trás contida no algoritmo de Cheng-Hsu [62], bem como na transformada de Hough, para o cálculo da curvatura local dos pontos do contorno. Este método foi desenvolvido como parte desta tese, sendo portanto original [79].

O método proposto por Cheng-Hsu utiliza o deslocamento ou passo de uma máscara  $3 \times 3$ , para frente (*forward*) e para trás (*backward*) sobre os pontos do contorno a ser modelado. Este deslocamento pode ser feito uma, duas ou mais vezes conforme desejado, caracterizando-se como 1-passo, 2-passos, 3-passos ou  $k$ -passos. Isto resulta num ganho considerável de informação sobre a curvatura local, possibilitando a localização dos vértices do modelo poligonal [62].

Os pontos de contorno sobre os quais as máscaras se deslocam para frente e para trás são mapeados para o espaço de Hough, onde estes pontos são representados por senóides. As interseções destas senóides determinam ângulos ou direções. Baseado na diferença destas direções o algoritmo encontra os vértices do polígono. Estas interseções são calculadas pelo método de Newton [79].

### 3.5.7 Algoritmo de Kung-Kaswand

A idéia básica deste algoritmo é usar apenas as informações contidas no código direcional de bordas. Nenhum controle da precisão do modelo está previsto neste algoritmo, o que restringe sobremaneira sua aplicação.

Considere uma seqüência de  $n$  pontos conectados  $L(S(i))$ , representando um contorno bidimensional qualquer, onde  $S(i)$  é o código direcional do  $i$ -ésimo ponto desta seqüência. Deseja-se detectar os “pontos críticos” ao longo da seqüência  $L$  de tal forma que esta seja aproximada por um conjunto de segmentos retos consecutivos [76].

Seja  $S(i)$  o código direcional do  $i$ -ésimo pixel da seqüência  $L$ . O código diferença para o  $i$ -ésimo pixel é definido como:

$$\begin{aligned} C(i) &= S(i) - S(i - 1), \quad i = 2, \dots, n \text{ módulo } 8, \text{ com} \\ C(1) &= 0, \end{aligned} \quad (3.6)$$

$C(i)$  indica a mudança de direção da seqüência  $L$ , isto é, a própria quantização da curvatura no pixel  $i$ . O código diferença, gerado desta forma, possui regiões nas quais seus valores são não nulos. Estas regiões são chamadas de NZS's (*non-zero segment*). Todos os outros parâmetros inerentes ao método são baseados nas NZS's [76, 82].

Este algoritmo se compõe de duas etapas, uma na qual são encontrados os pontos críticos iniciais (processamento das NZS's) e uma outra para localizar novos vértices entre os já detectados na primeira etapa mas que não foram selecionados por causa dos cancelamentos entre os pares 1, -1 ou -1, 1 no código diferença. A descrição completa deste algoritmo é feita no trabalho de Kung e Kaswand [76].

### 3.6 Resultados experimentais

Os resultados mostrados nesta seção foram obtidos com imagens reais de dimensões  $256 \times 252$  com 256 níveis de cinza. Duas imagens foram selecionadas para a construção dos modelos para apresentação, embora testes tenham sido realizados com mais de dez imagens das mais diferentes formas.

Foram implementados sete algoritmos distintos, entre os quais o método proposto neste trabalho, chamado de Cortez 1 [78]. Para este método e para o método de Cheng-Hsu também foram implementadas uma variante para cada um, chamadas de Cortez 2 [81] e Cheng-Hsu-S (seqüencial) [81], respectivamente. Para este último, utilizou-se máscaras de tamanho 2, 3 e 4 para ambas as versões, a original e a seqüencial.

Os algoritmos foram implementados em uma estação SUN SPARC-2 e executados sem nenhuma prioridade para este processo. Por esta razão foi considerado, para efeito de análise, o tempo médio de execução de cada algoritmo, calculado sobre dez execuções distintas.

O tempo de execução  $t_i$ ,  $i = 1, 2$  considerado é o tempo de obtenção da lista de pontos de contorno<sup>2</sup>, adicionado ao tempo de construção do modelo poligonal a partir desta lista. A lista é obtida pelo método de segmentação descrito na seção 3.2, exceto para o método original Cheng-Hsu. Como este método não utiliza a lista dos pontos de borda no processo de modelagem, leva-se em conta apenas o tempo de modelagem, a partir da imagem binária contendo o contorno a ser modelado.

O erro  $e_i$ ,  $i = 1, 2$  que será mostrado nas tabelas é o erro médio quadrático da aproximação, calculado a partir da distância de cada ponto  $a_i$  da borda do objeto ao correspondente segmento de reta do modelo poligonal (figura 3.9).

As tabelas exibem também a taxa de compressão dos dados  $\tau_i$ ,  $i = 1, 2$ , dada pela relação entre o número de pontos do contorno e o número de vértices ou de lados do modelo  $l_i$ ,  $i = 1, 2$ .

As razões  $t_2/t_1$  entre o tempo de processamento de cada algoritmo para as duas imagens utilizadas nos testes, estão incluídos na Tabela 3.4. Estas razões indicam a sensibilidade de cada algoritmo ao aumento do número de pontos da lista para as imagens usadas. Quanto maior for esta razão, maior é a sensibilidade do respectivo algoritmo. Tais razões estão relacionadas com a complexidade dos algoritmos [73].

As figuras 3.14 a 3.19, mostram os modelos produzidos pelos métodos Cortez 2 [64, 78], denominado apenas por Cortez por todo o resto do trabalho, Ramer<sup>3</sup> [74] e Split-Merge [75]. Os dois últimos métodos foram escolhidos para comparação porque são considerados clássicos, sendo utilizados em diversos sistemas de visão, propostos por outros autores [13, 38, 33]. Além disso, os demais algoritmos seqüenciais, mostrados nas tabelas, produzem modelos poligonais com muitos lados, não sendo portanto adequados

---

<sup>2</sup>Este tempo é igual para todos os algoritmos, pois, utilizou-se o algoritmo BTVE neste processo.

<sup>3</sup>Segundo Duda and Hart [83] este método foi inicialmente sugerido por G. E. Forsen [75].

Métodos	$t_1$ (ms)	$l_1$	$e_1$ (pixels)	$\tau_1$
Ramer	115	12	0.40	10.50
Split-Merge	135	12	0.40	10.50
Cortez 1	006	15	0.33	08.40
Cortez 2	008	09	0.43	14.00
Kung-Kaswand	005	22	0.22	05.73
Ray-Ray	008	17	0.26	07.41
T. Hough	114	11	0.51	11.45
Cheng-Hsu(2)-O	104	08	1.02	15.75
Cheng-Hsu(2)-S	14	08	1.02	15.75
Cheng-Hsu(3)-O	109	17	0.27	7.41
Cheng-Hsu(3)-S	21	17	0.27	7.41
Cheng-Hsu(4)-O	112	24	0.20	5.25
Cheng-Hsu(4)-S	25	24	0.20	5.25

Tabela 3.3: Resultados dos algoritmos para o objeto Arco com 126 pontos.

Métodos	$t_2$ (ms)	$l_2$	$e_2$ (pixels)	$\tau_2$	$t_2/t_1$
Ramer	270	18	0.56	21.50	2.35
Split-Merge	305	14	0.55	27.64	2.26
Cortez 1	008	17	0.56	22.76	1.33
Cortez 2	010	12	1.11	32.25	1.25
Kung-Kaswand	007	26	0.69	14.88	1.40
Ray-Ray	040	36	0.50	10.75	5.00
T. Hough	292	20	1.75	19.35	2.56
Cheng-Hsu(2)-O	132	14	1.77	27.64	1.18
Cheng-Hsu(2)-S	41	14	1.77	27.64	2.93
Cheng-Hsu(3)-O	153	28	0.64	13.82	1.30
Cheng-Hsu(3)-S	65	28	0.64	13.82	3.10
Cheng-Hsu(4)-O	169	35	0.61	11.06	1.30
Cheng-Hsu(4)-S	80	35	0.61	11.06	3.20

Tabela 3.4: Resultados dos Algoritmos para o objeto Fenda com 387 pontos.

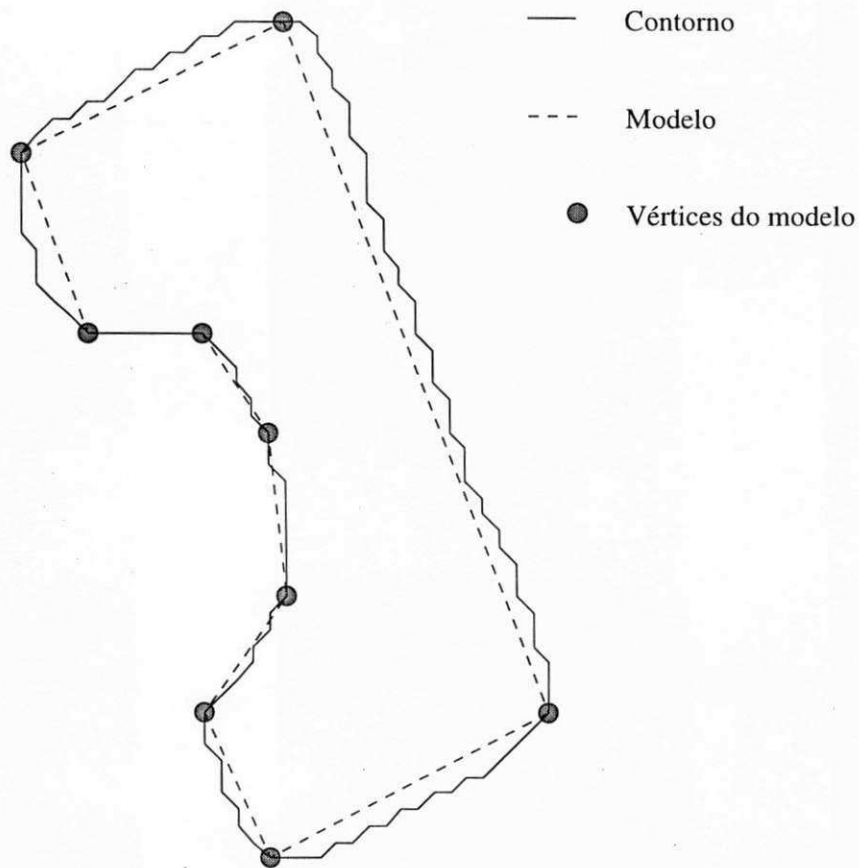


Figura 3.14: Modelo obtido pelo algoritmo Cortez para o objeto Arco.

para uso em sistemas de visão. Estes métodos, porém, podem servir como solução inicial para métodos iterativos.

### 3.7 Análise dos resultados

Um estudo comparativo de algoritmos de modelagem poligonal foi realizado por Junior e Linhares [84]. Neste trabalho, são examinados aspectos de operação (seqüencial ou iterativa) e de implementação de três algoritmos, não sendo fornecidos, entretanto, dados práticos que permitam uma avaliação de desempenho dos mesmos.

A presente avaliação é baseada nos dados experimentais apresentados na seção



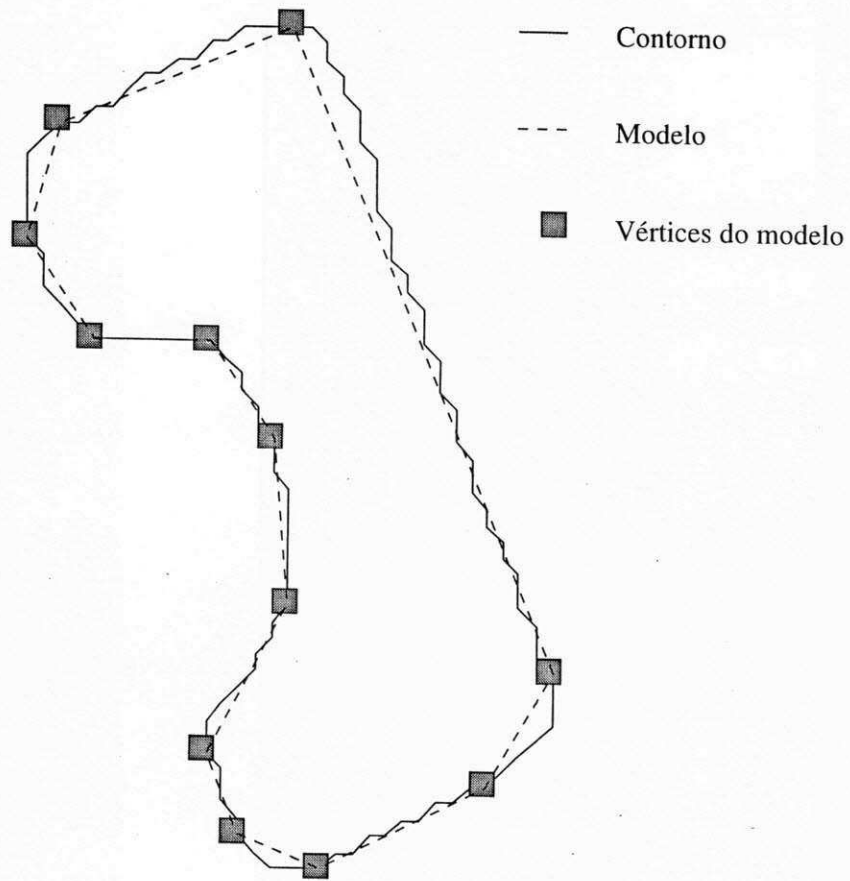


Figura 3.15: Modelo obtido pelo algoritmo Split-Merge para o objeto Arco.

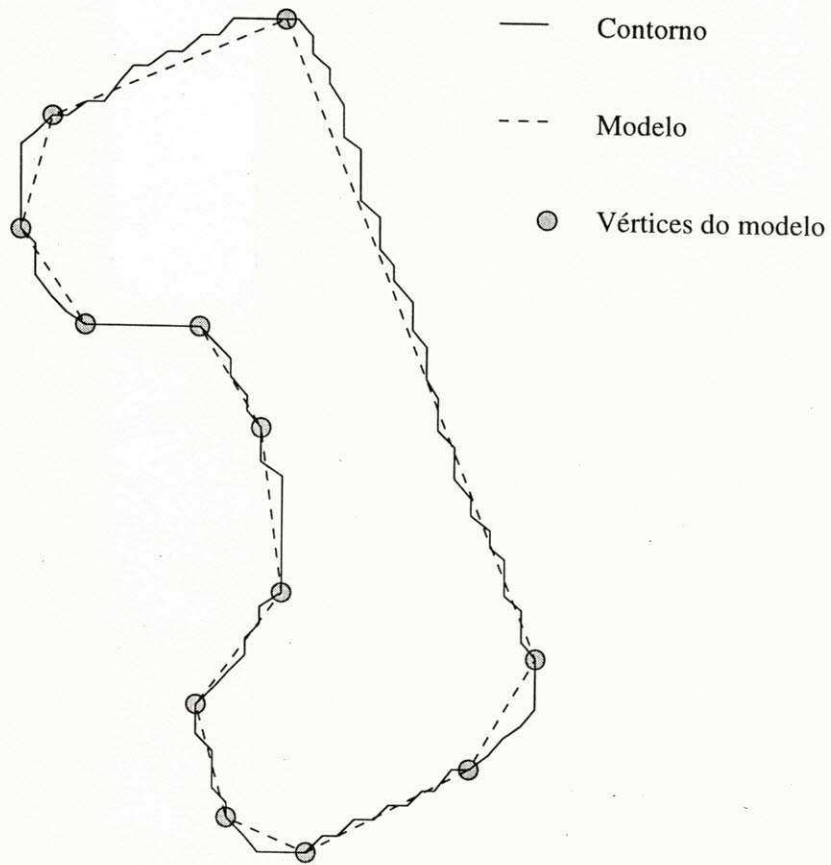


Figura 3.16: Modelo obtido pelo algoritmo Ramer para o objeto Arco.

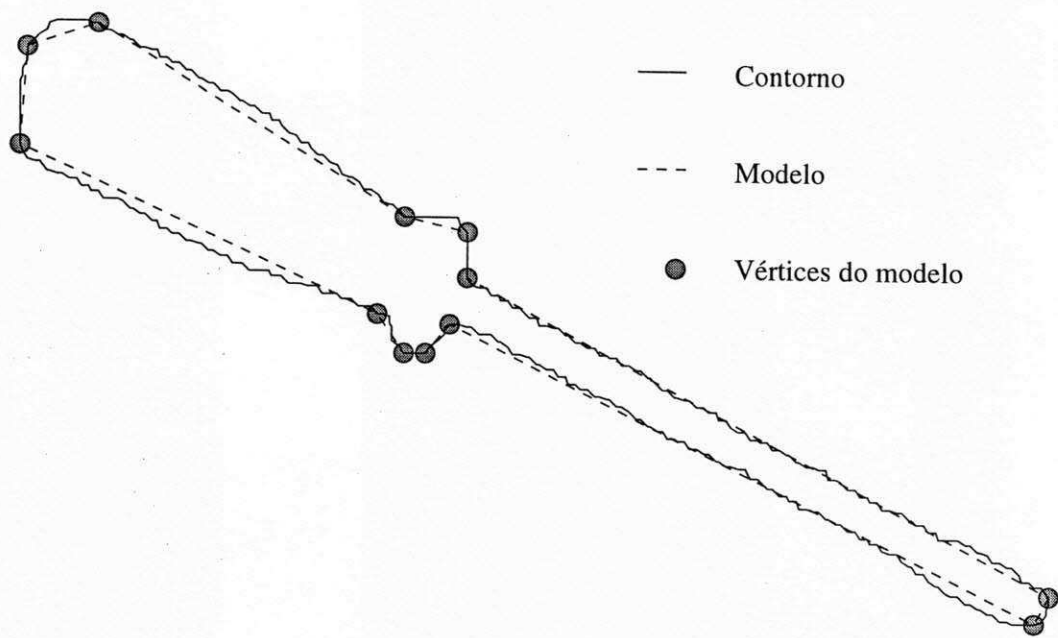


Figura 3.17: Modelo obtido pelo algoritmo Cortez para o objeto Fenda.

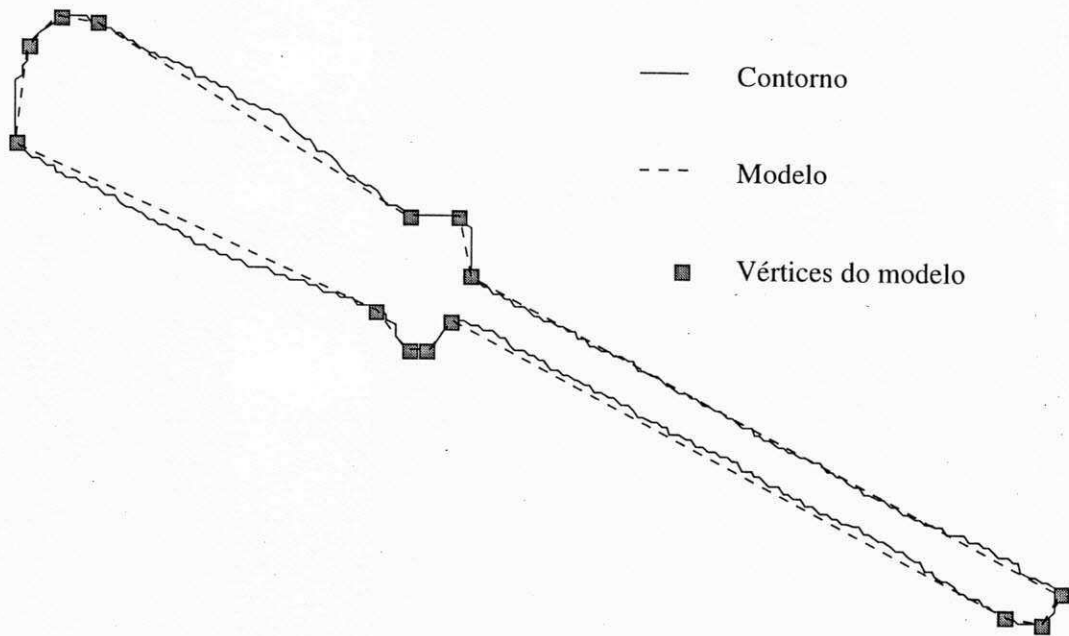


Figura 3.18: Modelo obtido pelo algoritmo Split-Merge para o objeto Fenda.

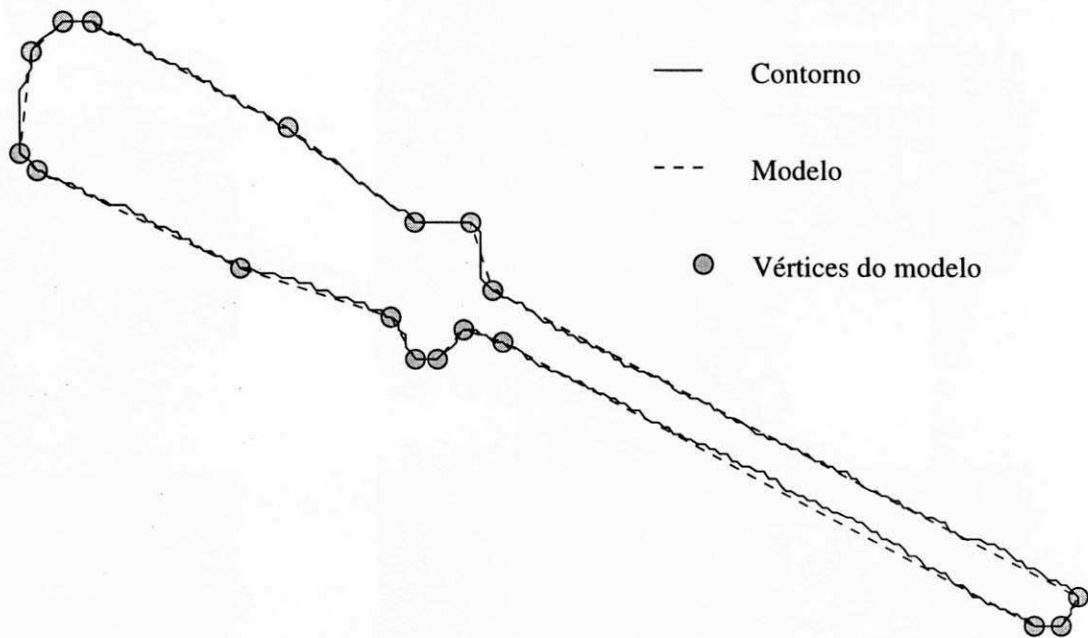


Figura 3.19: Modelo obtido pelo algoritmo Ramer para o objeto Fenda.

3.6. Como parâmetros de comparação foram utilizados o tempo de processamento, a variação do tempo de processamento com o aumento do número de pontos da lista, o número de segmentos poligonais gerados, a taxa de redução dos dados e o erro médio quadrático da aproximação. Os dois primeiros parâmetros se relacionam ao tempo de processamento e os outros à forma geométrica dos modelos produzidos.

Um parâmetro que também deve ser levado em consideração é a posição dos vértices no modelo, analisada a partir do conhecimento a priori do objeto modelado. Este é um parâmetro subjetivo mas que pode influir na avaliação dos algoritmos. Isto se justifica pelo fato de que um método pode produzir um modelo com um pequeno número de lados e com um erro dentro da faixa desejada, para o qual porém a posição dos vértices não coincide com a posição dos cantos do contorno do objeto modelado. A maneira de avaliar este parâmetro é comparando visualmente o contorno do objeto e o seu respectivo modelo.

### 3.7.1 Quanto ao tempo de processamento

De acordo com as Tabelas 3.3 e 3.4, os algoritmos seqüenciais juntamente com o algoritmo paralelo de Cheng-Hsu são mais rápidos, entre 10 a 15 vezes, do que os algoritmos iterativos. Portanto, no que diz respeito ao tempo de processamento, existe uma grande vantagem dos algoritmos seqüenciais e paralelos sobre os algoritmos iterativos.

Outro parâmetro determinante, em algoritmos usados para reconhecimento de formas, é o grau de sensibilidade do algoritmo quanto à quantidade de pontos de contorno. Em geral, este parâmetro é uma função da complexidade das operações realizadas pelo algoritmo. A razão  $t_2/t_1$  (última coluna da Tabela 3.4) expressa a complexidade de cada algoritmo, indicando que um bom algoritmo é aquele cuja razão  $t_2/t_1$  é a mais próxima de um [73]. Conforme este parâmetro, dos algoritmos implementados, os menos sensíveis ao crescimento do número de pontos de borda são os algoritmos Cortez e o método Cheng-Hsu original. Para um crescimento de mais de 200% do número de pontos de contorno, o tempo de processamento destes algoritmos crescem, em média, apenas 25% e 30%, respectivamente. Já para os métodos iterativos este tempo cresce, em média, mais de 220%.

### 3.7.2 Quanto ao número de lados dos modelos

Uma característica muito importante na escolha de um algoritmo para modelagem poligonal é a complexidade dos modelos produzidos, avaliada pelo número de lados do modelo. Quando o número de lados aumenta, o modelo pode se tornar inadequado, devido à baixa taxa de compactação dos dados. Por outro lado, um número de lados muito pequeno pode aumentar o erro ao ponto de tornar o modelo inútil para qualquer aplicação. Portanto, existe um compromisso entre o número de lados, o erro associado ao modelo e as características da aplicação específica.

Para aplicações de reconhecimento de formas, deseja-se produzir modelos com o menor número de lados possível, sem aumentar contudo, o erro de modelagem. Deste modo, é priorizada a formação de lados de grande comprimento para facilitar o processo

de reconhecimento. Um exemplo clássico deste procedimento é o sistema de visão desenvolvido por Ayache [4, 13].

Com base no que foi exposto antes e nas tabelas 3.3 e 3.4, torna-se evidente que os algoritmos Cortez e Split-Merge satisfazem às exigências das aplicações em reconhecimento de formas. Estes métodos produzem modelos com o menor número de lados, em comparação aos outros métodos, com um erro máximo menor do que 2.0 pixels.

### 3.7.3 Quanto à posição dos vértices nos modelos

A observação visual constitui um aspecto importante no processo de reconhecimento de formas [9] e como consequência para o processo de modelagem. Isto significa que as observações visuais dos modelos produzidos pelos algoritmos, associadas ao conhecimento das formas dos objetos a serem modelados (representado por seu contorno), podem ser utilizadas como parâmetro de avaliação dos algoritmos.

Ao observar as figuras 3.14 a 3.19, pode-se dizer que os algoritmos Cortez e Split-Merge escolhem melhor os vértices dos modelos, em relação aos cantos do contorno dos objetos, do que os demais métodos.

## 3.8 Conclusões

Pela análise dos resultados experimentais realizada na seção anterior, pode-se concluir que o método Cortez satisfaz todas as principais exigências para aplicações em reconhecimento de formas. Isto porque produz modelos compactos, com pequeno erro, num tempo de processamento muito menor que o método Split-Merge, considerado clássico para a aplicação mencionada. Além disso, apresenta uma menor ordem de complexidade, medida pela razão  $t_2/t_1$  e posiciona adequadamente os vértices nos modelos, conforme observações visuais.

## Capítulo 4

# Reconhecimento de formas 2D usando modelos poligonais

O processo de reconhecimento é o mais importante e complexo processo num sistema de visão artificial, embora seu desempenho dependa em grande parte do processo de modelagem. Reconhecimento consiste na análise e interpretação dos dados da cena para que sejam casados (ou não) com os dados já existentes, obtidos durante a etapa de treinamento (seção 2.4). O casamento é realizado entre os atributos extraídos da cena e os atributos dos objetos, previamente armazenados. Para sistemas baseados em modelos, os atributos são extraídos dos modelos da cena e dos modelos dos objetos, gerados durante a etapa de treinamento.

As posições espaciais dos modelos no momento do treinamento são diferentes entre si e diferentes do modelo da cena na etapa de reconhecimento. Portanto, as coordenadas dos modelos devem sofrer transformações geométricas para que possam ser efetuadas as operações necessárias ao reconhecimento e localização dos mesmos na cena.

Na próxima seção é apresentada a filosofia do método e nas seções 4.2 e 4.3 são descritos os processos de geração e avaliação das hipóteses, respectivamente.

## 4.1 Metodologia proposta

O método usado para o reconhecimento de formas é baseado na aproximação poligonal dos contornos. Tanto os contornos dos objetos como os contornos das cenas são descritos através desta aproximação, bem como são extraídas dela suas características geométricas (capítulo 3 subseção 3.5.3).

O método de reconhecimento se inicia pela etapa conhecida como geração de hipóteses, realizada ainda durante a fase de segmentação e modelagem da cena. As hipóteses são formadas, independentemente umas das outras, pelo casamento entre os atributos da cena e os atributos dos modelos. O método desenvolvido utiliza como atributo o comprimento de apenas um segmento. Para o processo de avaliação de hipóteses são empregadas cinco técnicas diferentes, as três primeiras compondo o primeiro método e as duas restantes compondo o segundo método.

As hipóteses geradas para cada modelo são agrupadas e ordenadas pelas suas características de similaridade. Utiliza-se para tanto uma função composta pelas variações ponderadas de dois tipos de erros. O primeiro é o erro entre os comprimentos dos segmentos casados da cena e do modelo. Este erro é medido pelo valor absoluto da diferença entre estes comprimentos, em percentagem do comprimento do segmento do modelo. O segundo é o erro de orientação, obtido pelo valor absoluto da diferença de orientação entre os segmentos casados da cena e do modelo. Estes erros são comparados à valores máximos permitidos, que dependem do valor de limiar  $L_d$  adotado para a aproximação poligonal (capítulo 3 subseção 3.5.3) e do limiar  $\epsilon_l$  para o erro entre os comprimentos dos segmentos da cena e do modelo.

Este processo de agrupamento permite diminuir o número de hipóteses a serem analisadas na etapa de reconhecimento. Isto é feito pela eliminação das hipóteses isoladas, ou seja, aquelas que não possuem similaridade com nenhuma outra. Além disso, ao reunir as hipóteses mais prováveis, ou seja, aquelas que melhor satisfazem o critério de similaridade, evita-se testar hipóteses desnecessariamente, reduzindo portanto o tempo de análise.

A análise das hipóteses, agrupadas segundo a medida de similaridade entre elas,



é feita para cada modelo e se encerra para um dado modelo se este for reconhecido, admitindo-se no máximo 50% de oclusão, ou se todas as hipóteses forem testadas e o respectivo modelo for rejeitado. Este percentual de oclusão (50%) é adotado em muitos sistemas, dentre eles o sistema desenvolvido por Ayache que determinou experimentalmente este valor de percentual [4]. A oclusão é medida pela relação entre o perímetro parcial  $P_a$ , resultante da soma dos comprimentos dos lados casados da cena e o perímetro efetivo  $P_e$ , dado pela soma dos comprimentos dos lados do modelo efetivamente usados no processo de geração das hipóteses, ou seja, maiores que o valor mínimo  $L_{min_k}$ . O percentual máximo permitido para a oclusão é previsto, do ponto de vista teórico, no trabalho realizado por Grimson e Huttenlocher [85]. Este resultado é obtido baseado no número de modelos e de atributos, no comprimento médio dos segmentos dos modelos e nos limiares constantes permitidos para os comprimentos e para os ângulos. Entretanto, uma vez que os limiares adotados neste trabalho de tese são variáveis, tal percentual não pode ser calculado pela metodologia desenvolvida por Grimson e Huttenlocher.

A estratégia adotada para a escolha dos diversos limiares é a de fixar apenas o valor do limiar  $\epsilon_l$ , responsável pelo casamento entre os segmentos da cena e os segmentos dos modelos. Os outros limiares são variáveis, dependendo do comprimento do segmento envolvido e também do limiar  $L_d$  (capítulo 3 subseção 3.5.3), adotado para definir a precisão do modelo durante o processo de modelagem do contorno. Esta escolha se justifica pelas mudanças que possam ocorrer com um dado modelo devido à variações de posição do objeto correspondente. Para tanto, admite-se que o erro máximo permitido, entre os pontos do contorno e o segmento reto obtido na modelagem, ocorra exatamente no ponto médio deste segmento. Esta é uma solução simplificada para o problema de variações que possam ocorrer no modelo de um dado objeto quando este se encontra em diferentes posições espaciais na cena. A solução ideal seria determinar precisamente o local onde ocorre o erro máximo para todos os segmentos do modelo. A informação adicional obtida desta maneira, seria usada na determinação dos limiares máximos variáveis para os ângulos associados a cada segmento do modelo.

Para efeito de comparação de resultados, foram selecionados dois métodos de avaliar as hipóteses, ambos verificados para cada modelo casado. O primeiro é aplicado em cada grupo de hipóteses, iniciando pelo par de hipóteses mais provável, até que o reconhecimento ocorra. O par de hipóteses mais provável é determinado por um dado critério de similaridade, utilizado no processo de agrupamento de hipóteses. Os parâmetros empregados como referência para a transformação geométrica, que mapeia sobre a cena os segmentos casados dos modelos, são aqueles definidos pela hipótese que resultou em um menor erro relativo entre os comprimentos dos segmentos casados. São incorporadas ao processo de avaliação, uma a uma, todas as hipóteses de um mesmo grupo que satisfazem a uma função discriminante. Após a análise de um grupo de hipóteses, é feito o teste para verificar se é satisfeito ou não o valor de limiar permitido para a oclusão. Com este método de avaliação de hipóteses, podem ser usadas ainda duas estratégias diferentes para a determinação da transformação geométrica: uma casando o ponto médio dos segmentos da hipótese de referência e a outra casando o ponto médio dos pontos médios dos segmentos referentes ao par de hipóteses.

O segundo método de avaliação de hipóteses é baseado numa transformação geométrica média, obtida inicialmente com base no primeiro par de hipóteses do grupo em análise. Neste método de avaliação de hipóteses, são incorporadas ao cálculo da transformação média, uma a uma, todas as hipóteses do grupo que satisfazem a uma função discriminante. Também são considerados os dois casos de casamentos dos pontos médios conforme descrito acima para o primeiro método.

As operações que compõem os métodos serão descritas detalhadamente nas próximas seções, iniciando-se pela geração de hipóteses, primeira tarefa executada na fase de reconhecimento.

## 4.2 Geração das hipóteses

A formação de uma hipótese se caracteriza pelo casamento entre um segmento da cena e um segmento de um determinado modelo. Este casamento ocorre se o valor absoluto da diferença entre os comprimentos dos segmentos casados, normalizado pelo

comprimento do segmento casado do modelo, for menor que um dado limiar. Este processo se inicia ainda durante a etapa de segmentação e modelagem do contorno da cena, descritos no Capítulo 3.

Considere o conjunto de todos os modelos gerados durante a etapa de treinamento  $\mathcal{M} = \{M_1, M_2, \dots, M_m\}$ , onde  $m \in \mathcal{N}^*$  é o número total de modelos ( $\mathcal{N}^*$  é o conjunto dos números naturais estritamente positivos) e  $C$  o modelo da cena a ser analisada. Seja o  $k$ -ésimo modelo composto pelo conjunto dos comprimentos dos segmentos, isto é,  $M_k = \{L_{k_0}, L_{k_1}, \dots, L_{k_n}\}$ , onde  $n \in \mathcal{N}^*$  é o número total de segmentos do  $k$ -ésimo modelo,  $k \leq m$ . O modelo da cena  $C = \{l_0, l_1, \dots, l_{n'}\}$  é constituído pelos comprimentos dos seus segmentos, onde  $n' \in \mathcal{N}^*$  é o número total de segmentos do contorno contido na cena. Neste processo de geração de hipóteses, todos os  $L_{k_I} \in M_k$ , com  $I \leq n$  e  $k = 1, 2, \dots, m$ , são verificados independentemente para o segmento atual  $l_i \in C$ , com  $i \leq n'$ . Uma dada hipótese  $h$  é formada, resultante do casamento entre o  $i$ -ésimo segmento da cena e o segmento  $I$ -ésimo do  $k$ -ésimo modelo se o erro  $el^{(h)}$  relativo ao comprimento dos segmentos da cena e do modelo satisfaz a desigualdade,

$$el^{(h)} = \frac{|l_i - L_I|}{L_I} \leq \epsilon_l, \quad (4.1)$$

onde  $\epsilon_l$  é um limiar fixo predefinido. Também são obtidas as variações de orientação  $\theta^{(h)}$  entre os sistemas de coordenadas da cena e do  $k$ -ésimo modelo, bem como o fator de escala  $fe^{(h)}$  dados por,

$$\begin{cases} \theta^{(h)} = \theta(l_i) - \theta(L_I) & e \\ fe^{(h)} = \frac{l_i}{L_I}, \end{cases} \quad (4.2)$$

onde  $\theta(\cdot)$  define a orientação do segmento em relação ao eixo- $x$  positivo, variando de  $0^\circ$  a  $360^\circ$  no sentido horário, como ilustrado na figura 4.1. Pode-se caracterizar a  $h$ -ésima hipótese para o  $k$ -ésimo modelo pelo vetor  $\vec{V}_k^{(h)} = [i, I, el^{(h)}, \theta^{(h)}, fe^{(h)}]$ . Os dois primeiros termos deste vetor se referem às respectivas ordens dos segmentos no modelo da cena e no  $k$ -ésimo modelo casado. Os outros três são obtidos a partir dos dois primeiros termos pela equação 4.2 e são fundamentais para o processo de controle do número de hipóteses e para a avaliação destas.

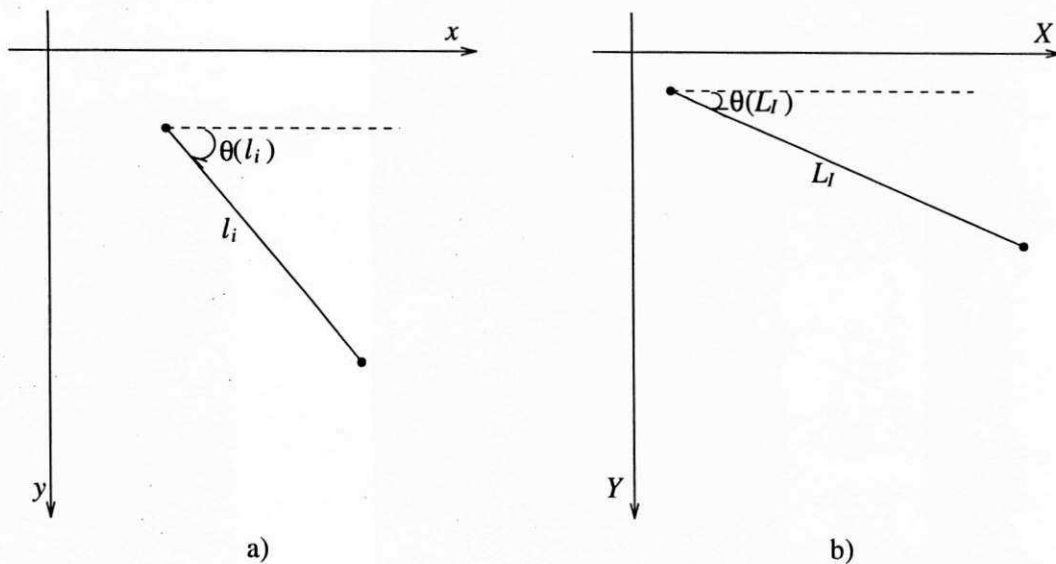


Figura 4.1: Orientação dos segmentos: a) na cena e b) no modelo.

O número de hipóteses, geradas da forma descrita acima, cresce proporcionalmente ao número de lados dos modelos e da cena. Para que não ocorra uma explosão deste número, devem ser adotados alguns critérios preventivos, como descrito a seguir.

#### 4.2.1 Controle do número de hipóteses

O número de hipóteses geradas influi decisivamente no tempo consumido para o casamento do modelo da cena com os modelos armazenados. É conveniente, portanto, reduzir ao máximo o número de hipóteses geradas e conseqüentemente, diminuir o tempo de processamento usado na verificação das hipóteses.

A redução do número de hipóteses é feita em duas etapas ainda durante a sua geração. Na primeira etapa, limita-se o número de segmentos  $L_I$  dos modelos a serem pesquisados, considerando-se somente os segmentos da cena  $l_i$  que são maiores do que um valor de limiar calculado para cada modelo durante a fase de treinamento, ou seja,  $l_i \geq L_{min_k}$ . Este limiar mínimo para o  $k$ -ésimo modelo  $M_k$  é calculado pela equação,

$$L_{min_k} = \overline{L}_k - \frac{\tau_k}{2}, \quad (4.3)$$

onde  $\overline{L}_k$  é o valor médio dos comprimentos dos segmentos que compõem o  $k$ -ésimo modelo e  $\tau_k$  é o seu desvio padrão. Além disso, a pesquisa só se inicia se  $l_i \geq L_{min_T}$ , onde  $L_{min_T}$  é o menor valor de limiar para o conjunto de modelos armazenados, ou seja,  $L_{min_T} = \min [L_{min_k}]$ , com  $k = 1, \dots, m$ .

A segunda etapa reduz o número de hipóteses geradas pelo agrupamento das hipóteses semelhantes para cada modelo, através de uma função de similaridade aplicada a cada par de hipóteses.

A primeira hipótese do primeiro grupo de hipóteses de um dado modelo é formada apenas satisfazendo a desigualdade 4.1. A partir de então, cada hipótese gerada, juntamente com a primeira hipótese do grupo, comporá um par de hipóteses para o qual a similaridade é testada.

Considere o par de hipóteses  $\{h, t\}$  para o  $k$ -ésimo modelo, com  $t \neq h$ , e  $t, h \leq h_k$ , caracterizadas pelos vetores  $\vec{V}_k^{(h)} = [i, I, el^{(h)}, \theta^{(h)}, fe^{(h)}]$  e  $\vec{V}_k^{(t)} = [j, J, el^{(t)}, \theta^{(t)}, fe^{(t)}]$ , onde  $h_k$  é o número total de hipóteses. Estas hipóteses resultam do casamento entre os segmentos  $l_i, l_j \in C$  da cena e os segmentos  $L_I, L_J \in M_k$ , respectivamente. Então, a função de similaridade para o par de hipóteses  $\{h, t\}$  é dada por,

$$s^{(h,t)}(\Delta el, \Delta \theta) = p1 \cdot \frac{|\Delta el|}{2 \cdot el_{max}} + p2 \cdot \frac{|\Delta \theta|}{ea_{max}}, \quad (4.4)$$

onde  $p1, p2$  são pesos, escolhidos de forma que a soma deles seja igual a um. Estes pesos ponderam as diferentes variações que possam vir a ocorrer nos comprimentos e orientações dos lados. Os parâmetros  $\Delta el$ ,  $el_{max}$ ,  $\Delta \theta$  e  $ea_{max}$  são definidos pelo conjunto de equações que se seguem,

$$\begin{cases} \Delta el^{(h,t)} = el^{(h)} - el^{(t)} \\ el_{max} = \epsilon_l \\ \Delta \theta^{(h,t)} = \theta^{(h)} - \theta^{(t)} \\ ea_{max}^{(h,t)} = \max[\arctan(\frac{2 \cdot L_d}{l_i}), \arctan(\frac{2 \cdot L_d}{l_j})], \end{cases} \quad e \quad (4.5)$$

onde  $L_d$  é o erro máximo permitido para a construção do modelo poligonal (capítulo 3 subseção 3.5.3), definido no processo de modelagem.

Diz-se que as hipóteses  $h$  e  $t$  são consideradas semelhantes se o valor da função  $s^{(h,t)}(.,.)$  for menor ou igual à soma dos pesos  $p1$  e  $p2$  (equação 4.4). Caso contrário

é rejeitada a similaridade, conseqüentemente, a hipótese  $t$  não fará parte do mesmo grupo que contém a hipótese  $h$ . Quanto menor for o valor da função de similaridade dada pela equação 4.4 maior será o grau de semelhança entre as hipóteses.

A partir da formação do primeiro grupo, uma hipótese qualquer passa a pertencer a um dado grupo se satisfizer o critério de similaridade definido pela equação 4.4 para pelo menos uma das hipóteses contida no grupo. Caso contrário, dará início a um novo grupo.

Um característica muito importante usada no agrupamento das hipóteses é considerar que o valor máximo  $ea_{max}^{(h,t)}$  usado como limiar na equação 4.4 não é fixo. Tal propriedade reflete a estratégia adotada para os limiares, a qual está baseada no fato de que o processo de modelagem pode produzir variações nos modelos de um mesmo contorno, quando este se encontra em diferentes posições espaciais. Parte-se do princípio de que o erro máximo permitido, entre os pontos do contorno e os seus respectivos segmentos retos produzidos durante o processo de modelagem, ocorre no ponto médio de cada um dos segmentos. Esta suposição possibilita definir o limiar para os ângulos de forma variável, dependendo do comprimento do segmento sob avaliação. Isto evita o uso de informações adicionais sobre a posição onde ocorreu o erro máximo para cada um dos segmentos do modelo. A figura 4.2 ilustra a idéia apresentada e fornece uma interpretação geométrica para a variação máxima permitida para o ângulo  $ea_{max}^{(h,t)}$ , utilizada na equação 4.4.

Além da importante característica mencionada acima, a equação 4.4 expressa a ligação entre os processos de modelagem e reconhecimento, através do limiar para a variação de orientação  $ea_{max}^{(h,t)}$ . Esta ligação evidencia de maneira objetiva a dependência entre aqueles processos.

Após o processo de geração e agrupamento de hipóteses, para todos os segmentos do contorno fechado em análise, inicia-se a etapa de avaliação de hipóteses para grupos de hipóteses com pelo menos um par de hipóteses. Terminado este processo, as hipóteses que não conseguiram se agrupar com nenhuma outra, através da função de similaridade (equação 4.4), deixarão de existir.

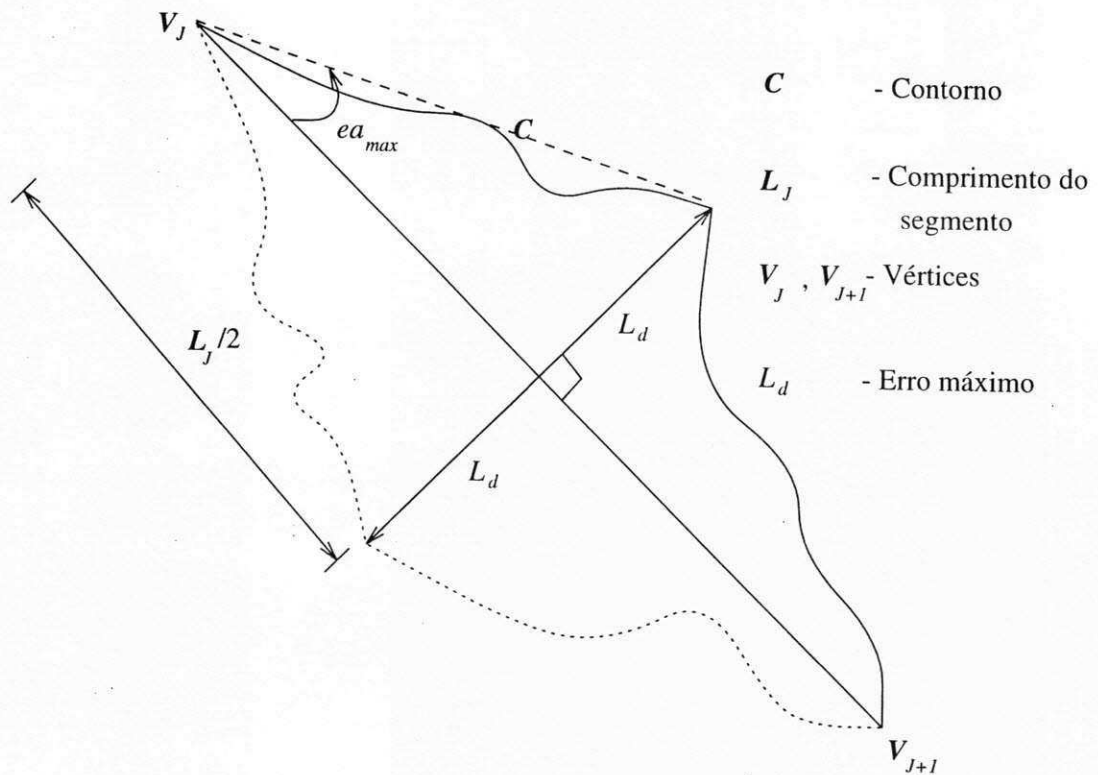


Figura 4.2: Ilustração geométrica dos parâmetros necessários ao cálculo dos valores dos limiares.

### 4.3 Avaliação das hipóteses

O processo de avaliação de hipóteses é responsável pela principal tarefa de um sistema de visão, isto é, a classificação das formas que são apresentadas ao sistema. É neste processo que o desempenho do sistema pode ser avaliado, principalmente pelo tempo utilizado na avaliação das hipóteses e pela classificação correta das formas.

A verificação das hipóteses é a tarefa que exige o maior esforço computacional para a maioria dos sistemas de visão, principalmente aqueles que utilizam processos iterativos. Um sistema muito citado na literatura que adota uma técnica iterativa foi desenvolvido por Ayache [4]. Este sistema pode consumir até dezenas de segundos para reconhecer um modelo.

No presente trabalho, as hipóteses são avaliadas pelo grau de similaridade entre elas, obtido durante o processo de agrupamento. Esta avaliação é feita de maneira independente para cada modelo, por uma função discriminante que depende da posição relativa entre os segmentos casados da cena e do modelo.

O processo de avaliação se inicia para o  $k$ -ésimo modelo  $M_k$ , com pelo menos um grupo de hipóteses semelhantes e na ordem de treinamento dos modelos, o qual é realizado de forma aleatória. Cada grupo a ser avaliado possui no mínimo um par de hipóteses. A primeira tarefa desta etapa é determinar os parâmetros da transformação geométrica  $T^{(*)}$  a ser empregada. Esta transformação é responsável pelo posicionamento do modelo na cena e pode ser representada pelo vetor  $\vec{T}^{(*)} = [\theta^{(*)}, fe^{(*)}, t_x^{(*)}, t_y^{(*)}]$ . Os dois primeiros parâmetros desta transformação representam o ângulo de rotação  $\theta^{(*)}$  e o fator de escala  $fe^{(*)}$  e dependem da orientação e do comprimento dos segmentos casados, respectivamente. Os dois últimos parâmetros  $t_x^{(*)}$  e  $t_y^{(*)}$  representam a translação nas direções  $x$  e  $y$ , respectivamente. Estes dependem dos dois primeiros e dos pontos escolhidos para realizar a translação. Em geral, usam-se os pontos médios dos segmentos casados para distribuir de forma equitativa o erro proveniente da diferença de comprimento entre estes segmentos. Nesta tese, a escolha destes pontos definirá dois métodos diferentes para avaliar as hipóteses.

Considere a avaliação da hipótese  $h$  gerada para o modelo  $M_k$ , caracterizada pelo



vetor  $\vec{V}_k^h = [i, I, el^{(h)}, fe^{(h)}]$ . A transformação geométrica  $\vec{T}^{(*)}$  será aplicada sobre o ponto médio  $P_m(X_I, Y_I)$  do modelo para fazer coincidir com os pontos  $P_m(x_i, y_i)$  da cena. A equação que caracteriza esta transformação [13, 33], é dada por

$$\begin{bmatrix} x_i^{(*)} \\ y_i^{(*)} \end{bmatrix} = \begin{bmatrix} t_x^{(*)} \\ t_y^{(*)} \end{bmatrix} + fe^{(*)} \cdot R(\theta^{(*)}) \cdot \begin{bmatrix} X_I \\ Y_I \end{bmatrix}, \quad (4.6)$$

onde

$$R(\theta^{(*)}) = \begin{bmatrix} \cos(\theta^{(*)}) & -\sin(\theta^{(*)}) \\ \sin(\theta^{(*)}) & \cos(\theta^{(*)}) \end{bmatrix} \quad (4.7)$$

correspondem respectivamente à rotação e os parâmetros  $t_x^{(*)}$  e  $t_y^{(*)}$  à translação nas direções  $x$  e  $y$ , respectivamente. O ponto médio  $P_m(x_i^{(*)}, y_i^{(*)})$  obtido é a projeção do ponto médio  $P_m(X_I, Y_I)$  do modelo sobre a cena.

Entretanto, os parâmetros de translação  $t_x^{(*)}, t_y^{(*)}$  ainda não são conhecidos, necessitando-se, portanto, que sejam determinados para que  $T^{(*)}$  seja definida completamente.

A determinação dos parâmetros de translação  $t_x^{(*)}$  e  $t_y^{(*)}$  é feita a partir da equação 4.6, conhecendo-se os dois primeiros parâmetros  $\theta^{(*)}$  e  $fe^{(*)}$  da transformação  $\vec{T}^{(*)}$  e os pontos médios  $P_m(X_I, Y_I)$  do modelo e  $P_m(x_i, y_i)$  da cena escolhidos para se superpor um ao outro. Então, os valores de  $t_x^{(*)}$  e  $t_y^{(*)}$  são calculados pela equação,

$$\begin{cases} t_x^{(*)} = x_i - fe^{(*)} \cdot [X_I \cdot \cos(\theta^{(*)}) - Y_I \cdot \sin(\theta^{(*)})] \\ t_y^{(*)} = y_i - fe^{(*)} \cdot [X_I \cdot \sin(\theta^{(*)}) + Y_I \cdot \cos(\theta^{(*)})]. \end{cases} \quad (4.8)$$

Estes parâmetros garantem a superposição dos pontos médios dos segmentos  $P_m(x_i, y_i)$  da cena e  $P_m(X_I, Y_I)$  do modelo e completam a definição da referida transformação  $T^{(*)}$ .

É importante resaltar que os parâmetros  $t_x^{(*)}$  e  $t_y^{(*)}$  encontrados através da equação 4.8 estão intrinsicamente ligados aos pontos da cena e do modelo que são superpostos um ao outro. Por outro lado, os parâmetros  $fe^{(*)}$  e  $\theta^{(*)}$ , obtidos durante a geração de hipóteses, estão relacionados apenas com os comprimentos dos segmentos casados da cena  $l_i$  e do modelo  $L_I$ .

A transformação  $T^{(*)}$  é aplicada a cada um dos segmentos do modelo contidos nos vetores que caracterizam as hipóteses em análise. Por exemplo, considere a hipótese  $h$

deste grupo, caracterizada pelo vetor  $\vec{V}_k^{(h)} = [i, I, el^{(h)}, \theta^{(h)}, fe^{(h)}]$ . Usando os parâmetros da transformação  $T^{(*)}$  e aplicando a equação 4.6 ao ponto médio  $P_m(X_I, Y_I)$  do  $I$ -ésimo segmento do modelo, obtêm-se a projeção  $P_m(X_{I_c}, Y_{I_c})$  deste ponto na cena, cujas coordenadas são dadas por,

$$\begin{cases} X_{I_c} = t_x^{(*)} + fe^{(*)} \cdot [X_I \cdot \cos(\theta^{(*)}) - Y_I \cdot \sin(\theta^{(*)})] \\ Y_{I_c} = t_y^{(*)} + fe^{(*)} \cdot [X_I \cdot \sin(\theta^{(*)}) + Y_I \cdot \cos(\theta^{(*)})] \end{cases} \quad (4.9)$$

Em seguida, são determinadas as variações de orientação  $\Delta\theta$  entre os segmentos casados do modelo projetado na cena e o da cena e de distância  $\Delta Dpm$  entre  $P_m(X_{I_c}, Y_{I_c})$  e o ponto médio do respectivo segmento casado na cena  $P_m(x_i, y_i)$ . Os valores destas variações são calculadas pela equação,

$$\begin{cases} \Delta\theta = \theta(l_i) - \theta(L_{I_c}) \\ \Delta Dpm = \{[x_i - X_{I_c}]^2 + [y_i - Y_{I_c}]^2\}^{\frac{1}{2}}, \end{cases} \quad (4.10)$$

onde  $\theta(L_{I_c})$  é o ângulo do  $I$ -ésimo segmento do modelo projetado na cena.

Para decidir se a hipótese em verificação é verdadeira ou não, os valores encontrados para  $\Delta\theta$  e  $\Delta Dpm$ , pela equação 4.10, são submetidos à função discriminante ponderada, dada por,

$$f(\Delta Dpm, \Delta\theta) = p1 \cdot \frac{\Delta Dpm}{ed_{max}} + p2 \cdot \frac{|\Delta\theta|}{ea_{max}}, \quad (4.11)$$

onde  $ed_{max} = 2 \cdot L_d$  é o valor máximo permitido para a distância entre os dois pontos médios e  $ea_{max}$  é o mesmo da equação 4.4. Para que uma hipótese seja considerada como verdadeira, o valor desta função  $f(.,.)$  deve ser menor ou igual a um. Quanto menor for o valor desta função, melhor é o casamento entre os segmentos pertencentes à hipótese. Caso a hipótese seja verdadeira, o comprimento do segmento da cena contido na hipótese é adicionado ao perímetro parcial do modelo para ser comparado ao perímetro efetivo do modelo em análise. Este mesmo procedimento é adotado para cada nova hipótese considerada verdadeira, pertencente ao grupo de hipóteses em verificação.

Ao final da análise de cada grupo de hipóteses, se aceito pelo menos um par de

hipóteses como verdadeiro, é associado ao mesmo um valor dado pela função discriminante de grupo  $f_g(\Delta Dpm, \Delta\theta)$ ,

$$f_g(\Delta Dpm, \Delta\theta) = \frac{p1}{h_v} \cdot \sum_{h_v} \frac{\Delta Dpm}{ed_{max}} + \frac{p2}{h_v} \cdot \sum_{h_v} \frac{|\Delta\theta|}{ea_{max}}. \quad (4.12)$$

O parâmetro  $h_v$  representa o número de hipóteses verdadeiras do referido grupo. Esta função é usada para auxiliar na decisão de casos de contornos com mais de um modelo presente (reconhecidos) e onde as hipóteses, geradas para os diferentes modelos, sejam conflitantes. Um exemplo seria o caso em que um mesmo segmento da cena se encontre presente nas hipóteses avaliadas como verdadeiras para modelos distintos.

Como próxima etapa, para os casos em que o valor da função, dada pela equação 4.11, é menor ou igual a um, é feito o teste para saber se o critério de oclusão é satisfeito, ou seja, se a soma dos segmentos casados é maior ou igual a 50% do perímetro efetivo do modelo em verificação. Caso não satisfeito este critério, a próxima hipótese do mesmo grupo satisfazendo a equação 4.11 é incorporada ao processo de verificação. Este procedimento é repetido até que todas as hipóteses do grupo sejam verificadas ou que seja satisfeito o critério de oclusão. No caso de mais de um modelo ter sido identificado como presente na cena, faz-se também a verificação para determinar se existem hipóteses conflitantes entre eles. Se existirem hipóteses em conflito, será reconhecido como presente na cena o modelo que apresentar menor valor de  $f_g$ , se o valor absoluto da diferença entre os percentuais de oclusão for menor do que 10%. Caso contrário, é escolhido aquele modelo que obteve menor percentual de oclusão.

O problema de avaliação de hipóteses será solucionado por dois métodos distintos, conforme a escolha dos parâmetros de orientação  $\theta^{(*)}$  e de escala  $fe^{(*)}$  usados pela transformação  $T^{(*)}$ . O primeiro destes métodos, emprega valores de  $\theta^{(*)}$  e  $fe^{(*)}$  iguais aqueles pertencentes ao vetor que caracteriza a hipótese mais provável do grupo (menor valor de  $el^{(\cdot)}$ ), ou seja, iguais a  $\theta^{(h)}$  e  $fe^{(h)}$ , com  $h$  sendo a hipótese mais provável.

No segundo método os parâmetros de  $T^{(*)}$  são calculados com base nos parâmetros  $\theta^{(*)}$  e  $fe^{(*)}$  das hipóteses em análise. Cada um dos dois métodos será também classificado conforme a determinação dos parâmetros  $t_x^{(*)}$  e  $t_y^{(*)}$ , caracterizando-se como técnicas distintas de avaliação de hipóteses. A primeira, efetiva o casamento entre os

pontos médios dos segmentos que compõem a hipótese mais provável (referência) e a segunda realiza pelo casamento entre os pontos cujas coordenadas são calculadas como a média aritmética das coordenadas dos pontos médios dos segmentos que formam as hipóteses em análises que satisfazem a equação 4.11.

As referências ao primeiro e segundo métodos será feita pela transformação utilizada  $T^{(h)}$  e  $\bar{T}$ , respectivamente. Os dois casos de determinação dos parâmetros de translação  $t_x^{(*)}$  e  $t_y^{(*)}$ , descritos acima, serão referenciados pelo acréscimo dos subscritos  $h$  para o primeiro caso e  $m$  para o segundo caso. Desta maneira, o primeiro método é composto por duas técnicas, cujas transformações são dadas por  $T_h^{(h)}$  e  $T_m^{(h)}$ , respectivamente. Da mesma maneira, o segundo método será formado por duas técnicas dadas por  $\bar{T}_h$  e  $\bar{T}_m$ .

Nas próximas subseções são descritos os dois métodos de avaliação de hipóteses.

### 4.3.1 Primeiro método de avaliação de hipóteses

O primeiro método de avaliação de hipóteses utiliza a transformação  $T^{(*)}$ , adotando-se como referência para os parâmetros  $\theta^{(*)}$  e  $fe^{(*)}$  os mesmos da hipótese  $h$  (menor erro  $el$ ), ou seja, estes parâmetros são iguais a  $\theta^{(h)}$  e  $fe^{(h)}$ , respectivamente.

Considere a análise, por este método, de um primeiro par de hipóteses  $\{h, t\}$  geradas para o modelo  $M_k$ , agrupadas conforme o procedimento apresentado na subseção 4.3.1. Estas hipóteses são caracterizadas por seus respectivos vetores  $\vec{V}_k^{(h)} = [i, I, el^{(h)}, \theta^{(h)}, fe^{(h)}]$  e  $\vec{V}_k^{(t)} = [j, J, el^{(t)}, \theta^{(t)}, fe^{(t)}]$ . Como mencionado acima, a determinação da transformação  $T^{(*)}$  e portanto a avaliação das hipóteses será feita por duas técnicas distintas.

A primeira técnica emprega uma transformação  $T^{(*)} = T_h^{(h)}$  cujos parâmetros  $\theta^{(*)}$  e  $fe^{(*)}$  são extraídos diretamente do vetor  $\vec{V}_k^{(h)}$ . A partir destes, pela aplicação da equação 4.8, são determinados os parâmetros  $t_x^{(*)}$  e  $t_y^{(*)}$ , utilizando o casamento dos pontos médios do  $i$ -ésimo segmento da cena e do  $I$ -ésimo segmento do modelo.

Submete-se, neste caso, o  $J$ -ésimo segmento casado do modelo, contido no vetor  $\vec{V}_k^{(t)} = [j, J, el^{(t)}, \theta^{(t)}, fe^{(t)}]$ , à transformação  $T_h^{(h)}$ . Deste modo, o segmento é projetado sobre o sistema de coordenadas da cena.

A segunda técnica de avaliação de hipóteses emprega uma transformação  $T^{(*)} = T_m^{(h)}$  cujos parâmetros  $\theta^{(*)}$  e  $f_e^{(*)}$  também são extraídos diretamente do vetor  $\vec{V}_k^{(h)}$ . Entretanto, os parâmetros referentes à translação,  $t_x^{(*)}$  e  $t_y^{(*)}$  são calculados através da equação 4.8, utilizando o casamento dos pontos médios calculados  $P_{m_c}(\bar{x}, \bar{y})$  da cena e  $P_{M_k}(\bar{X}, \bar{Y})$  do modelo. As coordenadas destes pontos são calculadas como a média aritmética das coordenadas dos pontos médios  $P_{m_c}(x_i, y_i)$  e  $P_{m_c}(x_j, y_j)$  dos segmentos da cena e  $P_{M_k}(X_I, Y_I)$  e  $P_{M_k}(X_J, Y_J)$  dos segmentos do modelo, respectivamente.

Para as cenas com contornos formados por mais de um objeto, ou seja, com objetos parcialmente oclusos, ocorre com frequência a fragmentação de segmentos. Este problema é solucionado pelo uso de uma função discriminante cujos valores de limiar de distância entre os pontos médios  $ed_{max}$  é acrescido de 50% em relação ao valor usado na equação 4.11, ou seja,  $ed_{max} = 3.L_d$ . Esta função é aplicada, após a verificação de um dado grupo de hipóteses, a todos os segmentos da cena e do modelo que não foram casados durante a etapa de geração de hipóteses.

O aumento do valor de limiar  $ed_{max}$  em 50% se baseia no percentual permitido de oclusão, que por sua vez se refere ao perímetro do modelo. Como o perímetro é dado pela soma dos lados, admite-se que cada segmento pode se encontrar até 50% ocluso, pois é este o percentual adotado para a oclusão máxima de objetos.

A referência a esta maneira de lidar com o problema de fragmentação será feita como sendo uma terceira técnica de avaliação de hipóteses. É uma técnica que utiliza o casamento entre os pontos médios das hipóteses em verificação para determinar se as hipóteses são ou não verdadeiras, levando em conta o problema de fragmentação de segmentos da cena.

Conhecendo-se todos os parâmetros da transformação  $T^{(*)}$ , esta é aplicada a cada um dos segmentos do modelo especificados pelos vetores que caracterizam cada par de hipóteses. Deste modo, os segmentos casados dos modelos são mapeados para o sistema de coordenadas da cena. Em seguida, são determinadas as variações de orientação  $\Delta\theta$  e de distância  $\Delta D_{pm}$  entre cada um dos pontos médios dos segmentos casados da cena e dos mapeamentos dos pontos médios dos segmentos do modelo. De posse destes valores, aplica-se então a função discriminante (equação 4.11) a todos os pares de hipóteses do

mesmo grupo, formados pela primeira hipótese (de referência) e cada uma das demais hipóteses.

### 4.3.2 Segundo método de avaliação de hipóteses

A idéia de utilizar uma transformação média neste segundo método é para posicionar na cena os segmentos casados do modelo de tal forma que diminua a distância média entre os respectivos pontos médios, sem exceder o erro  $L_d$  permitido por segmento.

A transformação  $T^{(*)}$  empregada é uma transformação média  $\bar{T}$  cujos parâmetros são inicialmente calculados a partir do primeiro par de hipóteses em avaliação. Posteriormente estes parâmetros serão atualizados a cada nova hipótese incorporada ao processo de avaliação, desde que esta hipótese satisfaça a função discriminante dada pela equação 4.11. Para determinar os parâmetros da transformação média  $\bar{T}$ , parte-se do princípio de que os parâmetros relativos ao produto,  $f e^{(h)}.R(\theta^{(h)})$  (equação 4.6) dependem apenas do casamento dos segmentos. Os parâmetros  $t_x^{(h)}$  e  $t_y^{(h)}$  referentes à translação, dependem também do casamento entre os pontos médios, conforme evidenciado pela equação 4.6. Desta forma, o produto  $f e^{(h)}.R(\theta^{(h)})$  pode ser isolado, tornando-se possível caracterizá-lo apenas por um dos produtos  $f e^{(h)}. \cos(\theta^{(h)})$  ou  $f e^{(h)}. \sin(\theta^{(h)})$ . Deseja-se calcular o valor médio, ou valor esperado, de um destes produtos, por exemplo  $E[f e^{(h)}. \cos(\theta^{(h)})]$ . Isto é suficiente para se determinar o valor esperado  $E[\cos(\theta^{(h)})] = \cos(\theta^{*(h)})$ , onde  $\theta^{*(h)}$  é o valor do ângulo que produz esta igualdade.

Para o conjunto  $H$  de hipóteses agrupadas à hipótese  $h$  para o  $k$ -ésimo modelo, tem-se que

$$\begin{cases} E[f e^{(H)}. \cos(\theta^{(H)})] = \frac{\sum_{w=h}^{h_v} f e^{(w)}. \cos(\theta^{(w)})}{h_k} = E[f e^{(H)}]. E[\cos(\theta^{(H)})], \\ \text{com } w \leq h_v, \end{cases} \quad (4.13)$$

onde  $f e^{(w)}$  e  $\theta^{(w)}$  são independentes para qualquer  $w \leq h_v$  e  $h_v$  é o número de hipóteses em verificação. Pode-se encontrar o valor médio  $E[f e^{(H)}] = \bar{f e}^{(H)}$  e, a partir deste,

calcular  $E[\cos(\theta^{(H)})] = \cos(\theta^{*(H)})$ , através da seguinte equação,

$$\begin{cases} \bar{f}e^{(H)} = \frac{\sum_{w=h}^{h_k} f e^{(w)}}{h_k}, \\ E[\cos(\theta^{(H)})] = \cos(\theta^{*(H)}) = \frac{\sum_{w=h}^{h_k} f e^{(w)} \cdot \cos(\theta^{(w)})}{\sum_{w=h}^{h_k} f e^{(w)}}. \end{cases} \quad (4.14)$$

Desta maneira, são determinados os parâmetros  $\theta^*$  e  $\bar{f}e$  da transformação média  $\bar{T}$ . Resta calcular os parâmetros de translação  $t_x^{(*)}$  e  $t_y^{(*)}$  para que esta transformação fique completamente definida. A forma de calcular os parâmetros de translação  $t_x^{(*)}$  e  $t_y^{(*)}$  caracterizam mais duas técnicas de avaliação de hipóteses, a quarta e quinta.

A quarta técnica utiliza uma transformação  $T^{(*)} = \bar{T}_h$ , ou seja, os parâmetros de translação  $t_x^{(*)}$  e  $t_y^{(*)}$  são iguais aos respectivos parâmetros  $t_x^{(h)}$  e  $t_y^{(h)}$  da hipótese de referência<sup>1</sup>.

A determinação dos parâmetros  $t_x^{(*)}$  e  $t_y^{(*)}$  utilizados pela quinta técnica é feita pelo cálculo da média aritmética entre os pontos médios da cena e do modelo contidos no par de hipóteses em avaliação. Portanto, esta técnica é caracterizada pela transformação  $T^{(*)} = \bar{T}_m$ .

O emprego da quinta técnica, adotando o procedimento de iniciar a avaliação de hipóteses pelo cálculo da transformada média a partir do primeiro par de hipóteses e atualizar este cálculo para cada hipótese do mesmo grupo que satisfaça a função discriminante, produz um aumento do número de hipóteses a serem verificadas. Este fenômeno introduz uma maior discrepância nos parâmetros de translação e conseqüentemente compromete o posicionamento relativo entre o modelo e a cena. A figura 4.3 ilustra este fenômeno para uma cena contendo o objeto Arco rotacionado de 60° graus em relação a sua posição de treinamento. O grupo de hipóteses responsável pelo casamento é composto da seguinte forma:

Modelo arco: Hip. [cena, modelo, theta, erro, fe, tx, ty]

Grupo: 0

Hip. 0: [9, 8, 60, 0.070, 0.930, 168, -62]

Hip. 1: [1, 9, 63, 0.071, 0.929, 176, -56]

<sup>1</sup>Aquela com menor valor de  $el$ .



Figura 4.3: Posicionamento do modelo casado na cena: a) usando o grupo completo de hipóteses e b) usando apenas as três primeiras hipóteses.

Hip. 2: [4,1,69,0.138,0.862,189,-37]

Hip. 3: [3,0,73,0.265,0.735,186,-6]

Ao analisar este grupo de hipóteses, verifica-se que a última delas, Hip. 3, possui um valor de parâmetro  $t_y$  distante dos valores associados às outras hipóteses do grupo. Este fato também é confirmado pela figura 4.3, por comparação do posicionamento do modelo do objeto na cena entre as figuras 4.3.a e 4.3.b. Isto evidencia a influência da dispersão entre os parâmetros de translação no posicionamento do modelo na cena.

O problema exposto acima pode ser solucionado por uma filtragem, ou eliminação por limiar, de hipóteses, considerando os parâmetros de translação. Isto pode ser feito logo após o processo de geração de hipóteses ou após o casamento do modelo do objeto na cena. Adotou-se esta última solução por envolver apenas o grupo de hipóteses responsável pelo casamento, reduzindo portanto o número de hipóteses e conseqüentemente o tempo de filtragem.

O fenômeno descrito acima pode ser produzido pelos dois valores dos parâmetros de translação, associados a cada hipótese. Portanto, o processo de filtragem deve ser



Parâmetros	1º Método			2º Método	
	Técnicas			Técnicas	
	1ª	2ª	3ª	4ª	5ª
$\theta$	$\theta^{(h)}$	$\theta^{(h)}$	$\theta^{(h)}$	$\theta^*$	$\theta^*$
$fe$	$fe^{(h)}$	$fe^{(h)}$	$fe^{(h)}$	$\bar{fe}$	$\bar{fe}$
$t_x$	$t_x^{(h)}$	$t_{\bar{x}}$	$t_{\bar{x}}$	$t_x^{(h)}$	$t_{\bar{x}}$
$t_y$	$t_y^{(h)}$	$t_{\bar{y}}$	$t_{\bar{y}}$	$t_y^{(h)}$	$t_{\bar{y}}$

Tabela 4.1: Resumo dos dois métodos e suas técnicas.

realizado sobre ambos os parâmetros,  $t_x$  e  $t_y$ , adotando-se como limiar mínimo o valor absoluto da média de cada um dos parâmetros. Portanto, somente continuam no grupo aquelas hipóteses cujos valores absolutos de  $t_x$  e de  $t_y$  são superiores aos respectivos valores absolutos de suas médias  $\bar{t}_x$  e  $\bar{t}_y$ .

Após submeter o grupo de hipóteses ao processo de filtragem descrito acima, calcula-se a transformada média entre as hipóteses que continuaram no grupo. Baseado nesta transformada é então feita uma reavaliação do grupo de hipóteses para determinar sua função discriminante  $f_g$ , o percentual de oclusão e o posicionamento do modelo sobre a cena.

A tabela 4.1 mostra de forma resumida os parâmetros que compõem a transformação geométrica utilizada por cada uma das cinco técnicas. Nesta tabela, os parâmetros  $t_{\bar{x}}$  e  $t_{\bar{y}}$  se referem aqueles obtidos pelo casamento dos pontos médios, cujas coordenadas são calculadas como a média aritmética das coordenadas dos pontos médios dos segmentos pertencentes às hipóteses em análises.

Ao concluir a avaliação das hipóteses para um determinado modelo, o mesmo deve ser localizado, caso tenha sido reconhecido. A descrição desta tarefa é feita na seção seguinte.

## 4.4 Localização de formas reconhecidas

Uma das tarefas mais importantes de um sistema de visão é a localização dos modelos reconhecidos. Esta é uma tarefa indispensável para sistemas que manipulam partes numa linha de produção.

O sistema proposto executa a tarefa de localizar formas reconhecidas empregando a mesma transformação  $T^{(*)}$  utilizada para o reconhecimento de uma determinada forma numa cena. Esta transformação é aplicada ao centróide, calculado a partir das coordenadas dos vértices de cada modelo durante o processo de modelagem descrito no capítulo 3, subseção 3.5.3. A posição da projeção deste centróide, no sistema de coordenadas da cena, caracterizará a posição do objeto reconhecido na cena.

# Capítulo 5

## Resultados experimentais

### 5.1 Introdução

Neste capítulo, a metodologia de reconhecimento de formas proposta é avaliada através de resultados obtidos experimentalmente. Esta avaliação é baseada em parâmetros objetivos medidos e na análise e classificação dos objetos presentes nas cenas utilizadas, tendo por objetivo demonstrar o desempenho de uma determinada técnica. Além disso, constitui uma importante forma de validação de novas técnicas desenvolvidas.

Os experimentos apresentados nesta tese utilizam três tipos de cenas. O primeiro tipo é composto por cenas reais, obtidas de imagens digitais. O segundo tipo corresponde à cenas geradas com auxílio do editor gráfico "XFIG", utilizando objetos cujos contornos foram obtidos de cenas reais através do algoritmo BTVE, descrito no capítulo 3. Finalmente, o terceiro tipo de cenas, é obtido pela rotação de imagens gradientes em torno do primeiro ponto de borda, sempre no sentido horário. Nestas últimas cenas estão presentes apenas um objeto. A idéia de empregar nos testes cenas geradas artificialmente é para melhor controlar a quantidade de oclusão e com isto ajudar na análise das técnicas desenvolvidas.

Para os resultados mostrados são considerados dois exemplos de objetos isolados e três com cenas onde os objetos aparecem superpostos uns aos outros.

Cada contorno fechado é analisado isoladamente, obedecendo a seqüência de obtenção dos mesmos nas cenas. Portanto, inicia-se pelo contorno que possuir o primeiro ponto de borda, localizado pela varredura linha-a-linha da imagem gradiente. Os demais contornos, que possam estar presentes na imagem gradiente, possuem os primeiros pontos de borda em linhas posteriores à linha do primeiro ponto de borda do contorno anterior.

A avaliação é realizada considerando um conjunto de treinamento de onze objetos, ou seja, são onze modelos poligonais e os seus atributos extraídos, conforme descrito na subseção 3.5.3 (figura 3.12).

Os valores de limiares empregados para o erro relativo  $\epsilon_l$  e a distância máxima permitida  $L_d$  são de 0.30 e 2.0 pixels, respectivamente. Aos pesos  $p_1$  e  $p_2$  e à oclusão máxima permitida foram atribuídos os valores 0.60, 0.40 e 50%, respectivamente.

Os resultados serão apresentados na seqüência: a) as figuras contendo os contornos e os modelos das cenas; b) resultados obtidos para as cinco técnicas, mostrados através de figuras nas quais se encontram superpostos os modelos da cena e dos objetos reconhecidos e c) parâmetros usados para avaliar o desempenho das técnicas, listados em tabelas.

A seção seguinte deste capítulo define os parâmetros empregados na avaliação de desempenho. Na seção 5.3 são mostradas as cenas a serem submetidas às várias técnicas de avaliação de hipóteses, na seção 5.4 são mostrados os resultados produzidos pela primeira, segunda e terceira técnicas. A seção 5.5 mostra os resultados obtidos pela aplicação da quarta e quinta técnicas e finalmente na seção 5.6 são analisados os resultados e tiradas algumas conclusões.

## 5.2 Parâmetros para avaliação de desempenho

O parâmetro mais importante na avaliação de um sistema de visão para aplicações em linhas de produção industriais é o tempo gasto para reconhecer e localizar corretamente os objetos nas cenas. Para estas e outras aplicações em tempo real, este tempo deve

ser o menor possível. Nesta tese, o parâmetro tempo é composto por duas parcelas: a) o tempo empregado nos processos de segmentação, modelagem do contorno e geração de hipóteses chamado de  $t_{smg}$  e b) o tempo empregado para a avaliação das hipóteses (classificação) e localização do modelo na cena,  $t_{cl}$ . Os valores destes tempos são obtidos calculando o tempo médio de dez execuções distintas para cada um dos processos mencionados acima.

Um critério objetivo para avaliar uma determinada técnica de reconhecimento de formas é a precisão do posicionamento relativo entre as formas poligonais da cena e as formas poligonais dos objetos. Este critério pode ser medido por dois parâmetros: as variações média de distância  $\Delta\bar{D}_{pm}$ , em pixels, e de orientação  $\Delta\bar{\theta}$ , em graus, entre os pontos médios dos segmentos casados da cena e do modelo identificado. O valor máximo permitido para o parâmetro  $\Delta\bar{D}_{pm}$  não deve exceder o valor de limiar  $L_d$  empregado no processo de modelagem. Além disso, é usada também uma função que expressa uma combinação ponderada destes parâmetros. Esta é a função discriminante de grupo  $f_g$  (equação 4.11), que será tanto menor quanto mais coincidente for a posição das formas presentes na cena e nos modelos reconhecidos. Nesta tese são utilizados, na avaliação das técnicas de reconhecimento, estes dois parâmetros médios e a função discriminante de grupo  $f_g$ . Outro parâmetro a ser medido para avaliar as técnicas é o percentual de oclusão  $p_{oc}$ , o qual participa no processo de decisão sobre a classificação dos objetos.

Além dos parâmetros descritos acima, é medida também a taxa de redução de hipóteses  $\tau_h$ , obtida pelo emprego do mecanismo de controle do número de hipóteses geradas e agrupadas. Esta taxa é dada pela relação entre o número total de hipóteses geradas sem utilizar aquele mecanismo e o número total de hipóteses a serem analisadas, utilizando o controle do número de hipóteses.

A terceira técnica, dentre as descritas no capítulo 4, possui a propriedade de poder incorporar, ao perímetro parcial, o comprimento de segmentos que ainda não foram casados. Tal propriedade somente pode ser avaliada pelo perímetro parcial máximo  $P_{p(max)}$  obtido, principalmente para objetos que foram rejeitados pelo critério de oclusão mas que possuem pelo menos um par de hipóteses verdadeiras. Para os casos de objetos

reconhecidos esta propriedade pode ser avaliada pelo percentual de oclusão  $p_{oc}$ .

Na seção seguinte são mostradas as cenas que serão analisadas e medidos os seus parâmetros, para possibilitar uma avaliação objetivas das cinco técnicas desenvolvidas.

### 5.3 Cenas a serem analisadas

As cenas 1 e 2, contendo apenas objetos isolados, são do terceiro tipo. As cenas 3, 4 e 5, contém contornos com mais de um objeto presente, sendo a cena 3 do primeiro tipo e as cenas 4 e 5 do segundo tipo.

Os tempos, em milissegundos, gastos nos processos de segmentação, modelagem, extração de atributos e geração de hipóteses  $t_{smg}$  para as cenas 1, 2, 3, 4 e 5 foram 48, 49, 300, 19 e 20 ms, respectivamente.

As figuras mostradas neste capítulo, referentes à classificação, estão ampliadas duas vezes para facilitar a compreensão visual dos processos de modelagem e de casamento entre as formas. As demais figuras mostram os objetos no tamanho real.

As cenas são compostas por objetos isolados (cenas 1 e 2) e por objetos parcialmente oclusos (cenas 3, 4 e 5). Os contornos e os modelos dos objetos que compõem estas cenas são mostrados na figura 5.1.

Considere inicialmente as cenas 1 e 2 do terceiro tipo contendo um objeto cada. A primeira é um objeto retangular denominado Bloco, apropriado para se obter um modelo poligonal. A segunda é um objeto denominado Arco, com regiões de contorno curvas e retas. As figuras 5.2.a e 5.2.b mostram as cenas 1 e 2, respectivamente. Cena 1 é obtida pela rotação de  $300^\circ$ , em torno do primeiro ponto de borda, da imagem gradiente que originou o contorno mostrado na figura 5.2.a. Cena 2 é composta pelo objeto Arco rotacionado de  $60^\circ$ , adotando-se o mesmo procedimento.

Considere a cena 3, do primeiro tipo, contendo três contornos fechados, como ilustrado na figura 5.3. No primeiro contorno os objetos Bloco, Cadeado, Chave, Fenda e Triang2 se encontram superpostos. Os outros dois contornos, contendo objetos isolados, não serão analisados neste trabalho. A análise desta cena será feita apenas para

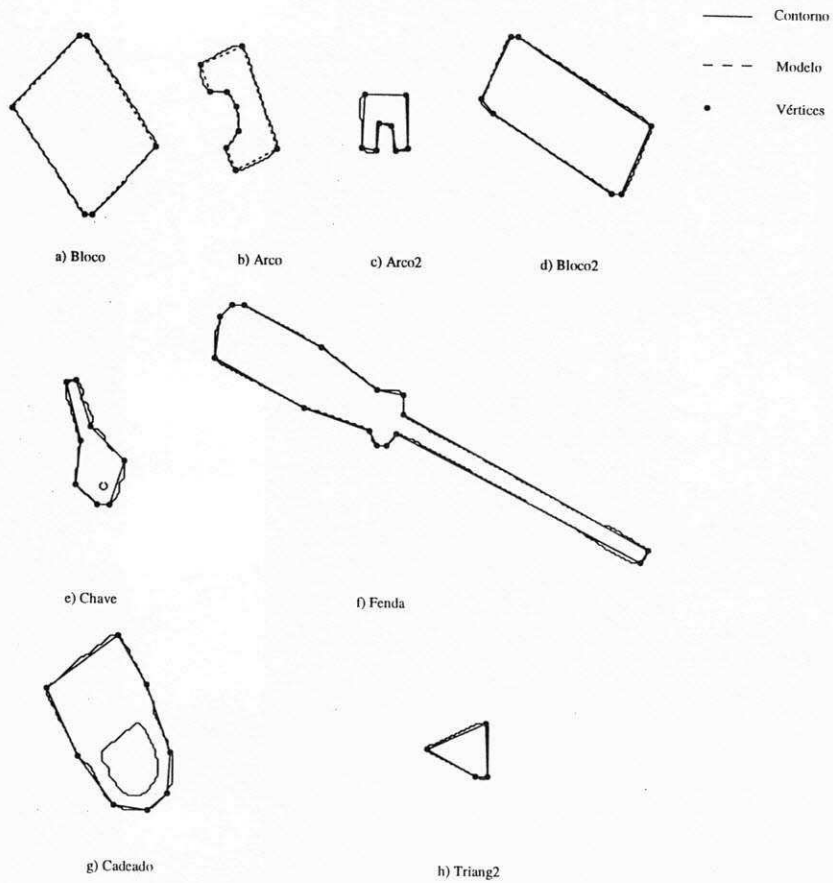


Figura 5.1: Conjunto de objetos que compõem as cinco cenas.

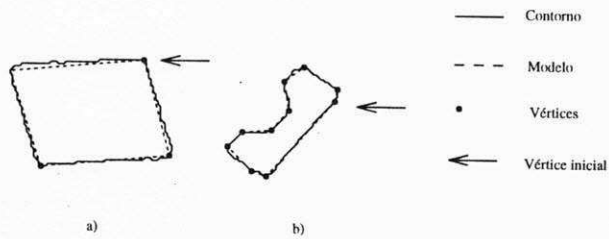


Figura 5.2: Contornos e modelos poligonais das cenas: a) cena 1 e b) cena 2.

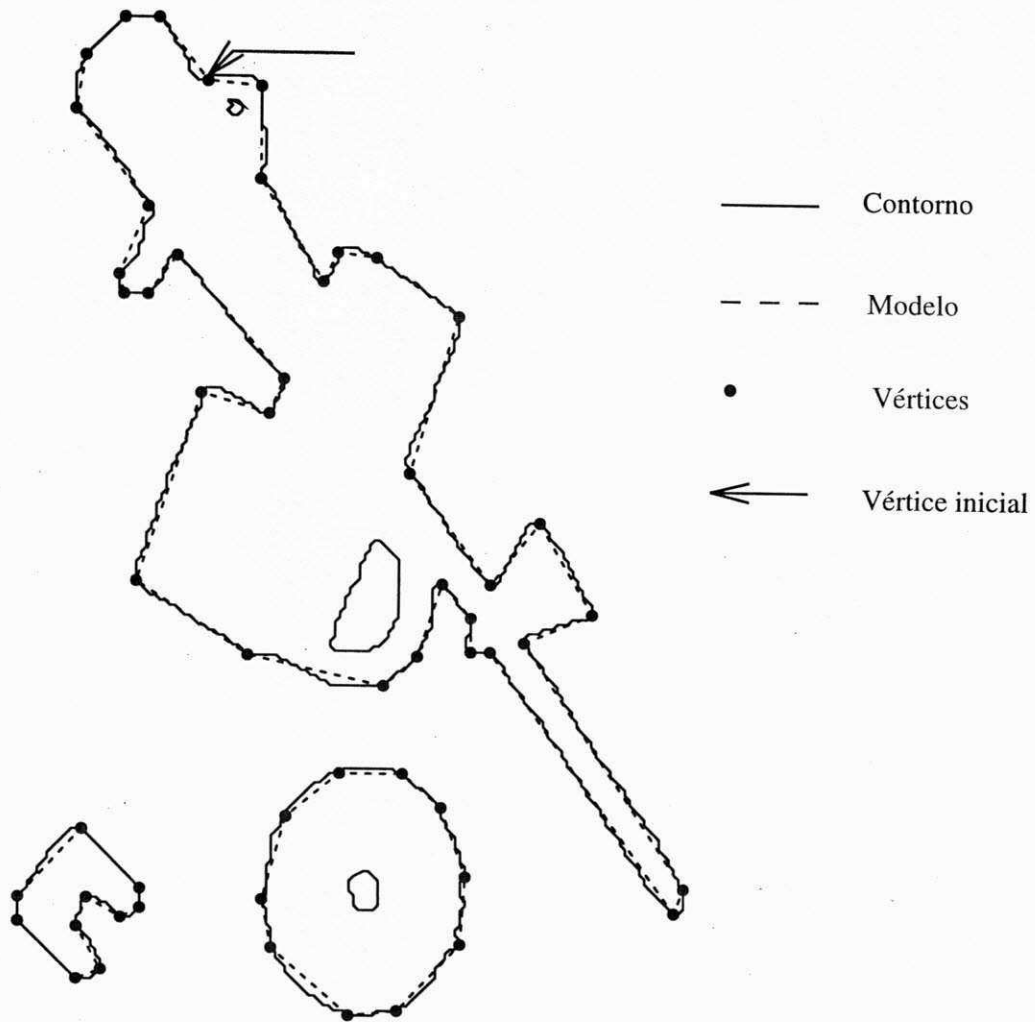


Figura 5.3: Contorno e modelo poligonal da cena 3, contendo cinco objetos presentes.



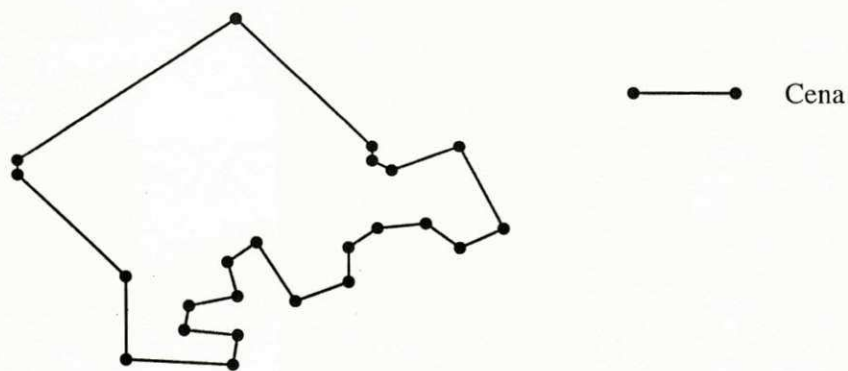


Figura 5.4: Modelo poligonal da cena 4.

o primeiro contorno fechado.

Cena 4 é do segundo tipo, composta pela superposição dos objetos Arco, Arco2 e Bloco. A figura 5.4 mostra o modelo poligonal desta cena.

Finalmente, a cena 5 é do segundo tipo, composta pela superposição dos objetos Arco, Bloco e Fenda. A figura 5.5 mostra o modelo poligonal desta cena.

Na seção seguinte são apresentados os resultados obtidos para as três primeiras técnicas (primeiro método). Os resultados das técnicas referentes ao segundo método são mostrados na seção 5.5. O desempenho destas técnicas é avaliado através de figuras, pela superposição dos modelos reconhecidos ao modelo da cena e de tabelas, contendo os valores medidos dos parâmetros definidos na seção 5.2.

## 5.4 Resultados obtidos pelas primeira, segunda e terceira técnicas

O processo de agrupamento de hipóteses produz grupos com pelo menos duas hipóteses. Cada grupo é ordenado pelo erro relativo  $el^{(\cdot)}$ . A primeira hipótese de cada grupo é aquela com menor valor de erro  $el^{(\cdot)}$ .

Para cada uma das técnicas, os modelos reconhecidos são posicionados sobre a cena empregando sua respectiva transformada, ou seja,  $T_h^{(h)}$  para a primeira técnica e  $T_m^{(h)}$

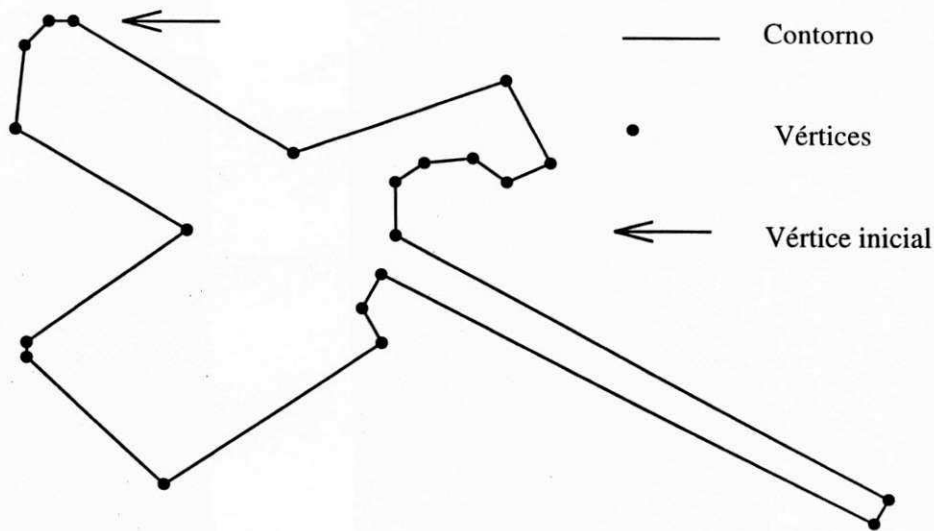


Figura 5.5: Modelo poligonal da cena 5.

para a segunda e terceira técnicas, descritas no capítulo 4.

Os resultados experimentais obtidos pelo emprego das três primeiras técnicas são mostrados a seguir. Estas três técnicas compõem o primeiro método de avaliação de hipóteses e são caracterizadas pela maneira de obter os parâmetros de translação, como descrito na subseção 4.3.1.

As três técnicas aplicadas às cenas 1 e 2 produzem os resultados mostrados na figura 5.6. Pode-se observar que para a cena 1 os resultados obtidos são praticamente idênticos (figuras 5.6.a e 5.6.b). Para a cena 2, a primeira técnica foi incapaz de reconhecer o objeto Arco, enquanto que os resultados produzidos pela segunda (figura 5.6.c) e terceira (figura 5.6.d) técnicas indicam um melhor desempenho daquela. Esta conclusão é comprovada pelos menores valores de  $\Delta \bar{D}_{pm}$ ,  $\Delta \bar{\theta}$  e  $f_g$ , produzidos pela segunda técnica, de acordo com a tabela 5.1. Entretanto, a terceira técnica reduz o percentual de oclusão  $p_{oc}$  de 40% para 19%, em relação à segunda técnica, às custas de valores maiores daqueles parâmetros.

Com relação à cena 3, os resultados produzidos pelas três técnicas são mostrados nas figuras 5.7, 5.8 e 5.9. Pode-se observar que o desempenho das três técnicas é bastante semelhante o que é comprovado pelos dados da tabela 5.1. As três técnicas

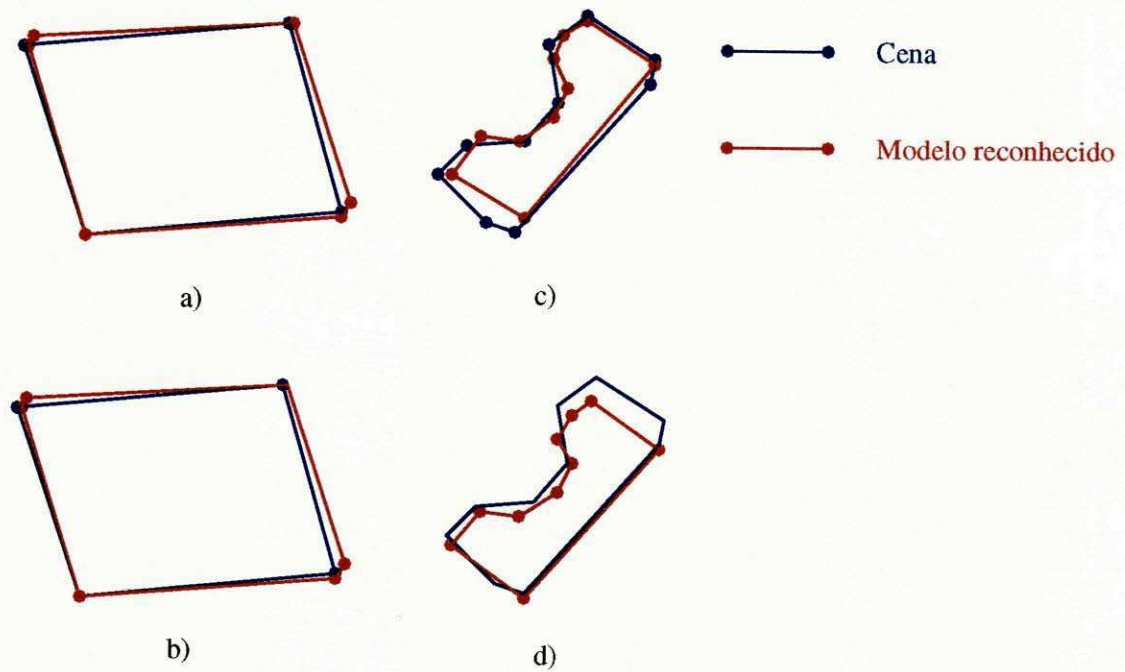


Figura 5.6: Modelos poligonais das formas reconhecidas superpostos aos modelos das cenas: a) cena 1 pela aplicação da primeira técnica, b) cena 1 pela aplicação da segunda e terceira técnicas, c) cena 2 pela aplicação da segunda técnica e d) cena 2 pela aplicação da terceira técnica (incluindo fragmentação).

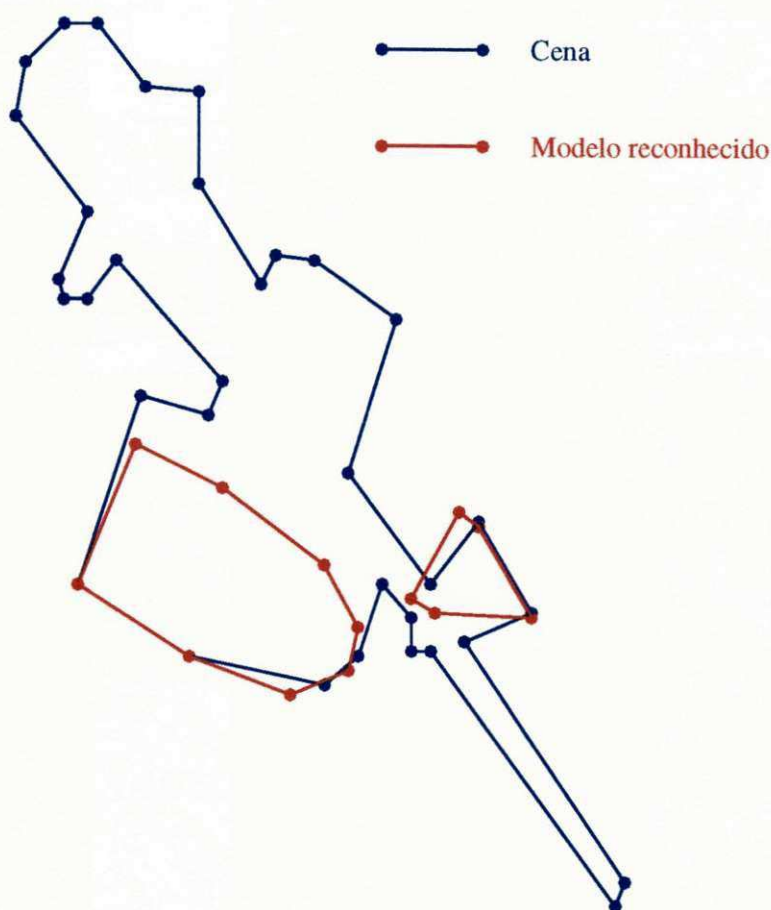


Figura 5.7: Resultado obtido pela aplicação da primeira técnica à cena 3.

foram capazes de identificar os objetos que deram origem a pelo menos um par de hipóteses verdadeiras, ou seja, o Cadeado e o Triang2. A terceira técnica comete um erro acidental ao identificar o objeto Chave numa posição incorreta. Isto ocorreu como consequência de uma hipótese gerada entre segmentos do objeto e da cena não pertencente aquele. Além disso, uma hipótese adicional é gerada como resultado de considerar o problema da fragmentação descrito na subseção 4.3.1 do capítulo 4, contribuindo ainda mais para o erro cometido.

A figura 5.10 mostra os resultados obtidos pela aplicação das três técnicas à cena 4. A única diferença notável de desempenho neste caso é com relação ao tempo  $t_{cl}$  (tabela 5.1), que aumenta com a complexidade da técnica, sendo portanto maior para

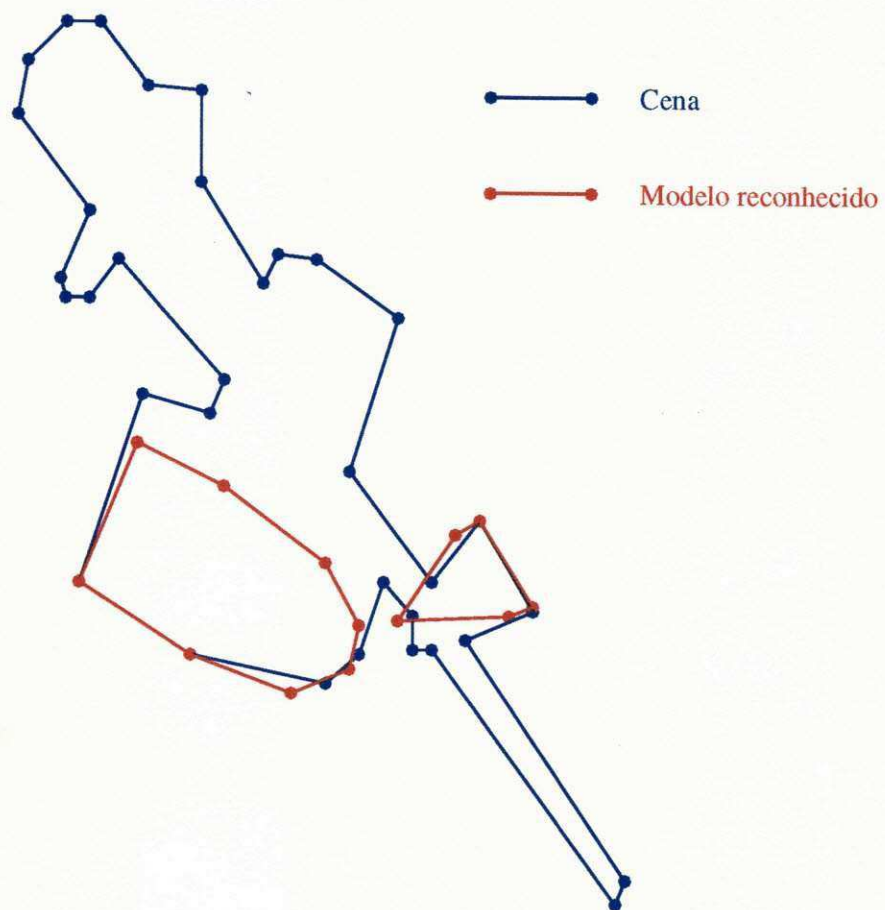


Figura 5.8: Resultado obtido pela aplicação da segunda técnica à cena 3.

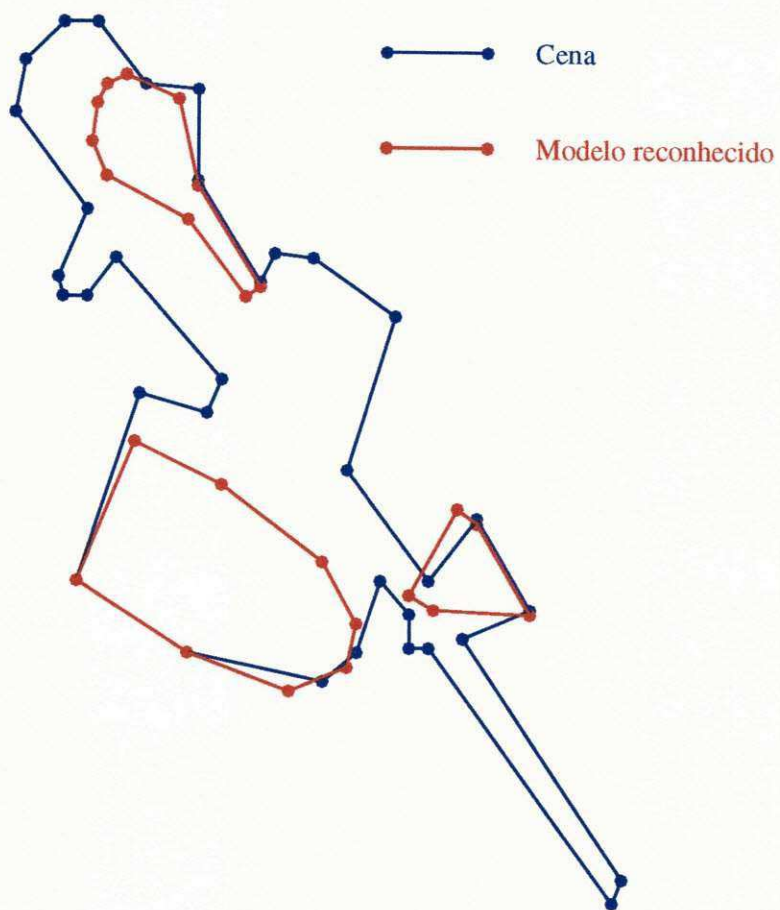


Figura 5.9: Resultado obtido pela aplicação da terceira técnica à cena 3.

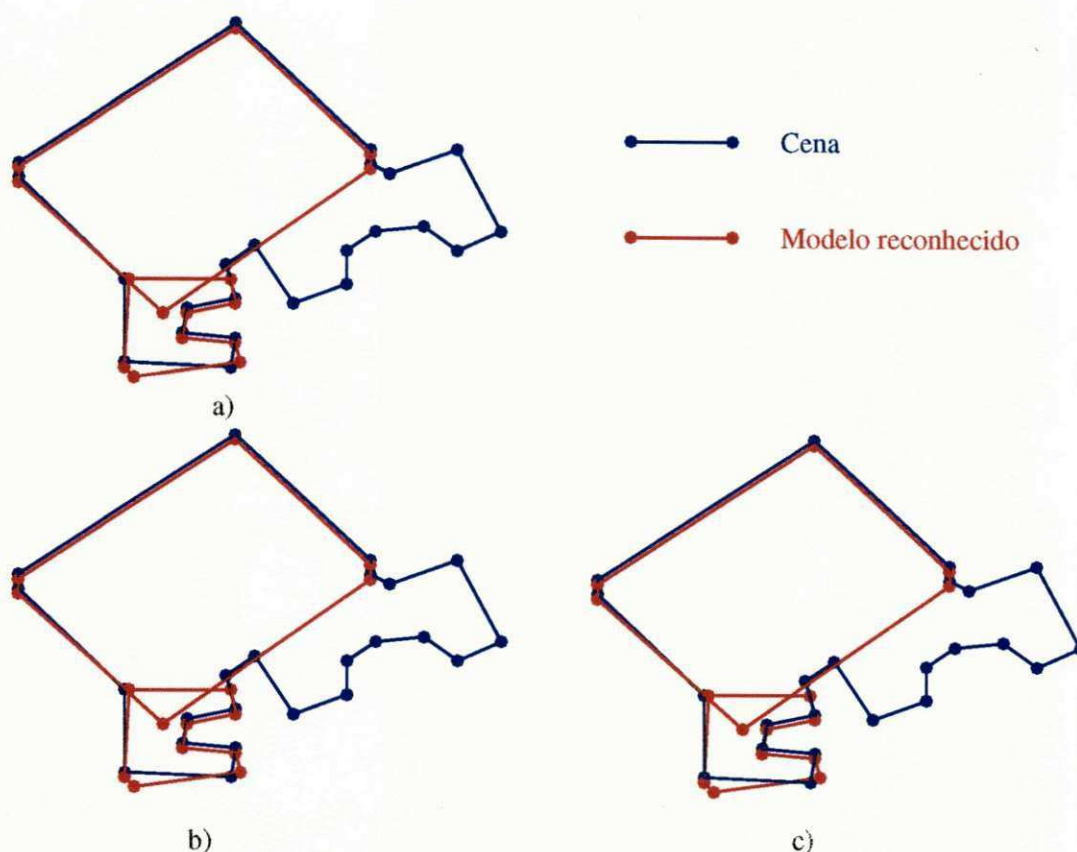


Figura 5.10: Resultado obtido pela aplicação das técnicas à cena 4: a) primeira, b) segunda e c) terceira.

a terceira técnica. Nenhuma delas foi capaz de reconhecer o objeto Arco, presente na cena.

Finalmente, os resultados obtidos com a cena 5 pela aplicação da primeira, segunda e terceira técnicas são mostrados nas figuras 5.11, 5.12 e 5.13, respectivamente. Novamente, as três técnicas apresentam desempenho praticamente idênticos. Os três objetos presentes na cena foram corretamente identificados. As pequenas diferenças de posicionamento que podem ser observadas nas figuras são irrelevantes na prática, como indicado pelos valores de parâmetros listados na tabela 5.1.

Os parâmetros medidos pela aplicação da primeira, segunda e terceira técnicas de

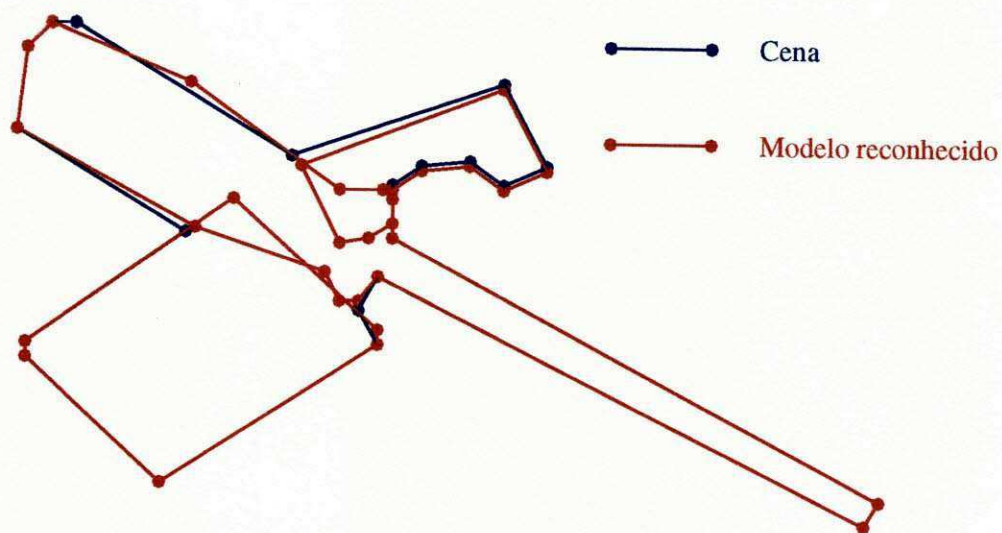


Figura 5.11: Resultado obtido pela aplicação da primeira técnica à cena 5.

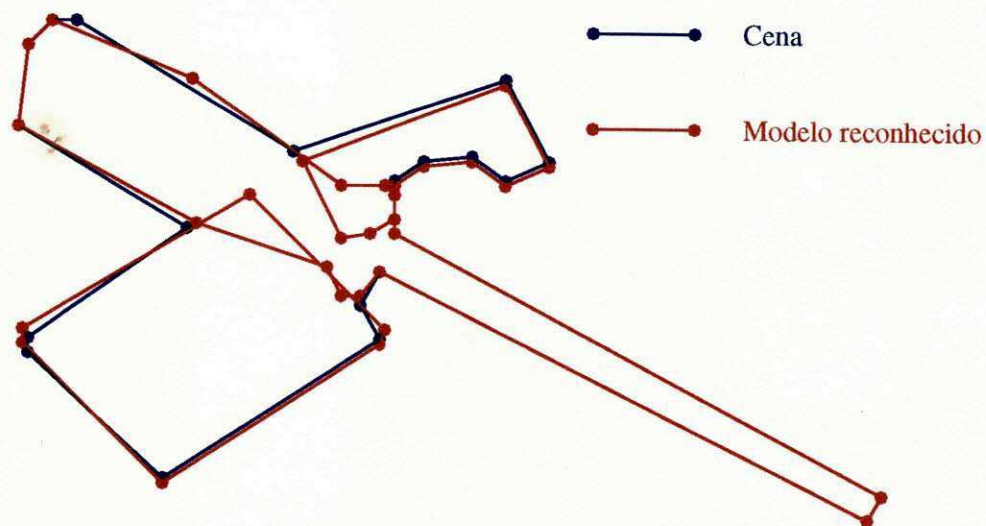


Figura 5.12: Resultado obtido pela aplicação da segunda técnica à cena 5.



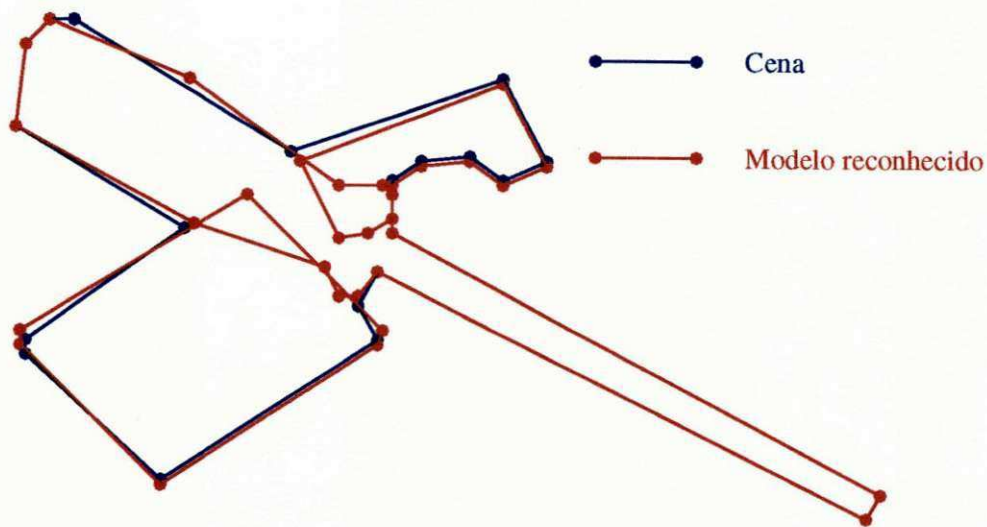


Figura 5.13: Resultado obtido pela aplicação da terceira técnica à cena 5.

avaliação de hipóteses são listados na tabela 5.1<sup>1</sup>.

Portanto, os resultados acima analisados indicam que dentre as três técnicas que compõem o primeiro método de avaliação de hipóteses, a segunda e a terceira técnicas apresentam um melhor desempenho do ponto de vista de identificação dos objetos presentes nas cenas. Isto é conseguido às custas de um maior tempo de processamento.

Para as cenas 3, 4 e 5, nas quais as três técnicas obtêm a mesma taxa de redução de hipóteses  $\tau_h$ , o tempo de classificação e localização  $t_{cl}$  aumenta com a complexidade da técnica.

Apesar de cometer um erro de classificação, a terceira técnica é a única que consegue reduzir o percentual de oclusão. Isto é realizado às custas de um acréscimo no tempo de verificação de hipóteses, dado a sua maior complexidade, e/ou de um acréscimo na função de grupo  $f_g$ . Este fato é comprovado na cena 5, com o objeto Fenda, e na cena 2. Também na cena 4 para o objeto Arco (presente na cena), apesar de não identificado, esta técnica consegue aumentar o perímetro parcial máximo  $P_{p(max)}$  deste objeto para 49%. As outras técnicas obtêm um perímetro parcial máximo para este objeto de apenas 24%.

<sup>1</sup>\* Significa que o objeto não foi reconhecido, portanto os parâmetros não podem ser medidos.

Primeira técnica							
Cenas	$t_{cl}$ (ms)	$\tau_h$	Objetos	$\Delta \bar{D}_{pm}$ (pixels)	$\Delta \bar{\theta}^o$	$f_g$	$p_{oc}(\%)$
1	020.4	1.56	Bloco	1.06	01	0.64	00
2	046.4	2.91	Arco	*	*	*	*
3	294.0	1.86	Cadeado	2.43	04	1.49	48
			Triang2	2.00	04	1.22	49
4	103.5	2.41	Arco2	1.00	04	0.62	39
			Bloco	2.00	01	1.20	35
5	104.3	2.47	Arco	0.48	00	0.29	31
			Bloco	2.22	00	1.33	30
			Fenda	0.35	01	0.22	47
Segunda técnica							
1	019.7	3.00	Bloco	0.33	00	0.20	00
2	050.3	3.32	Arco	2.27	06	1.40	40
3	366.0	1.86	Cadeado	2.43	04	1.49	48
			Triang2	1.12	04	0.70	49
4	118.7	2.41	Arco2	1.00	04	0.62	39
			Bloco	2.00	01	1.20	35
5	080.0	2.47	Arco	0.40	00	0.24	31
			Bloco	1.20	02	0.73	30
			Fenda	0.25	01	0.16	47
Terceira técnica							
1	019.6	1.56	Bloco	0.50	01	0.31	00
2	062.0	2.91	Arco	3.16	10	1.97	19
3	402.0	1.86	Cadeado	1.22	05	0.76	48
			Chave	1.61	04	1.00	48
			Triang2	1.12	05	0.70	49
4	165.2	2.41	Arco2	0.75	04	0.47	39
			Bloco	0.64	00	0.39	35
5	113.5	2.47	Arco	0.40	00	0.24	31
			Bloco	1.20	02	0.73	29
			Fenda	0.71	01	0.43	34

Tabela 5.1: Resultados obtidos pela primeira, segunda e terceira técnicas para as cinco cenas.

Do ponto de vista de posicionamento dos modelos na cena, a terceira técnica produz os melhores resultados. Este fato pode ser observado por um maior equilíbrio entre as variações angular  $\Delta\bar{\theta}$  e de distância  $\Delta\bar{D}_{pm}$  para quase todos os objetos nas cenas, listados na tabela 5.1.

Os resultados apresentados nesta seção podem ser melhorados pelo emprego de técnicas mais complexas, cujos resultados são mostrados na seção seguinte.

## 5.5 Resultados obtidos pelas quarta e quinta técnicas

Os resultados apresentados nesta seção se referem à quarta e quinta técnicas, descritas na subseção 4.3.2. Apenas a quinta técnica emprega o processo de filtragem de hipóteses por limiar descrito na subseção 4.3.2.

Os parâmetros medidos, resultantes da aplicação da quarta e quinta técnicas são listados na tabela 5.2.

A quarta técnica de avaliação de hipóteses não foi capaz de identificar o objeto Arco presente na cena 2. Na figura 5.14.a é ilustrado o resultado obtido com a cena 1, onde o objeto Bloco foi identificado. Também nesta figura, pode-se observar os resultados produzidos pela quinta técnica nas cenas 1 e 2, 5.14.b e 5.14.c, respectivamente. A diferença favorável à quinta técnica é comprovada pelos valores dos parâmetros listados na tabela 5.2. Observa-se também que esta técnica distribui mais uniformemente o erro de posicionamento entre os segmentos da cena e do modelo. Isto acarreta um menor valor de  $f_g$  e portanto em um melhor posicionamento do objeto sobre a cena.

Os resultados obtidos com a cena 3 são claramente favoráveis à quinta técnica, como pode ser observado nas figuras 5.15 e 5.16. Esta superioridade é verdadeira tanto no que se refere ao reconhecimento (a quarta técnica foi incapaz de identificar o Triang2) quanto em termos de posicionamento dos modelos na cena, indicados por menores valores dos parâmetros  $\Delta\bar{D}_{pm}$ ,  $\Delta\bar{\theta}$  e  $f_g$  (tabela 5.2).

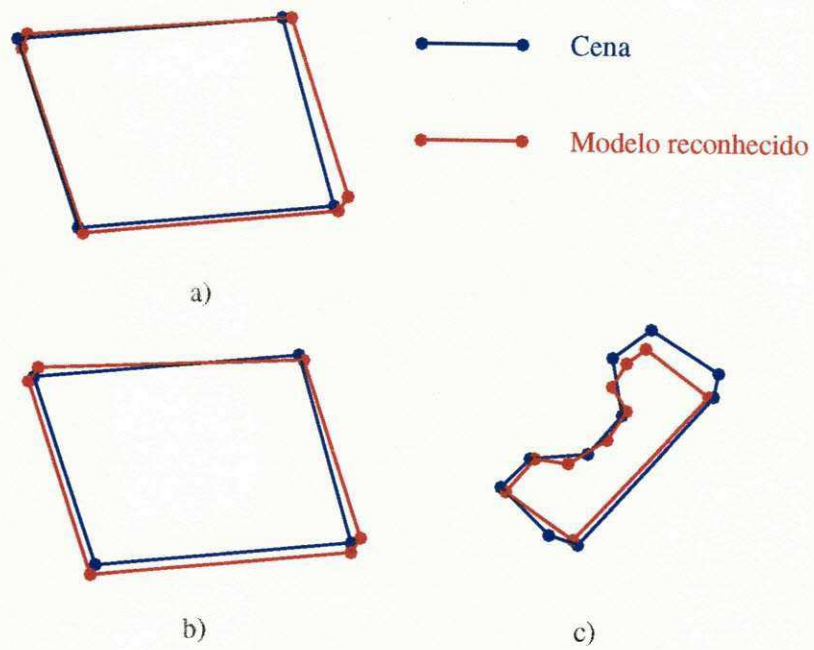


Figura 5.14: Modelos poligonais das duas cenas contendo os objetos reconhecidos resultante da aplicação das duas técnicas: a) quarta técnica à cena 1, b) quinta técnica à cena 1 e c) quinta técnica à cena 2.

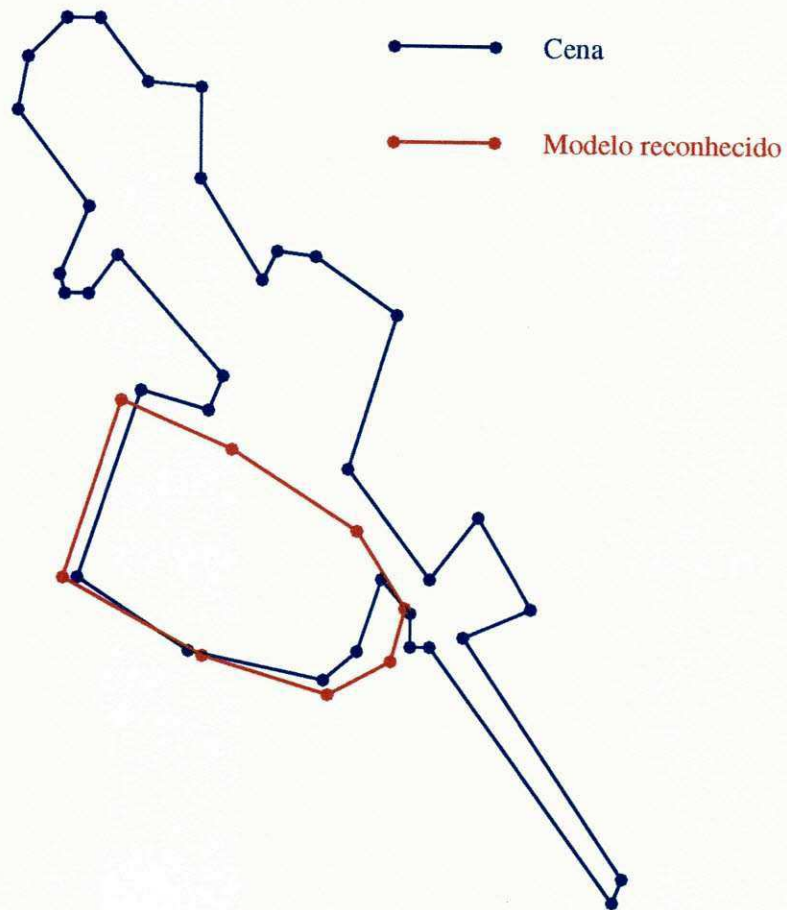


Figura 5.15: Resultados obtidos pela quarta técnica aplicada à cena 3.

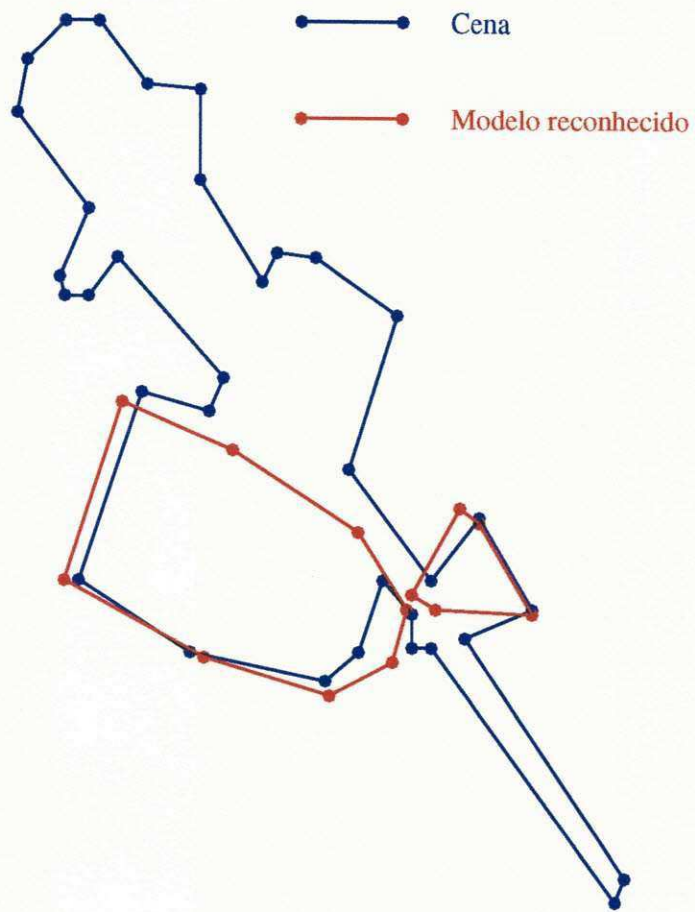


Figura 5.16: Resultados obtidos pela quinta técnica aplicada à cena 3.

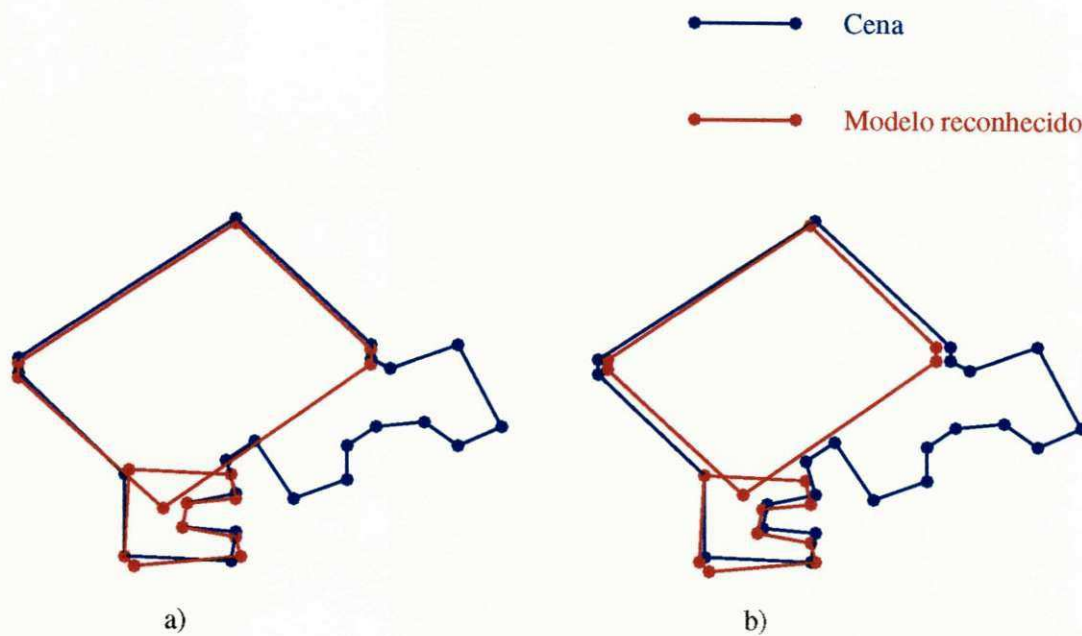


Figura 5.17: Resultados obtidos pelas duas técnicas aplicadas à cena 4: a) quarta técnica e b) quinta técnica.

Com a cena 4, como já havia ocorrido antes com o primeiro método, novamente o objeto Arco não foi reconhecido. Este resultado é mostrado na figura 5.17 para as duas técnicas que compõem o segundo método. Para os demais objetos presentes na cena, o desempenho das duas técnicas é bastante semelhante.

Finalmente, os testes realizados com a cena 5 levaram ao reconhecimento de todos os objetos presentes, com ambas as técnicas utilizadas. Isto é mostrado nas figuras 5.18 e 5.19. Apenas para o objeto Fenda, a quarta técnica resultou em um posicionamento ligeiramente superior ao da quinta técnica, comprovado pelos valores dos parâmetros listados na tabela 5.2.

Dentre as duas técnicas que compõem o segundo método de avaliação de hipóteses, a quinta é a única que apresenta um desempenho plenamente satisfatório do ponto de vista de identificação dos objetos presentes nas cenas.

Os resultados indicam que para as cenas 3, 4 e 5, nas quais as duas técnicas obtêm a mesma taxa de redução de hipóteses  $\tau_h$ , o tempo de classificação e localização  $t_{cl}$

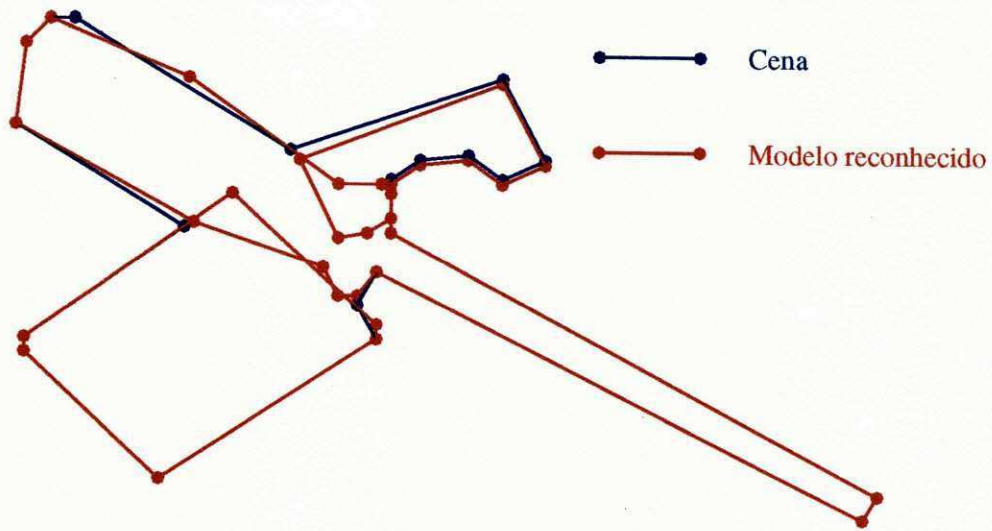


Figura 5.18: Resultados obtidos pela quarta técnica aplicada à cena 5.

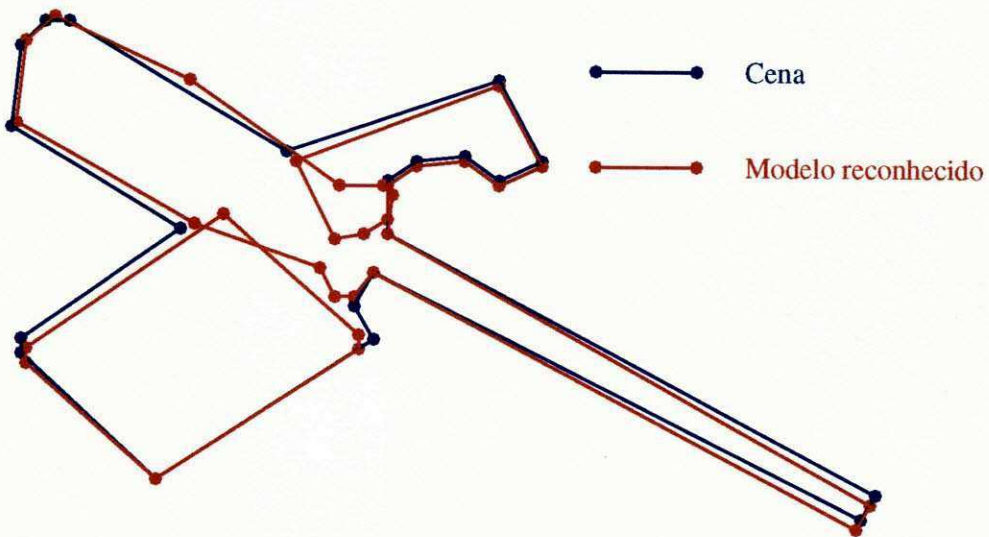


Figura 5.19: Resultados obtidos pela quinta técnica aplicada à cena 5.



Quarta técnica							
Cenas	$t_{cl}$ (ms)	$\tau_h$	Objetos	$\Delta \bar{D}_{pm}$ (pixels)	$\Delta \bar{\theta}^\circ$	$f_g$	$p_{oc}(\%)$
1	013.6	3.00	Bloco	1.18	01	0.71	00
2	034.4	2.91	Arco	*	*	*	*
3	242.9	1.71	Cadeado	1.35	03	0.83	48
			Triang2	*	*	*	*
4	81.6	2.41	Arco2	1.05	03	0.65	39
			Bloco	1.50	00	0.90	35
5	079.5	2.47	Arco	0.40	00	0.24	31
			Bloco	1.66	00	1.00	30
			Fenda	0.00	01	0.00	47
Quinta técnica							
1	014.8	1.56	Bloco	1.09	01	0.66	00
2	035.3	2.91	Arco	1.36	05	0.85	40
3	262.9	1.71	Cadeado	0.91	03	0.57	48
			Triang2	0.47	02	0.30	49
4	085.4	2.41	Arco2	0.71	03	0.44	46
			Bloco	0.67	00	0.40	31
5	097.2	2.47	Arco	0.33	00	0.20	31
			Bloco	0.71	01	0.43	30
			Fenda	0.47	01	0.29	47

Tabela 5.2: Resultados obtidos pela quarta e quinta técnicas para as cinco cenas.

aumenta com a complexidade da técnica.

Do ponto de vista de posicionamento dos modelos na cena, a quinta técnica produz os melhores resultados. Este fato pode ser observado por um maior equilíbrio entre as variações angular  $\Delta\bar{\theta}$  e de distância  $\Delta\bar{D}_{pm}$  para quase todos os objetos nas cenas, listados na tabela 5.2. Como resultado, em geral, o valor da função discriminante de grupo  $f_g$  é menor para esta técnica do que para a quarta técnica.

Os resultados obtidos pela aplicação das cinco técnicas de avaliação de hipóteses são analisados em conjunto na seção seguinte, bem como são tiradas algumas conclusões.

## 5.6 Análise geral dos resultados e conclusões

O tempo gasto para realizar a tarefa de reconhecimento e localização do modelo na cena é função do número de hipóteses e da complexidade da técnica usada no processo de verificação de hipóteses.

Os resultados apresentados nas seções anteriores indicam que a velocidade de processamento das técnicas depende da taxa de redução de hipóteses  $\tau_h$ . Isto significa que técnicas mais complexas podem ser mais rápidas, desde que apresentem uma maior taxa de redução de hipóteses. Um exemplo deste fato ocorre para a cena 1, como pode ser observado comparando-se os dados referentes à primeira e quarta técnicas, listados nas tabelas 5.1 e 5.2, respectivamente.

Para cenas nas quais as mesmas taxas  $\tau_h$  são obtidas, o tempo de execução das tarefas de reconhecimento e localização cresce, em geral, com a complexidade das técnicas, como esperado. Este fato pode ser comprovado pelos dados mostrados na tabela 5.1 para as cenas 3, 4 e 5 para o primeiro método de avaliação. Comprovado também pelos dados listados na tabela 5.2 para o segundo método e por análise das mesmas cenas.

A aplicação da quarta técnica às cenas 2 e 5 produziu resultados diferentes para o mesmo objeto Arco presentes nas duas cenas, ou seja, este objeto é reconhecido na cena 5 e não o é na cena 2. Este fato se explica pela natureza das duas cenas,

por ser gerada artificialmente a partir do modelo original, a cena 5 preserva muitos segmentos daquele, facilitando desta forma a sua identificação. Por outro lado, a cena 2 é obtida pela rotação, em torno do primeiro ponto de borda, da imagem gradiente do objeto original e passa por todo o processo de segmentação descrito na seção 3.3. Este processo produz maiores variações nos modelos se comparado ao processo de obtenção do modelo da cena 5.

A terceira técnica, em geral, reduz o percentual de oclusão  $p_{oc}$ , como pode ser observado nos dados listados nas tabelas 5.1 e 5.2 para as cenas 2, 3 (objeto Cadeado) e 5 (objetos Bloco e Fenda). Na tentativa de reduzir este percentual, esta técnica pode cometer erros como ocorreu com a cena 3, na qual é identificado acidentalmente o objeto Chave (presente na cena), cujos segmentos casados não correspondem aos do modelo. Isto resultou num errado posicionamento do modelo (figura 5.9). Entretanto, na análise da cena 4, consegue acumular um perímetro parcial máximo  $P_{p(max)}$  de 47% do objeto Arco, presente na cena, contra apenas 24% obtidos pelas outras técnicas.

Quanto à classificação dos objetos presentes nas cenas, aparecendo isolados ou oclusos, a segunda e quinta técnicas executam esta tarefa sem cometer erros de classificação. A primeira e quarta técnicas são incapazes de reconhecer o objeto Arco, presente na cena 2. A terceira técnica comete apenas o erro acidental descrito anteriormente.

A quinta técnica reduz as variações médias de distância e de orientação, produzindo um posicionamento relativo mais centralizado entre os segmentos dos objetos reconhecidos e os segmentos da cena. Este fato pode ser comprovado também por observação visual das figuras apresentadas na seção anterior e pelo valor da função discriminante de grupo  $f_g$ , listados nas tabelas.

Com base nos resultados mostrados na seção anterior e nos comentários acima, podem ser tiradas as seguintes conclusões:

1. o tempo de execução das tarefas de reconhecimento e localização dos objetos presentes nas cenas é inversamente proporcional à taxa de redução de hipóteses, indicando a importância do emprego da estratégia de controle do número de hipóteses geradas,

2. a *terceira* técnica pode possibilitar o reconhecimento de objetos presentes na cena que não foram identificados pelas outras técnicas, por incorporar segmentos do modelo e da cena até então não casados,
3. as técnicas terceira e quinta produzem menores erros de localização dos objetos, avaliados pelas variações médias de orientação e de distância,
4. portanto, as *terceira* e *quinta* técnicas de avaliação de hipóteses ou uma combinação destas, são as mais adequadas para uso no sistema de visão proposto.

## Capítulo 6

# Conclusões, contribuições e trabalhos futuros

Muito embora diversos sistemas de visão 2D (bidimensional) tenham sido desenvolvidos nos últimos anos, esta ainda constitui uma importante área de pesquisa, como pode ser comprovado pelo constante fluxo de trabalhos publicados. Esta relevância deve-se ao fato de que as soluções de problemas em 2D conduzem muitas vezes às soluções de problemas em 3D e pela existência de um grande número de situações reais que podem ser resolvidas satisfatoriamente em 2D.

O trabalho apresentado nesta tese abordou aspectos não ainda explorados do problema de reconhecimento de formas bidimensionais em imagens, através do casamento de modelos poligonais. Em consequência, vários resultados originais foram obtidos, trazendo contribuições para o estado da arte nesta área. As principais destas contribuições são listadas a seguir:

1. o desenvolvimento de um algoritmo de segmentação por rastreamento de bordas, chamado de BTVE (*boundary tracking* com vizinhança expandida) [71];
2. o desenvolvimento de um algoritmo seqüencial para a modelagem de contornos 2D, o qual produz modelos com precisão e complexidade equivalentes aos algoritmos considerados como clássicos, mas com um tempo de processamento inferior

[86, 81],

3. o estabelecimento de uma nova metodologia de reconhecimento de formas em 2D. Os seguintes aspectos inovadores podem ser destacados nesta metodologia:

- (a) realiza simultaneamente as etapas de segmentação, modelagem e geração de hipóteses,
- (b) permite reconhecer objetos isolados ou parcialmente visíveis, com o emprego do algoritmo seqüencial de modelagem de contornos,
- (c) utiliza processos de controle e agrupamento de hipóteses baseados em critérios objetivos que dependem do processo de modelagem e do conjunto de modelos usados, reduzindo o tempo de processamento,
- (d) introduz técnica para verificação de hipóteses que permite melhorar o posicionamento relativo entre os modelos casados e o modelo da cena, resultando em maior precisão na localização das formas reconhecidas,
- (e) aborda o problema da fragmentação de segmentos, em decorrência da oclusão de objetos num mesmo contorno, pelo acréscimo no valor do limiar  $ed_{max}$  (equação 4.11) no processo de verificação de hipóteses,
- (f) introduz dois métodos que se subdividem em cinco técnicas distintas para a validação de hipóteses,
- (g) pode ser aplicada em situações onde um baixo tempo de processamento é exigido (da ordem de 0.5 segundos).

Adicionalmente, com relação às técnicas de avaliação de hipóteses referida no item f) acima, podem ser destacados os seguintes aspectos inovadores:

1. o emprego de funções discriminantes para agrupar e ordenar as hipóteses,
2. as funções discriminantes utilizam um limiar máximo variável para o ângulo entre segmentos casados. Este limiar possibilita a integração entre os processos de modelagem e de reconhecimento, através das funções discriminantes.

Entretanto, o problema aqui abordado não pode ainda ser considerado esgotado. Abaixo são listados sugestões para trabalhos futuros:

1. implementação de uma técnica de verificação de hipóteses, resultante da integração das terceira e quinta técnicas testadas neste trabalho,
2. elaboração de um filtro para os parâmetros de translação, baseado nas propriedades geométricas dos modelos produzidos durante o processo de treinamento,
3. desenvolvimento de um processo de agrupamento de segmentos semelhantes pertencentes aos modelos, a ser aplicado ainda no processo de treinamento. Para tanto, seria utilizada a mesma técnica de agrupamento de hipóteses. Espera-se com isto obter uma redução no tempo de geração de hipóteses,
4. desenvolvimento de algoritmos paralelos, segundo a técnica de fluxo de dados, para os processos de segmentação, modelagem e reconhecimento. Implementação destes algoritmos em VLSI, de acordo com a metodologia de "codesign software-hardware". Este desenvolvimento resultará em um sistema de visão de alto desempenho, adequado para aplicações em tempo real. Trabalhos nesta direção já estão em andamento no LAPS/DEE,
5. aplicar a metodologia desenvolvida para o caso de reconhecimento tridimensional através de modelos poliedrais. Este trabalho também já foi iniciado na COPELE/LAPS/DEE,
6. adaptar o sistema desenvolvido para utilização em inspeção automática de peças.

Como continuação imediata do trabalho apresentado nesta tese, a metodologia de modelagem e reconhecimento de formas proposta será implementada em máquinas com capacidade para multiprocessamento. Pretende-se com isso utilizar plenamente a característica paralela do método, que permite realizar simultaneamente as tarefas envolvidas. Este projeto será desenvolvido em conjunto pela UFPB-Campus II e a

Universidade Federal do Ceará (UFC)<sup>1</sup>. Na UFC já se encontra instalado um Centro de Processamento de Alto Desempenho (CENAPAD) equipado com uma máquina paralela, capaz de realizar o processamento desejado.

---

<sup>1</sup>Envolvendo os Departamentos de Engenharia Elétrica e de Computação das duas instituições.



## Referências

- [1] Chin, R. T., and Dyer, C. R. (1986). "Model-Based Recognition in Robot Vision". *ACM Computing Surveys*, 18(1):67-108, March.
- [2] Filho, O. M. (1990). "Reconhecimento de Caracteres Ópticos". *Jornada EPUSP/IEEE em Computação Visual*, pages 191-199.
- [3] Frias, B. C. e Guigoitia, C. F. B. (1987). "*Visión para Computadoras*". Facultad de Ingeniería - Universidad de Buenos Aires, Buenos Aires.
- [4] Ayache, N. (1983). "Un Système de Vision Bidimensionnelle en Robotique Industrielle". *L'Université de Paris-Sud Centre D'Orsay, Thèse de Docteur Ingénieur*.
- [5] Rosenfeld, A. (1988). "Computer Vision: basics principles". *Special Issue on Computer Vision, Proceedings IEEE*, 76(8), August.
- [6] Marshall, A. D., and Martin, R. R. (1992). "*Computer Vision, Models and Inspection*". "World Publishing Company - Series in Robotics and Automated Systems - Vol. 4, London - England.
- [7] Barrera, J. (1992). "Uma Abordagem Unificada para Problemas de Visão Computacional: a Morfologia Matemática". *Instituto de Pesquisas Espaciais - INPE, Tese de Doutorado*, Julho.
- [8] Mascarenhas, N.D.A. e Velasco, F.R.D. (1989). "*Processamento Digital de Imagens*". "Ministério da Ciência e Tecnologia e Instituto Nacional de Pesquisas Espaciais - 2ª Edição".

- [9] Suetens, P., Fua, P., and Hanson A. J. (1992). "Computational Strategies for Object Recognition". *ACM Computing Surveys*, 24(1):5-61, March.
- [10] Li, H., Derklerck, R., Cuyper B., Hermanus, A., Nyssen, E., and Cornelis, J. (1995). "Object Recognition in Brain CT-Scans: Knowledge-Based Fusion of Data from Multiple Feature Extractors". *IEEE - Transactions on Medical Image*, 14(2):212-229, June.
- [11] Bins, L. S. A. (1989). "Uso de Casamento Estrutural para Registro de Imagens de Satélite". *Instituto de Pesquisas Espaciais - INPE, Dissertação de Mestrado*.
- [12] Fukumi, M., Omatu, S., Takeda, F., and Kosaka, T. (1992). "Rotation-Invariant Neural Pattern Recognition System with Application to Coin Recognition". *IEEE - Transactions on Neural Network*, 3(2):272-279, March.
- [13] Ayache, N., and Faugeras, O. D. (1986). "Hyper: A New Approach for the Recognition and Positioning of Two-Dimensional Objects". *IEEE - Trans. Pattern Anal. Mach. Intell*, 8(1):44-54, January.
- [14] Wong, E. K. (1992). "Model Matching in Robot Vision by Subgraph Isomorphism". *Pattern Recognition*, 25(3):287-303.
- [15] Chen, Z., and Ho, S. (1991). "Computer Vision for Robust 3D Aircraft Recognition with Fast Library Search". *Pattern Recognition*, 24(5):375-390.
- [16] Tsang, K. M., and To, F. W. (1994). "Three Dimensional Shape Recognition". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, WA2.2:25-29, 9-11 November.
- [17] Zhang, G., Thirion, E., and Wallace, A. M. (1993). "Model Matching Using Multiple Data Sources". *SPIE - Machine Vision Applications in Industrial Inspection*, 1907:2-12.
- [18] La Valle, S. M., and Hutchinson, S. A. (1995). "A Bayesian Segmentation Methodology for Parametric Image Models". *IEEE - Trans. Pattern Anal. Mach. Intell*, 17(2):211-217, February.

- [19] Paik, J. K., Brailean, J. C., and Katsaggelos, A. K. (1992). "An Edge Detection Algorithm Using Multi-State Adalines". *Pattern Recognition*, 25(12):1495-1504.
- [20] Yang, H., Poso, S., and Halme, A. (1994). "A Region-Segmentation Method Based on Data of Samples". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, WA1.1:1-5, 9-11 November.
- [21] L, T., Itoh, H., and Seki, H. (1994). "Filtering images for edge detection using fuzzy reasoning". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, WA1.4:15-19, 9-11 November.
- [22] Yu, S., and Tsai, W. (1992). "Relaxation by the Hopfield Neural Network". *Pattern Recognition*, 25(2):197-209.
- [23] Hara, K., Zha, H. B., and Nagata, T. (1994). "Contour Segmentation by Using an Extended Koch Neural Network". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, TE7.2:1309-1313, 9-11 November.
- [24] Asano, T., Kenwood, G., Mochizuki, J., and Hata, S. (1986). "Color Image Recognition Using Chrominance Signals". *Proc. 8th International Conference Pattern Recognition*.
- [25] Mital, D. P., Leng, G. W., and Khwang, T. E. (1990). "Colour Vision for Industrial Applications". *10th International Conference TECON, Bombaim - India*, pages 548-551, November.
- [26] Sepeda, I. H. e Stemmer, M. R. (1994). "Um Sistema de Reconhecimento de Peças Baseado em Redes Neurais". *10<sup>o</sup> Congresso Brasileiro de Automática, Rio de Janeiro-RJ*, 2:1225-1227, Setembro.
- [27] Rillo, A.H.R.C. (1989). "Um Sistema de Visão Binária para Reconhecimento de Peças Isoladas e Parcialmente Oclusas". *Laboratório de Sistemas Integráveis - DEE - EPUSC - São Paulo*.

- [28] Turney, J. L., Mudge, T. N., and Volz, R. A. (1985). "Recognizing Partially Occluded Parts". *IEEE - Trans. Pattern Anal. Mach. Intell*, (7):410-421, July.
- [29] Bhanu, B., and Faugeras, O. D. (1984). "Shape Matching of Two-Dimensional Objects". *IEEE - Trans. Pattern Anal. Mach. Intell*, (6):137-152, March.
- [30] Ayache, N., and Faugeras, O. D. (1982). "Recognition of Partially Visible Planar Shapes". *Proc. 6th Conference Pattern Recognition, Munich*, Septembre.
- [31] Fukuda, T., Ishizuka, M., Hasegawa, O., Asama, H., Nagamune, T., and Endo, I. (1990). "Vision System for Animal Cell Recognition in a Bio-Engineering Process". *10th International Conference TECON, Bombaim - India*, pages 552-557, November.
- [32] Maes, M. (1991). "Polygonal Shape Recognition Using String-Matching Techniques". *Pattern Recognition*, 24(5):433-440.
- [33] Bousofiane, F., and Bertrand, G. (1992). "A New Method for Recognizing and Locating objects by Searching Longest Paths". *IAPR Workshop on Machine Vision, Tokyo*, pages 445-448.
- [34] Wen, W. and Lozzi, A. (1992). "Recognition and Inspection of Two-dimensional Industrial Parts Using Subpolygons". *Pattern Recognition*, 25(12):1427-1434.
- [35] Stein, F., and Medioni, G. (1992). "Structural Indexing: Efficient 2-D Object Recognition". *IEEE - Trans. Pattern Anal. Mach. Intell*, (12):1198-1204, December.
- [36] Sethi, I. K., and Ramesh, N. (1992). "Local Association Based Recognition of Two-Dimensional Objects". *Machine Vision and Applications*, (5):265-276.
- [37] Tsang, P. W. M., Yuen, P. C., and Lam, F. K. (1992). "Recognition of Occluded Objects". *Pattern Recognition*, 25(10):1107-1117.
- [38] Nasrabadi, N. M., and Li, W. (1991). "Object Recognition by a Hopfield Neural Network". *IEEE - Transactions on Systems, Man, and Cybernetics*, 21(6):1523-1535, November/December.

- [39] Liu, H., and Srinath, M. D. (1990). "Corner Detection from Chain-code". *Pattern Recognition*, 23(12):51-68.
- [40] Basak, J., and Pal, S. K. (1995). "PsyCOP: A Psychologically Motivated Connectionist System for Object Perception". *IEEE - Transactions on Neural Network*, 6(6):1337-1354, November.
- [41] Cortez, P. C. e Carvalho, J. M. (1992). "Reconhecimento de Objetos Bidimensionais Usando Modelos Poligonais". *V Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, 1(329-336), Águas de Lindóia-SP, Novembro.
- [42] Bebis, G. N., and Papadourakis, G. M. (1991). "Object Recognition Using Invariant Object Boundary Representations and Neural Network Models". *Pattern Recognition*, 25(1):25-44.
- [43] Yong, C. K., and Rockett, P. I. (1994). "A Modular High Order Neural Network for Vision Inspection". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, TE7.1:1304-1308, 9-11 November.
- [44] Pao, D. C. W., Li, H. F., and Jayakumar, R. (1992). "Shapes Recognition Using the Straight Line Hough Transform: Theory and Generalization". *IEEE - Trans. Pattern Anal. Mach. Intell*, 14(11):1076-1089, November.
- [45] He, Y., and Kundu, A. (1991). "2-D Shape Classification Using Hidden Markov Model". *IEEE - Trans. Pattern Anal. Mach. Intell*, 13(11):1172-1184, November.
- [46] Veltman, S. R., and Prasad, R. (1994). "Hidden Markov Models Applied to On-Line Handwritten Isolated Character Recognition". *IEEE - Transactions on Image Processing*, 3(3):314-319, May.
- [47] Tchoukanov, I., Safaee-Rad, R., Smith, K. C., and Benhabib, B. (1992). "The Angle-of-Sight Signature for Two-dimensional Shape Analysis of Manufactured Objects". *Pattern Recognition*, 25(11):1289-1305.

- [48] Das, M., Paulik, M. J., and Loh, N. K. (1990). "A Bivariate Autoregressive Modeling Technique for Analysis and Classification of Planar Shapes". *IEEE - Trans. Pattern Anal. Mach. Intell*, 12(1):97-103, January.
- [49] Chutatape, O. (1994). "On the Use of Fourier Descriptor with Neural Network for Pattern Recognition". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, TP7.6:1084-1088, 9-11 November.
- [50] Blumenkrans, A. (1991). "Two-Dimensional Object Recognition Using a Two-Dimensional Polar Transform". *Pattern Recognition*, 24(9):879-890.
- [51] Perantonis, S. J., and Lisboa, P. J. G. (1992). "Translation, Rotation, and Scale Invariant Pattern Recognition by High-Order Neural Networks and Moment Classifiers". *IEEE - Transactions on Neural Network*, 3(2):241-251, March.
- [52] Schmitt, M., and Mattioli, J. (1991). "Shapes Recognition Combining Mathematical Morphology and Neural Networks". *SPIE - Applications of Artificial Neural Networks II*, 1469:392-403.
- [53] Trahanias, P. E. (1992). "Binary Shape Recognition Using the Morphological Skeleton Transform". *Pattern Recognition*, 25(11):1277-1288.
- [54] Pitas, I., and Venetsanopoulos, A. N. (1992). "Morphological Shape Representation". *Pattern Recognition*, 25(6):555-565.
- [55] Vlontzos, J. A., and Kung, S. Y. (1992). "Hidden Markov Models for Character Recognition". *IEEE - Transactions on Image Processing*, 1(4):539-543, October.
- [56] Umeyama, S. (1988). "An Eigendecomposition Approach to Weighted Graph Matching Problems". *IEEE - Trans. Pattern Anal. Mach. Intell*, 10(5):695-703, September.
- [57] Bins, L. S. A., Velasco, F. R. D. e Monteiro, A. M. V. (1989). "Object Detection Using Structural Matching". *projeto ESTRA-SID informática no Instituto de Pesquisas Espaciais*.

- [58] Li, S. Z. (1992). "Matching: Invariant to Translations, Rotations and Scale Changes". *Pattern Recognition*, 25(6):583-594.
- [59] Christmas, W. J., Kittler, J., and Petrou, M. (1995). "Structural Matching in Computer Using Probabilistic Relaxation". *IEEE - Trans. Pattern Anal. Mach. Intell.*, 17(8):749-764, August.
- [60] Suganthan, P. N., Teoh, E. K., and Mital, D. P. (1994). "Optimal Mapping of Graph Matching onto Hopfield Network". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, TA7.2:808-812, 9-11 November.
- [61] Barrow, H. G., and Tenenbaum, J. M. (1981). "Computational Vision". *Proc. IEEE*, vol. 69, (5), May.
- [62] Cheng, F. H., and Hsu, W. H. (1988). "Parallel Algorithm for Corner Finding on Digital Curves". *Pattern Recognition Letters*, (8):47-53, July.
- [63] Cortez, P. C., Carvalho, J. M. e Morais, M. E. (1993). "Comparação de 3 Algoritmos de Modelagem Poligonal de Contornos Bidimensionais". *Universidade Federal da Paraíba - Relatório Interno do LAPS - COPELE*, Julho.
- [64] Cortez, P. C. (1992). "Reconhecimento de Objetos Bidimensionais Usando Modelagem Poligonal". *Universidade Federal da Paraíba - Dissertação de Mestrado*, Agosto.
- [65] Arman, F., and Aggarwal, J. K. (1993). "Model-Based Object Recognition in Dense-Range Images - A Review". *ACM Computing Surveys*, 25(1):5-43, March.
- [66] Shapiro, V. A. (1994). "Optimal Binarization a Robust Approximation Scheme". *The Third International Conference on Automation, Robotics and Computer Vision - ICARV'94*, WP1.4:163-167, 9-11 November.
- [67] Araújo, A. A. (1987). "Filtros Espaciais: Estudo Comparativo e Aplicação em Segmentação e Classificação de Imagens". *Universidade Federal da Paraíba - Tese de Doutorado*, Julho.

- [68] Barros, M. A., and Araujo, A. A. (1990). "Suavização espacial: uma Revisão". *Jornada EPUSP/IEEE em Computação Visual*, pages 81–102, Dezembro.
- [69] Langevin, F. e Facon, J. (1993). "SAIC: Segmentação Automática de Imagens do Cérebro". *VI Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, 1(343-347), Recife-PE, Outubro.
- [70] Castleman, K. R. (1979). "*Digital Image Processing*". Prentice-Hall, New Jersey.
- [71] Cortez, P. C. e Carvalho, J. M. "Look-ahead Boundary Tracking Segmentation of 2D Contours". *XI Congreso Chileno de Ingenieria Electrica, Punta Arenas-Chile*, Nov. 1995.
- [72] Aguiar Neto, B. G. (1987). "*Signalaufbereitung in Digitalen Sprachubertragungssystemen*". PhD thesis, Technischen Universitat Berlin.
- [73] Figueredo, L. H. e Carvalho, P. C. P. (1991). "*Introdução à Geometria Computacional*". IMPA - 18<sup>o</sup> Colóquio Brasileiro de Matemática, Rio de Janeiro-RJ.
- [74] Ramer, U. (1972). "An Iterative Procedure for the Polygonal Approximation of Plane Curves". *COMGP*, (1):224–256, January.
- [75] Pavlidis, T., and Horowitz, S. L. (1974). "Segmentation of Plane Curves". *IEEE - Transactions on Computers*, 23(8):860–870, August.
- [76] Kung, S. H. Y., and Kaswand, T. (1983). "Critical Points on a Perfectly 8- or 6-Connected thin Binary Line". *Pattern Recognition*, (3):297–306.
- [77] Ray, B. K., and Ray, K. S. (1993). "Determination of Optimal Polygon from Digital Curve Using  $L_1$  Norm". *Pattern Recognition*, (4):505–509.
- [78] Cortez, P. C. e Carvalho, J. M. (1992). "Algoritmo para Modelamento Poligonal de Objetos em 2D". *9<sup>o</sup> Congresso Brasileiro de Automática, Vitória-ES*, 1:240–244, Setembro.



- [79] Cortez, P. C. e Carvalho, J. M. (1995). "Modelagem Poligonal de Contornos 2D Usando a Transformada de Hough". 13<sup>o</sup> *Simpósio Brasileiro de Telecomunicações, Águas de Lindóia-SP*, Setembro.
- [80] Zhu, P., and Chirlian, P. M. (1995). "On Critical Point Detection of Digital Shapes". *IEEE - Trans. Pattern Anal. Mach. Intell*, 17(8):737-748, August.
- [81] Cortez, P. C. e Carvalho, J. M. (1994). "Algoritmo de Modelagem Poligonal de Contornos Bidimensionais: uma Avaliação Comparativa Objetiva". 10<sup>o</sup> *Congresso Brasileiro de Automática, Rio de Janeiro-RJ*, 2:1019-1024, Setembro.
- [82] Cortez, P. C. e Carvalho, J. M. (1993). "Algoritmos Seqüenciais para Modelagem Poligonal de Contornos Bidimensionais". *Universidade Federal da Paraíba - Relatório Interno do LAPS - COPELE*, Dezembro.
- [83] Duda, R. O., and Hart, P. E. (1973). "*Pattern Classification and Scene Analysis*". Wiley, New York.
- [84] Junior, R. M. C., and Linhares, O. D. (1990). "Métodos Numéricos de Aproximação Poligonal: uma Análise Comparativa de 3 métodos". *Jornada EPUSP/IEEE em Computação Visual*, pages 61-80, Dezembro.
- [85] Grimson, W. E. L., and Huttenlocher, D. P. (1991). "On the Verification of Hypothesized Matches in Model-Based Recognition". *IEEE - Trans. Pattern Anal. Mach. Intell*, 13(12):1201-1213, December.
- [86] Cortez, P. C. e Carvalho, J. M. (1993). "Avaliação de Três Algoritmos de Modelagem Poligonal de Contornos Bidimensionais". *VI Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, 2(21-24), Recife-PE, Outubro.

ERROR: timeout  
OFFENDING COMMAND: timeout

STACK: