

Reconhecimento de Objetos 3D Através de Atributos de Superfícies de Modelos Poliedrais

Maria de Fátima Santos Farias

Tese de Doutorado submetida à 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.

Área de Concentração: Processamento da Informação

Orientador, Ph. D.
João Marques de Carvalho

Campina Grande, Paraíba, Brasil

©Maria de Fátima Santos Farias, Dezembro de 1997

7056
611.3(213)
F824x

F 824x



F224r Farias, Maria de Fatima Santos
Reconhecimento de objetos 3D através de atributos de
superfícies de modelos poliedrais / Maria de Fatima Santos
Farias. - Campina Grande, 1997.
125 f.

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

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

CDU 621.3(043)

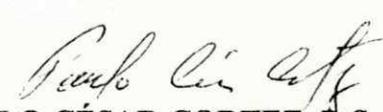
**RECONHECIMENTO DE OBJETOS 3D ATRAVÉS DE ATRIBUTOS DE
SUPERFÍCIES DE MODELOS POLIEDRAIS**

MARIA DE FÁTIMA SANTOS FARIAS

Tese Aprovada em 19.12.1997

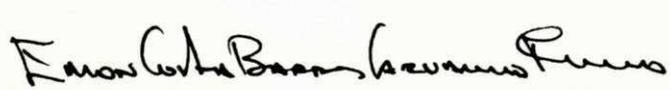

JOÃO MARQUES DE CARVALHO, Ph.D., UFPB
Orientador

ELMAR UWE KURT MELCHER, Dr., UFPB
Componente da Banca


PAULO CÉSAR CORTEZ, D.Sc., UFC
Componente da Banca


ANNA HELENA REALI COSTA RILLO, Dr., EPUSP
Componente da Banca


DÍBIO LEANDRO BORGES, Ph.D., UFG
Componente da Banca


EDSON COSTA DE BARROS CARVALHO FILHO, Ph.D., UFPE
Componente da Banca

CAMPINA GRANDE - PB
Dezembro - 1997

Agradecimentos

Agradeço, primeiramente à Deus, por poder concluir mais uma etapa nesta minha caminhada.

Ao professor João Marques pela valiosa orientação e dedicação, imprescindíveis em todas as fases deste trabalho.

Aos professores Paulo C. Cortez e Elmar Melcher pelas contribuições e sugestões.

Aos funcionários do DEE/UFPB: Rinaldo, Luís Carlos, Paulo Márcio, Ângela e Marcos, pelo espírito de cooperação.

Aos colegas de laboratório: Alberto, Robson, Madeiro, Robson2, Talvanes, Thomás, Yuri, Leonel e especialmente às amigas Luciana, Joseana, Rosângela e Kíssia pelas dicas e bate-papos.

Aos meus familiares, especialmente às minhas irmãs Hilda e Tazinha pelo incentivo e apoio.

Ao DEE/UFMA pela confiança em mim depositada e à CAPES que proporcionou o suporte financeiro deste trabalho.

Ao meu esposo e filhos, Beto, Aline e Bruno, pelo carinho e compreensão.

Enfim, à todos que não foram citados mas que contribuíram, de alguma forma, no desenvolvimento deste trabalho.

Muito Obrigada.

Resumo

Esta tese apresenta um método de reconhecimento de objetos 3D poliedrais através de uma técnica multi-vista utilizando atributos geométricos de superfícies. O método de reconhecimento consiste em casar atributos locais, globais e relacionais das superfícies, possibilitando a identificação tanto do modelo presente na cena quanto das superfícies visíveis nas vistas. Testes com cenas reais e simuladas foram realizados, algumas contendo objetos isolados e outras com mais de um objeto. As taxas de confiabilidade nestas cenas foram muito boas, tendo sido verificados casos com 100% de acerto. Além da identificação, o sistema determina também a posição espacial em que o modelo identificado se encontra na cena.

Abstract

This thesis presents a method of polyhedral 3D object recognition based on a multi-view procedure using geometric attributes from surfaces. The recognition method consists on matching global, local, and relational attributes in order to make possible the identification of both the models present in the scene and the visible surfaces in the views. Experiments of real and simulated scenes were carried on, where some scenes contain isolated objects and another ones contain more than one object. The reliability rate was very good, in some cases 100% of the objects and surfaces were correctly recognized. After the identification, the system establishes the position of the model in the environment of the scene.

Índice

1	Introdução Geral	1
1.1	Aspectos Gerais de Visão Computacional	1
1.2	Sistemas de Visão e Reconhecimento de Objetos 3D	2
1.3	Objetivos da Tese e Organização do Texto	6
2	Reconhecimento 3D Através de Modelos - Revisão Bibliográfica	9
2.1	Introdução	9
2.2	Construção da Biblioteca de Modelos	12
2.2.1	Modelagem por sistemas CAD	13
2.2.2	Representação de bordas de superfície	15
2.2.3	Representação de volume	18
2.2.4	Representação de Imagem Gaussiana Estendida	20
2.2.5	Representação de Varredura	21
2.2.6	Outras Representações	22
2.2.7	Conclusões	23
2.3	Extração de Características da Cena	24
2.3.1	Aquisição de imagens do objeto	24
2.3.2	Descrição do objeto na cena	26
2.3.3	Tipos e representações de características	27
2.4	Casamento de Características - Reconhecimento de Objetos	28
2.5	Análise de Alguns Métodos	30
2.6	Conclusões	39

3	Modelagem de Objetos 3D	42
3.1	Introdução	42
3.2	Construção das Formas 3D	43
3.3	Atributos Geométricos	44
3.4	Conclusões	46
4	Aquisição e Extração de Dados da Cena	48
4.1	Introdução	48
4.2	Transformações Geométricas da Câmera	51
4.3	Pré-processamento e Segmentação das Imagens	53
4.4	Técnica multi-vista	55
4.4.1	Análise das Vistas	56
4.4.2	Cálculos dos atributos de superfícies	61
4.4.3	Eliminação das superfícies duplicadas	63
5	Método de Identificação e Localização	65
5.1	Introdução	65
5.2	Características dos Objetos	66
5.3	Identificação	67
5.3.1	Geração de hipóteses	67
5.3.2	Verificação de hipóteses	69
5.4	Localização	71
6	Testes Efetuados e Resultados Obtidos	76
6.1	Introdução	76
6.2	Cenas Simuladas	77
6.2.1	Cenas com objetos isolados	78
6.2.2	Cenas com mais de um objeto	79
6.3	Cenas Reais	91
6.4	Pré-Processamento das Cenas Reais	92
7	Conclusões e Contribuições	108

A Passos do Algoritmo de Identificação

121

Lista de Tabelas

6.1	Superfícies 3D (cruzamento 2D).	79
6.2	Geração de hipóteses.	82
6.3	Verificação - Cena 1.	83
6.4	Verificação - Cena 2.	83
6.5	Verificação - Cena 3.	84
6.6	Verificação - Cena 4.	84
6.7	Verificação - Cena 5.	84
6.8	Localização e tempo de processamento.	84
6.9	Superfícies 3D (cruzamento 2D).	86
6.10	Geração de hipóteses.	88
6.11	Verificação - Cena 6.	89
6.12	Verificação - Cena 7.	90
6.13	Verificação - Cena 8.	90
6.14	Verificação - Cena 9.	90
6.15	Localização e tempo de processamento.	90
6.16	Superfícies 3D (cruzamento 2D).	92
6.17	Reconhecimento das imagens reais.	94
6.18	Cenas Simuladas.	95
6.19	Cenas Reais.	95

Lista de Figuras

1.1	Diagrama em blocos do sistema de reconhecimento.	8
2.1	Representação de bordas de superfícies.	16
2.2	Árvore de CSG	20
2.3	Estrutura para técnica de triangulação.	25
2.4	Casamento de características 1.	29
3.1	Estrutura básica de um modelo.	43
3.2	Vista 3D dos modelos.	45
4.1	Estrutura de aquisição das imagens.	49
4.2	Diagrama em blocos de análise da cena.	50
4.3	Relação entre comprimento real e projetado na negativa	53
4.4	Análise Multi-vista.	56
4.5	Diagrama da análise da vista superior.	57
4.6	Diagrama da análise da vista lateral.	59
4.7	Diagrama da análise da vista frontal.	60
5.1	Passos para a localização.	72
5.2	Ângulo e vetor de rotação.	74
5.3	Ângulo entre as arestas (a) a_m adiantada (b) a_m atrasada.	75
6.1	Cena 1.	80
6.2	Cena 2.	80
6.3	Cena 3.	81

6.4	Cena 4.	81
6.5	Cena 5.	83
6.6	Cena 6.	86
6.7	Cena 7.	87
6.8	Cena 8.	87
6.9	Cena 9.	89
6.10	Cena 10 - superior, lateral e frontal.	92
6.11	Cena 11 - superior, lateral e frontal.	93
6.12	Cena 12 - superior, lateral e frontal.	93
6.13	Cena 13 - superior e lateral.	94
6.14	(a) Imagem superior (b) gradiente e (c) contorno final da cena 10.	96
6.15	(a) Imagem lateral (b) gradiente e (c) contorno final da cena 10.	97
6.16	(a) Imagem frontal (b) gradiente e (c) contorno final da cena 10.	98
6.17	(a) Imagem superior (b) gradiente e (c) contorno final da cena 11.	99
6.18	(a) Imagem lateral (b) gradiente e (c) contorno final da cena 11.	100
6.19	(a) Imagem frontal (b) gradiente e (c) contorno final da cena 11.	101
6.20	(a) Imagem superior (b) gradiente e (c) contorno final da cena 12.	102
6.21	(a) Imagem lateral (b) gradiente e (c) contorno final da cena 12.	103
6.22	(a) Imagem frontal (b) gradiente e (c) contorno final da cena 12.	104
6.23	(a) Imagem superior (b) gradiente e (c) contorno final da cena 13.	105
6.24	(a) Imagem lateral (b) gradiente e (c) contorno final da cena 13.	106
6.25	(a) Imagem frontal (b) gradiente e (c) contorno final da cena 13.	107

Capítulo 1

Introdução Geral

1.1 Aspectos Gerais de Visão Computacional

A capacidade de discernir objetos, analisar, entender e inspecionar o que está sendo visto, averiguar seus movimentos e navegar no espaço tri-dimensional através da visão são tarefas utilizadas com bastante facilidades pelo ser humano. Entretanto, o sentido da visão do ser humano é resultado de um processo extremamente complexo e ainda não compreendido totalmente pela Ciência.

Com o desenvolvimento tecnológico da sociedade, o homem tem procurado investigar cada vez mais sobre processos que viabilizam a automatização de certas tarefas. Progressos importantes já foram obtidos na área de Inteligência Artificial, tornando possível construir máquinas com capacidade cada vez maior para se comunicar com o meio externo e processar inteligentemente as informações recebidas por sensores.

Fazer máquinas que podem “ver” tem sido um tema de grande interesse entre pesquisadores nestas últimas três décadas. Muito embora ainda não se consiga dotar uma máquina com a capacidade total do sistema visual humano, muitos progressos foram e estão sendo alcançados ao construir máquinas que realizam tarefas visuais específicas.

Sem dúvida, uma máquina com capacidade de visão inteligente pode ser aplicada em diversas áreas, tais como atividades militares, automação industrial, pesquisas espaciais e equipamentos médicos. Visão Computacional, ou Visão Artificial, tem como principal

objetivo extrair informações de uma cena através da análise das imagens da mesma e ser capaz de “raciocinar” com base nas mesmas informações, sem a ajuda de um assistente humano [Mar92].

Em Visão Computacional, a maioria das técnicas utilizam imagens formadas por radiação visível, de uma cena iluminada por luz natural ou artificial, através de um sensor que incorpora um sistema ótico, o qual transforma a imagem da cena numa imagem plana, perpendicular ao campo de visão do sensor. Outras técnicas utilizam imagens desenvolvidas por Raios-X, luz infra-vermelho, luz ultra-violeta, bem como imagens térmicas, de sonar e de radar [Mar92, Ros88].

Em geral, a aplicação de Visão Computacional envolve técnicas de Processamento Digital de Imagens, Reconhecimento de Padrões e Inteligência Artificial [Cor96].

Técnicas de Processamento Digital de Imagens consistem da manipulação de dados contidos nas imagens para melhorar ou intensificar certas propriedades tornando-as mais nítidas para os seres humanos e mais adequadas para seu armazenamento e transmissão no computador [Cor96].

Reconhecimento de Padrões classifica os dados resultantes da descrição da imagem da cena de acordo com certos tipos de critérios, regras, formas ou medidas pré-estabelecidas.

A Inteligência Artificial procura dotar os computadores de procedimentos de raciocínio semelhantes aos seres humanos na solução de problemas análogos [Cor96].

Uma outra área relacionada com a visão computacional, que é importante mencionar, é a computação gráfica, que produz imagem artificial pelo computador e pode ser usada para criar modelos úteis nos sistemas de Visão.

1.2 Sistemas de Visão e Reconhecimento de Objetos 3D

Uma *imagem monocromática*, ou *imagem de intensidade* ou simplesmente *imagem*, é definida como sendo uma função da intensidade de luz, uma vez que as imagens são captadas por sensores através da energia radiante do ambiente. Matematicamente, é

uma função bi-dimensional e contínua, ou seja, pode ser descrita por $f(x, y)$, onde x e y denotam as coordenadas espaciais do sistema de referência da imagem captada e o valor de f em qualquer ponto (x, y) é proporcional ao brilho (ou nível de cinza) da imagem naquele ponto. A *imagem digital* é discretizada (amostrada) nas coordenadas espaciais e no brilho. É considerada uma matriz de pontos (linhas \times colunas) cujos elementos correspondem ao nível de cinza do ponto correspondente na imagem. Os elementos desta matriz são denominados *elementos de imagem*, *elementos de quadro* ou *pixels* (de “*picture elements*”).

Para muitas aplicações, o uso de representações bi-dimensionais para objetos reais não é suficiente, requerendo o desenvolvimento de métodos mais precisos que descrevam o objeto em três dimensões.

Uma imagem tri-dimensional (3D) tem arranjo similar à imagem bi-dimensional (2D), sendo que contém informações sobre a profundidade do objeto, ou seja, a distância de um ponto do objeto ao plano da imagem. Tal distância pode ser obtida diretamente ou indiretamente da imagem. Na medida direta, o nível de cinza do pixel na imagem indica a distância do objeto ao sensor nas coordenadas correspondentes. Na medida indireta, a distância é calculada com base nas propriedades geométricas como por exemplo, a orientação das arestas ou superfícies encontradas [Nit88].

Atualmente, a principal área de aplicação para a imagem digital 3D é a área de equipamentos médicos. Podem ser citados como exemplo, os *scanners* de tomografia computadorizada e os de ressonância magnética os quais são amplamente utilizados para produzir imagens digitais 3D de órgãos internos do corpo humano [Udu90].

Srihari [Sri81] apresenta um estudo sobre os principais tipos de representação computacional de imagem digital 3D, onde considera que qualquer representação da imagem 3D requer uma amostragem do volume, para extrair um conjunto discreto de valores do brilho da imagem.

O menor elemento de volume da imagem digital 3D é denominado *voxel* (de “*volume element*”). Uma imagem digital 3D é um mapeamento (função) que associa cada *voxel* a um valor do tipo real, inteiro ou binário. Quando a imagem digital 3D é obtida por amostragem de uma função contínua, a amostragem é seguida pelo processo de quantização, que torna possível a representação da imagem com um número finito de

bits. Quando a imagem 3D consiste de um fólio de fatias (“*slices*”) de imagens 2D, a imagem correspondente ao espaço entre as fatias é geralmente obtida por processos de *interpolação* de imagens [Sri81].

A *imagem digital binária 3D* é definida através de uma função característica $f(V)$, onde V representa um *voxel* descrito em coordenadas espaciais (x, y, z) . O domínio da função f é o conjunto de todos os *voxels* da imagem e o contra-domínio é o conjunto $\{0, 1\}$. Em geral, na imagem digital binária 3D, o conjunto de *voxels* $S = \{V | f(V) = 1\}$ refere-se ao objeto da cena e o conjunto complementar, $\bar{S} = \{V | f(V) = 0\}$, refere-se ao plano de fundo da cena que contém o objeto.

A aquisição da imagem digital 3D de objetos e sua posterior representação, para uso em sistemas de Visão baseia-se no princípio de “reconhecimento por análise de contornos 3D” (“*recognition by 3D shape analysis*”). Tais métodos envolvem, freqüentemente, alguns problemas práticos tais como dificuldades no sistema de calibração do sensor 3D, complexidade na extração de atributos e no tratamento digital dos dados de volume [Tsu90, Osh83].

Se o objetivo é apenas classificar o objeto, é desejável evitar a análise de contornos 3D diretamente. A estratégia utilizada em inúmeras investigações é a do “reconhecimento 3D através da análise de contorno 2D”. A vantagem desta abordagem é que inúmeras técnicas de análise 2D estão disponíveis, existindo uma vasta literatura a respeito, além de possibilitar uma considerável economia nos cálculos requeridos. Na maioria dos casos, duas ou mais câmeras são utilizadas para produzir descrição 3D de uma cena através da combinação de características das imagens 2D de cada câmera. Esta técnica é conhecida como visão estéreo (“*stereo vision*”) ou multi-vista (“*multi-views*”) [Mar92, Chi92b].

Em Visão estéreo, a maior dificuldade está em estabelecer um casamento confiável e rápido dos pontos de correspondência em imagens da cena captadas de pontos de observação diferentes. Neste caso, é importante conhecer a geometria da estrutura de aquisição das imagens que irão determinar a correspondência de pontos entre elas.

Nitzan [Nit88] apresenta um estudo sobre as principais técnicas de aquisição de imagens, extração e pré-processamento de dados 3D para aplicações em Visão Computacional 3D. Uma das técnicas mais utilizadas é a da triangulação estéreo (“*triangu-*

lution stereo”), que utiliza duas câmeras direcionadas à cena, com ou sem projetor de luz [Liu90b, Sha96, Chu94]. Também são muito utilizadas as técnicas multi-vista que fazem uso de mesa giratória para captar várias imagens de um objeto com uma única câmera [Bha84].

A abordagem mais utilizada em sistemas de reconhecimento de objetos é a de sistema de visão baseado em modelos, em que características extraídas de objetos numa cena são casadas com características de modelos previamente armazenadas. No reconhecimento de objetos 3D as características do objeto na cena e dos modelos de objetos são formuladas em três dimensões.

Se um objeto (sólido), nunca visto antes, for apresentado a um ser humano, ele é capaz de extrair e organizar informações (características) sobre o mesmo, procurando observá-lo sob diferentes pontos de vista, tentando conhecer sua geometria, textura, funcionalidade, etc [Bes85].

Uma vez familiarizado com os objetos já apresentados, o ser humano pode normalmente identificá-los a partir de pontos arbitrários de observação, bem como descrever sua localização e orientação no espaço, sem necessidade de investigação adicional. Este procedimento pode ser denominado, de modo amplo, como *processo de reconhecimento de objetos* [Bes85].

Analogamente, a definição de reconhecimento de objetos por computador engloba basicamente a identificação e a localização dos objetos que se apresentam numa determinada cena, procurando definir quais são, e em que posição espacial se encontram os mesmos. Neste caso, o sistema deve ter um conhecimento *a priori* do objeto a ser identificado, cuja descrição em geral é feita através de atributos geométricos. Este conhecimento *a priori* é conhecido como o modelo do objeto, e o sistema que faz uso de modelos é denominado sistema de visão baseado em modelos.

A abordagem que se baseia em modelos, em geral, é caracterizada pelo fato do reconhecimento ser obtido através do casamento entre as características extraídas do objeto na cena, com características encontradas nos modelos armazenados.

Chin-Dyer [Chi86], Besl-Jain [Bes85], Arman-Aggarwal [Arm93] e Requicha [Req80] apresentam um estudo abrangente sobre as diversas técnicas para sistemas de Visão Baseado em Modelos, identificando as principais etapas incorporadas em tais técnicas.

1.3 Objetivos da Tese e Organização do Texto

Trabalhos realizados no Laboratório de Automação e Processamento de Sinais (LAPS) da UFPB nestes últimos anos [Cor96, Vil97] mostram as vantagens em se trabalhar com modelos poligonais para efetuar reconhecimento e localização de formas bi-dimensionais de objetos. Os modelos poligonais são fáceis de serem representados e reconhecidos, além de possuírem baixo custo computacional com relação a outros métodos, qualidades que os tornam adequados para aplicações em tempo real. Dissertações de Mestrado visando a implementação dos algoritmos de modelagem poligonal em *hardware* para processamento de imagens em tempo real já estão em fase de conclusão [Vil97, Mel97, Mel96b] no LAPS.

Para muitas aplicações, principalmente na automação industrial onde se utiliza visão para robôs em linha de montagem e inspeção visual, o uso de representações em 2D para objetos reais não é suficiente, requerendo o desenvolvimento de métodos mais precisos para modelagem e reconhecimento de objetos através de formas em três dimensões. Visando a integração de um método (*software*) de reconhecimento de objetos 3D com os projetos desenvolvidos em 2D no LAPS, surgiu a proposta desta tese. A integração do *software-hardware* permitirá o reconhecimento de objetos 3D em tempo real, utilizando modelos poliedrais.

O objetivo desta tese é apresentar um método de reconhecimento de objetos em cenas tri-dimensionais baseado em técnica multi-vista, cuja identificação ocorre através de casamento de atributos geométricos extraídos de imagens bi-dimensionais dos objetos na cena. Estes atributos são casados com os encontrados em modelos tri-dimensionais dos objetos, previamente construídos e armazenados. O método consiste de três etapas:

1. Construção de modelos,
2. Extração de características do objeto na cena,
3. Casamento de características (identificação e localização).

O diagrama em blocos é mostrado na figura 1.1. A biblioteca de modelos consiste de modelos tri-dimensionais e um conjunto de atributos geométricos de superfícies correspondente a cada um dos modelos. Na etapa de extração de características são captadas

três imagens do objeto as quais passam por processos de digitalização e extração de algumas características bi-dimensionais. Tais características são reunidas e um conjunto de atributos geométricos 3D correspondentes às superfícies visíveis do objeto é elaborado.

Com as informações provenientes do conjunto de atributos da cena, o bloco de identificação faz uma busca na biblioteca de modelos, determinando qual o modelo que contém aqueles atributos, através do casamento de características. Nesta etapa, o sistema define não apenas o modelo mas também as superfícies do mesmo correspondentes às superfícies visíveis na cena. Com as informações do modelo identificado e da cena, o sistema determina a localização espacial do modelo no ambiente da cena e produz o relatório de saída.

O casamento de características ocorre em duas fases: a geração de hipóteses e a verificação das hipóteses geradas. As hipóteses são geradas a partir de atributos geométricos locais e globais das superfícies e são verificadas (confirmadas ou não), a partir de atributos relacionais.

O Capítulo 2 faz uma revisão bibliográfica a respeito das técnicas mais utilizadas nos sistemas de reconhecimento de objetos 3D baseados em modelos. O Capítulo 3 apresenta o método utilizado neste trabalho para a modelagem dos objetos. O Capítulo 4 apresenta a técnica utilizada para a aquisição e extração dos dados das imagens da cena. O Capítulo 5 apresenta o método de reconhecimento de objetos desenvolvido. O Capítulo 6 analisa os resultados conseguidos com os testes efetuados com imagens reais e sintéticas. Por fim, o Capítulo 7 apresenta as características do sistema, as conclusões do trabalho desenvolvido, as contribuições originais e sugestões para trabalhos futuros.

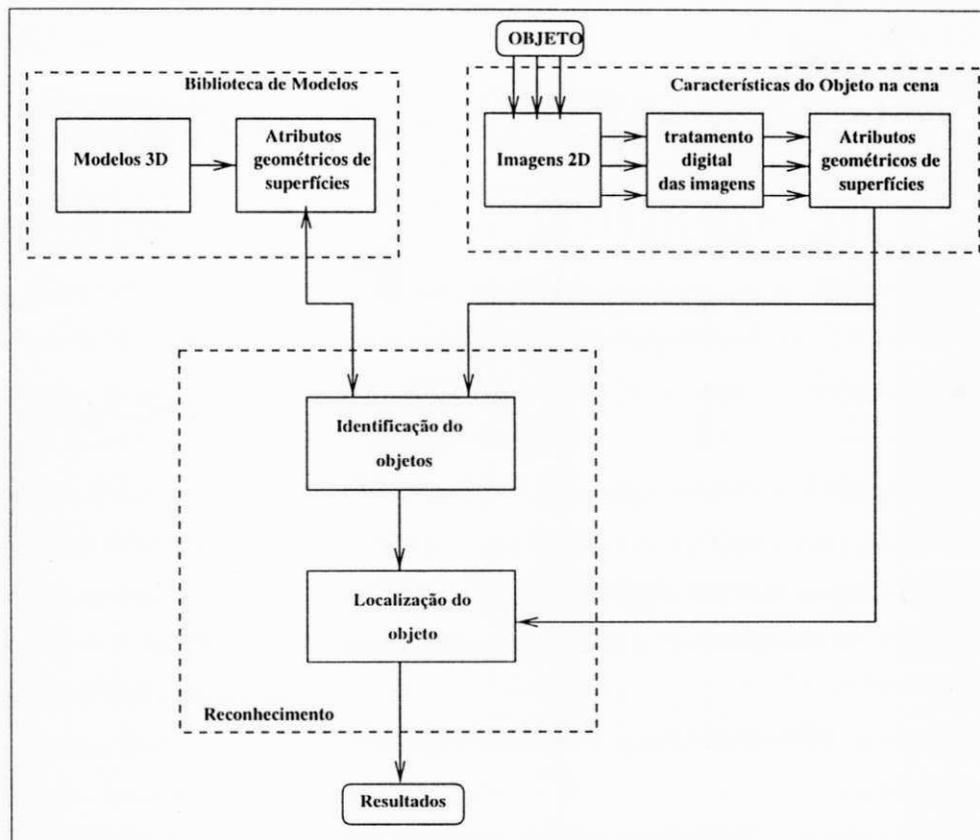


Figura 1.1: Diagrama em blocos do sistema de reconhecimento.

Capítulo 2

Reconhecimento 3D Através de Modelos - Revisão Bibliográfica

2.1 Introdução

O reconhecimento de objetos 3D baseado em modelos tem sido estudado por vários pesquisadores, e é considerada uma das mais importantes linhas de pesquisa na área de Visão Computacional. O reconhecimento envolve basicamente a construção de modelos de objetos em 3D, cujas descrições espaciais estão relacionadas a um sistema de coordenadas tri-dimensionais.

Os métodos de reconhecimento de objetos 3D, a partir de análise de contornos 2D, podem ser classificados em duas categorias:

1. Métodos de reconhecimento através de múltiplas vistas,
2. Métodos de reconhecimento através de uma vista simples.

Os métodos de múltiplas vistas [Aya94, Chu94, Bha84, Maj89, Liu90b, Xie94, Wen93] procuram casar contornos de bordas das projeções 2D obtidas de várias posições de visualizações para reconstruir a estrutura do objeto (modelo do objeto) em 3D. São métodos que possuem boa tolerância às sombras que aparecem nas imagens quando se utiliza a projeção apenas de uma vista. A dificuldade destes métodos é encontrar uma boa correspondência dos pontos do objeto entre várias imagens.

Nos métodos de vista simples [Bol86, Nev77, Fly91, Osh83, Bro79, Bro83, Low87], a maior dificuldade está na identificação e localização do objeto através de um número reduzido de informações 3D contidas numa única vista, o que aumenta a complexidade de cálculos.

A estrutura física para aquisição das imagens de tais categorias requer técnicas geométricas confiáveis envolvendo câmeras (sensores), luz e objeto. As estruturas mais utilizadas são as de triangulação e retorno de feixe, descritas mais adiante.

Para a extração de características das imagens são necessárias as etapas de pré-processamento e segmentação. Para determinar o método de segmentação deve-se, primeiro, indicar quais características serão extraídas. Em geral, estas características são representadas por características de volume, de superfícies ou de cilindro generalizado [Chi86, Bes85, Arm93].

A biblioteca de modelos dos objetos pode ser construída de dois modos: através de características extraídas do próprio objeto (mesma estrutura de aquisição da imagem da cena) ou através de sistemas de desenvolvimento de modelos auxiliados por computador, como o CAD (de "*Computer Aided Design*"), por exemplo. Alguns trabalhos tais como os de Bolles-Horaud [Bol86], Park et al. [Par93], Majumdar et al. [Maj89], Zhang et al. [Zha93], Flynn et al. [Fly91], Hoffman et al. [Hof89] e Gunnarsson-Printz [Gun87], utilizam sistemas CAD no método de reconhecimento de objetos 3D, quer seja na elaboração de vistas, quer seja na extração de atributos.

As correspondências entre os elementos dos dados da cena com dados encontrados nos modelos do objeto caracterizam a etapa de identificação (ou reconhecimento) do objeto na cena. Um conjunto de dados extraídos da cena podem gerar inúmeras correspondências com os dados dos modelos. Nesta etapa, é importante definir critérios para restringir o espaço de busca por correspondências consistentes.

Grimson [Gri90] mostra como restrições geométricas de dados de contorno, associadas com dados de subconjuntos de contorno, evitam o problema de grandes quantidades de correspondências durante o processo de casamento. Por exemplo, se apenas três arestas forem extraídas da cena, dependendo do modelo, isto pode gerar centenas de correspondências com arestas de diversos modelos. Mas se, adicionalmente, for conhecido o ângulo entre a primeira e a segunda aresta, certamente esta informação restringe

o espaço de pesquisa a um número menor de correspondências. Se for conhecido ainda o ângulo entre as demais arestas, o espaço de pesquisa é bem menor, tornando o sistema bem mais econômico, computacionalmente.

Relacionando-se o modo de construção dos modelos, com o modo de aquisição da imagem da cena e o tipo de representação utilizado para descrição de objetos e modelos vários trabalhos de reconhecimento ou reconstrução de modelos têm sido publicados. Estas diferenças somadas às estratégias utilizadas no algoritmo de reconhecimento personalizam cada sistema, de modo que se tornam específicos em sua atuação.

Besl-Jain [Bes85] descrevem formalmente uma estrutura para reconhecimento de objetos de propósito geral baseada em modelos. Esta estrutura envolve os seguintes passos: formação da imagem, descrição simbólica, modelagem, interpretação da imagem (ou casamento) e síntese. Em resumo, o processo de formação da imagem cria a base de dados do objeto baseado nos princípios físicos da aquisição da imagem e no tipo de sensor que está sendo utilizado. A descrição define os atributos relevantes do objeto. A modelagem apresenta os modelos dos objetos do mundo real através de descrições geométricas. A etapa de casamento envolve a geração de hipóteses, através do casamento de atributos das descrições de dados do objeto do mundo real com atributos dos modelos armazenados e a verificação das hipóteses, através do casamento de outros atributos mais restritos, tentando identificar corretamente o modelo do objeto. Na síntese, o sistema reproduz os dados do modelo identificado no ambiente do sistema de referência do objeto, além de permitir a checagem automática do modelo identificado comparando-o ao objeto para verificar a necessidade de realimentação de dados para novas investigações.

Arman-Aggarwal [Arm93] descrevem um sistema de Visão baseado em modelos composto pelas seguintes etapas: coleta de dados, representação de dados, construção de modelos dos objetos e casamento de descrições. Na coleta de dados definem-se o tipo de sensor e a estrutura utilizada para adquirir a imagem do objeto. Na representação de dados define-se o tipo adequado dos dados característicos dos modelos para o casamento e armazenamento. Na construção de modelos do sistema define-se a abordagem que deve ser utilizada para construir os modelos do sistema. O processo de casamento estabelece, num primeiro passo, correspondências entre o conjunto de

descrições do modelo do objeto e os conjuntos de descrições dos modelos armazenados. No segundo passo, utilizam-se correspondências estabelecidas para derivar uma transformação geométrica para que o modelo seja reconstruído na mesma orientação do objeto na cena.

Chin-Dyer [Chi86] apresentam um estudo de vários sistemas que fazem reconhecimento de peças industriais agrupando as representações em categorias, de acordo com sua dimensionalidade no espaço (2D, $2\frac{1}{2}$ D e 3D). Associado à cada categoria, Chin-Dyer propõem um sistema mais genérico para reconhecimento de objetos baseado em apenas três etapas: construção da biblioteca de modelos do objeto, extração de características do objeto na cena e casamento de características (identificação e localização). Em resumo, o reconhecimento de objetos proposto por Chin-Dyer, ocorre da seguinte maneira: são extraídos atributos geométricos das imagens dos objetos na cena. Tais atributos são indexados à biblioteca de modelos, onde é feita uma seleção (busca) dos modelos possuidores daqueles atributos, de acordo com certa função de erro. Deste modo, são geradas hipóteses de correspondências de atributos. Estas hipóteses são refinadas e confirmadas, ou recusadas, implicando no reconhecimento do objeto e sua respectiva localização, ou então na aquisição de uma nova imagem do objeto na cena para novas buscas.

As propostas de sistemas de reconhecimento de Besl-Jain, Arman-Aggarwal e Chin-Dyer são muito parecidas. Em se tratando de uma descrição genérica, adota-se a proposta de Chin-Dyer, como base neste trabalho. Nas seções seguintes serão abordadas as principais técnicas utilizadas em cada etapa do sistema proposto por Chin-Dyer.

2.2 Construção da Biblioteca de Modelos

Em aplicações industriais, frequentemente é necessário reconhecer objetos em três dimensões (3D). O desempenho de tais sistemas depende essencialmente da escolha apropriada do método que descreve completamente o objeto em 3D. Ou seja, depende do método que produz o modelo mais próximo possível do objeto real [Mar92].

A representação empregada para descrever objetos em sistemas de reconhecimento, define a estratégia do método utilizado para o casamento de atributos entre os modelos

armazenados e o modelo do objeto na cena. Portanto, uma vez definidos os modelos, podem ser desenvolvidos vários sistemas de visão capazes de reconhecê-los.

Na literatura encontram-se duas abordagens para a construção de modelos 3D. Uma abordagem utiliza o próprio objeto para gerar o modelo, a outra utiliza os recursos dos sistemas CAD (de “*Computer-Aided Design/Drafting*”) para elaborar os modelos. Na primeira abordagem, são extraídas com antecedência características das imagens obtidas de vários pontos de vista do objeto, integrando-as de forma coerente para que possam apresentar informações do objeto sob todos os ângulos de vista.

Na segunda abordagem, os sistemas CAD permitem que um usuário construa os modelos dos objetos no computador. Além da geração e visualização em 3D no computador, o CAD pode extrair dos modelos as características geométricas comumente utilizadas.

Como os sistemas que utilizam as técnicas de construção de modelos a partir do próprio objeto empregam as mesmas técnicas para a extração de características, esta abordagem será descrita na Seção 2.3. Na presente seção, será descrita apenas a modelagem por sistemas CAD.

2.2.1 Modelagem por sistemas CAD

Com o desenvolvimento da robótica e da automação industrial o interesse por esquemas de modelagem automática tem aumentado, principalmente em aplicações para as engenharias mecânica, civil e elétrica. Os avanços nas tecnologias CAD (de *Computer-Aided Design/Drafting*) e CAM (de *Computer-Aided Manufacturing*) têm sido essenciais na melhoria da qualidade e da produtividade industrial. A chave de tais sucessos é a integração dos modelos CAD no desenvolvimento de sistemas de robôs inteligentes para tarefas de reconhecimento, inspeção visual, manipulação e montagem de peças industriais.

Para a construção da biblioteca de modelos de objetos nos sistemas de reconhecimento podem ser utilizados sistemas CAD, tanto para projetar a geometria do objeto (sintetizar imagens), quanto para derivar características (extrair atributos) essenciais nos procedimentos de reconhecimento.

Os sistemas CAD oferecem facilidades gráficas para construir modelos em 2D e 3D, operando com grande precisão e escalas apropriadas para plotagem e visualização. O CAD permite manipular os modelos construídos de várias formas, como rotacionar e transladar para ter acesso às vistas (projeções) do objeto, transformar formas 2D em sólidos por meio de extrusão e revolução, estender curvas, combinar características simples em sólidos complexos com as operações booleanas de união, subtração e interseção, adicionar características a sólidos através de chanfros e arredondamentos, além de definir fontes de iluminação, adicionar ou retirar textura e cores.

A construção de modelos por CAD possui algumas vantagens sobre os métodos que constroem modelos a partir de imagens do objeto, são elas: elaboração de modelos com grandes aproximações e livres dos erros que podem ocorrer na aquisição da imagem, geração de partes não visíveis numa vista, facilidade para manusear uma ampla classe de objetos, facilidade para gerar bases de dados geométricos e existência de bibliotecas padronizadas de modelos adequadas para diversas áreas da engenharia. Além disto, o CAD pode simular situações (cenas) com os modelos antes mesmo da elaboração da estrutura física de aquisição das imagens do objeto, para testes e estudos do sistema.

Os sistemas CAD estão divididos em três grupos. No primeiro grupo, destacam-se os sistemas que são usados em *mainframes*. Estes são computacionalmente poderosos e com variedades de manipulação de modelos. É o caso do CATIA, do I-DEAS, do CADAM e outros. No segundo grupo estão aqueles utilizados nos PC's, com capacidade menor que o primeiro grupo mas munidos das tarefas essenciais ao projeto de modelos. É o caso do AutoCAD, do CAD KEY, do Personal Designer e outros. No terceiro grupo estão aqueles que operam especificamente em estações de trabalho como o ANVIL 500, o CADRA-III, o CIM CAD 3D e outros [Arm93].

Os diversos sistemas CAD diferem entre si pelo poder de manipular as características geométricas dos modelos, mostrá-las sob pontos de vistas supondo vários ângulos de iluminação, operações de edição, velocidade de operação, custo e disponibilidade de programas auxiliares. Alguns CAD também utilizam formatos padronizados de armazenamento de dados para permitir a troca de modelos entre sistemas diferentes.

Os primeiros modelos de sólidos 3D de sistemas CAD, provavelmente, utilizaram a representação de armação de linhas (*wire-frames*), onde o objeto é representado

por um conjunto de arestas (linhas) interligadas. Tal representação, por si só, não era suficiente para definir as faces do objeto e portanto poderia confundir o usuário. Na literatura encontram-se várias formas de representação de sólidos 3D para sistemas CAD, das quais as mais importantes se enquadram nas seguintes classes:

- representação de bordas de superfície,
- representação de volume,
- representação de imagem Gaussiana estendida,
- representação de varredura.

2.2.2 Representação de bordas de superfície

A representação de bordas de superfície (de “*Boundary*”), ou B-rep, é a mais utilizada nos sistemas CAD e em computação gráfica. Nesta representação, um sólido pode ser descrito por três formas: um conjunto de superfícies (\mathcal{S}), um conjunto de curvas espaciais (\mathcal{T}) e grafo relacional (\mathcal{G}) descrevendo as conectividades entre as superfícies [Fly91].

Com relação a representação B-rep, Besl-Jain [Bes85] mostram que uma superfície genérica 3D é descrita na seguinte forma:

$$\mathcal{S} = \{(x, y, z) : f(x, y, z) = 0\}, \quad (2.1)$$

a qual é conhecida como a representação implícita de uma superfície. Se o vetor gradiente ∇f existe e é contínuo não nulo para cada ponto (x, y, z) , então \mathcal{S} é uma superfície lisa (“*smooth*”). Dizer que $f(x, y, z) = 0$, significa que o ponto (x, y, z) pertence à superfície \mathcal{S} .

As superfícies planas são representadas por uma função com quatro coeficientes:

$$f_p(x, y, z) = Ax + By + Cz + D, \quad (2.2)$$

onde (A, B, C) são constantes que especificam a direção da normal da superfície, enquanto D representa a distância do plano até a origem do sistema de coordenadas, desde

que A, B e C estejam normalizados. Esta forma é a mais simples da representação B-rep e o conjunto destas superfícies resulta num objeto poliedral ou aproximadamente poliedral. A Figura 2.1 ilustra um objeto poliedral com suas superfícies constituintes. Neste caso, os elementos de \mathcal{S} são as faces do sólido, os de \mathcal{T} são as arestas de cada face e \mathcal{G} mostrará a estrutura de grafos relativa às propriedades geométricas que compõem o objeto, ou seja: as faces, as arestas e os vértices[Bes85].

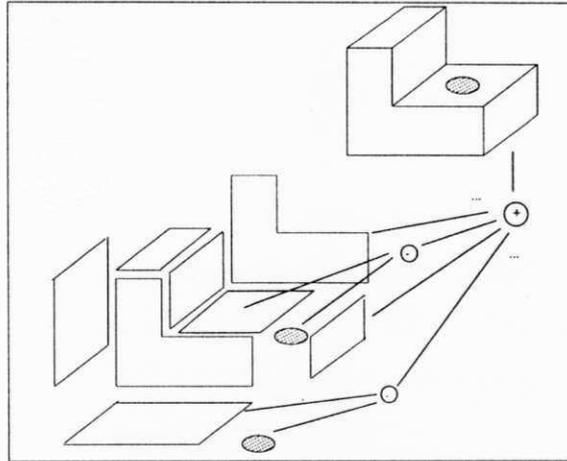


Figura 2.1: Representação de bordas de superfícies.

Para descrições de superfícies curvas, Besl-Jain [Bes85] utilizam a representação explícita de superfície com 10 coeficientes, conhecida como superfície quádrlica ou quadrática (de “*quadric surface*”), definida por:

$$f_q(x, y, z) = Ax^2 + By^2 + Cz^2 + Jxy + Hyz + Izx + Ux + Vy + Wz + D, \quad (2.3)$$

onde (A, B, C) descrevem o contorno da superfície enquanto (J, H, I, U, V, W) são parâmetros necessários para localização e orientação da superfície do objeto no espaço.

Ajustando-se alguns coeficientes, as superfícies quádrlicas podem ser classificadas em vários tipos. As mais conhecidas são [Bes85]:

- elipsóide (se $A > 0, B > 0, C > 0, D = -1$)
- parabolóide elíptico (se $A > 0, B > 0, W = -1$)

- parabolóide hiperbólico (se $A > 0, B < 0, W = -1$)
- hiperbolóide de 1 lâmina (se $A > 0, B > 0, C > 0, D = -1$)
- hiperbolóide de 2 lâminas (se $A > 0, B < 0, C < 0, D = -1$)
- cone quádrico (se $A > 0, B > 0, C < 0$).

Para descrição de superfícies de forma livre, a forma implícita não é utilizada, a menos que as superfícies sejam decompostas numa coleção de remendos (“*surface patches*”) homogêneos. Uma abordagem alternativa padronizada é utilizar a representação paramétrica, da forma:

$$\mathcal{S} = \{(x, y, z) : x = h(u, v), y = g(u, v), z = f(u, v) | (u, v) \in \text{Dominio } D \subseteq \mathbb{R}^2\} \quad (2.4)$$

onde f, g, h são funções escalares de duas variáveis (u, v) . Um objeto representado por níveis de cinza em imagens de intensidade (2D) ou em arranjo de imagens (3D), pode ser descrito usando a equação 2.4, da seguinte forma:

$$\mathcal{S} = \{(x, y, z) : x = u, y = v, z = f(u, v) \in D \subseteq \mathbb{R}^2\}, \quad (2.5)$$

onde apenas o valor de z é função de um ponto no plano X - Y [Bes85].

Muitos tipos de equações paramétricas de superfícies são utilizados em sistemas CAD. As diferenças entre elas ocorrem devido às diferentes definições para as funções $f(u, v), g(u, v), h(u, v)$. Besl [Bes85] cita, dentre outras, algumas destas superfícies: superfície de remendos bicúbica de Ferguson, superfície de remendos bicúbica de Bezier, superfície *spline* paramétrica, superfície *B-spline* racional.

Dentre as citadas, as superfícies *B-spline* racional são mais abrangentes e podem representar tanto as superfícies de forma livre quanto as de primitivas quádricas e as de aproximação poliedral, através de uma forma matemática proposta por Tiller [Til83]. Além disto, esta representação é adotada como um padrão IGES (de “*Initial Graphics Exchange Specification*”) para superfícies 3D de sistemas CAD/CAM industrial. Maiores detalhes sobre estas superfícies podem ser encontradas em Tiller [Til83], Lehmann [Leh70] e Figueiredo [Fig91].

2.2.3 Representação de volume

A representação de volume descreve um sólido baseado no espaço volumétrico que ele ocupa. Segundo Flynn [Fly91], um meio de representar um objeto volumetricamente é colocá-lo em algum sistema de coordenadas tri-dimensionais e dividir o volume que ele ocupa em elementos menores de volume, tal que estes elementos menores englobem partes do objeto.

A maioria das descrições volumétricas empregam cubos ou paralelepípedos (box-contornado) conhecido como *voxels*. A dimensão do voxel é relativa, um único *voxel* pode ocupar um grande espaço ao representar um objeto ou ainda um grande número de *voxels* pode ser requerido para representar um determinado objeto. O critério de erro para definir a dimensão do *voxel* vai depender da aproximação desejada no projeto.

Nos sistemas CAD, as técnicas mais utilizadas para representação de volume são: *octree*, superquádrico e técnica de geometria de sólidos construtivos.

A *octree* é uma técnica recursiva de decomposição de um espaço cúbico em sub-cubos. Em geral, esta técnica associa o objeto com uma representação de árvore, onde o espaço total é representado por um nó simples chamado raiz. Se todo o volume do objeto é homogêneo (totalmente cheio ou totalmente vazio), a raiz não se decompõe e compreenderá a descrição completa do espaço (um cubo). Caso contrário, se houver partes não homogêneas, o referido cubo é dividido em oito sub-cubos iguais (octantes). Cada octante, por sua vez, engloba parte do objeto e é analisado para verificar também se está homogêneo. Se estiver, o nó correspondente na árvore não se decompõe, é considerado um nó terminal. Se não estiver homogêneo, é necessário subdividir o octante em análise em outros oito sub-cubos, e assim por diante. O processo continua até que todos os pontos de nós sejam homogêneos ou alguma limitação seja alcançada. Então, a representação de árvore descreverá o objeto, através das ramificações de seus pontos de nós. Maiores detalhes sobre esta técnica podem ser encontrados nos trabalhos de Oshima [Osh83] e Sheneier [She87]

A *superquádrica* (de “*superquadric*”) é uma extensão da função quadrática de superfície. Um exemplo típico de uma superquádrica é um elipsóide descrito pelos parâmetros de contorno ϵ_1 e ϵ_2 , tal que estes parâmetros possam deformar o elipsóide,

transformando-o em outro sólido, um paralelepípedo ou um objeto em forma de banana, por exemplo. Este contorno deformado é freqüentemente referido como um superelipsóide. O superelipsóide então, pode curvar-se, inclinar-se e afinar-se para obter o modelo desejado, uma vez que pode ser descrito de forma paramétrica.

Solina [Sol90] desenvolveu um método para adequar as superquádricas aos arranjos de dados utilizados em Visão por Computador, considerando a seguinte forma implícita para descrevê-los:

$$f(x, y, z) = \left[\left(\frac{x}{a_1} \right)^{2/\epsilon_2} + \left(\frac{y}{a_2} \right)^{2/\epsilon_2} \right]^{\epsilon_2/\epsilon_1} + \left(\frac{z}{a_3} \right)^{2/\epsilon_1} - 1 = 0, \quad (2.6)$$

onde, os parâmetros a_1, a_2, a_3 determinam o tamanho da superquádrica nas direções das coordenadas x, y e z , respectivamente. Os parâmetros ϵ_1 e ϵ_2 indicam o contorno do objeto, os quais são expressos em coordenadas polares e quando combinados, indicam o grau das deformações (“squareness”) do objeto, ou seja o tamanho, inclinação, afinamento e curvatura (enroscamento). O parâmetro ϵ_1 está relacionado com a deformação latitudinal e ϵ_2 com a deformação longitudinal. Estas características volumétricas criam uma ampla variedade de primitivas curvas, tais como esferas, cilindros, paralelepípedos, elipsóides e modelos de contornos deformados. Se $\epsilon_1 = \epsilon_2 = 1$, a função define um elipsóide e, se, além disto, $a_1 = a_2 = a_3$, a função define uma esfera. Se $\epsilon_1 \ll 1$ e $\epsilon_2 = 1$ tem-se um cilindro. Se ambos ϵ_1 e $\epsilon_2 \ll 1$ tem-se um paralelepípedo. Valores de ϵ_1 ou ϵ_2 ou ambos forem maiores que 2, têm-se modelos de contornos deformados. Solina [Sol90] mostra algumas figuras de superquádricas sem e com as deformações conseguidas através da aplicação da equação 2.6.

Alguns trabalhos utilizam as primitivas de superquádricas em problemas de ajustamento de superfície na reconstrução de objetos 3D [Bol91, Bes92], outros em modelagem de objetos de contornos curvos [Arm93, Bor96].

A técnica de Geometria de Sólidos Construtivos (CSG, de “Constructive Solid Geometry”) é especificada como sendo um conjunto finito de primitivas volumétricas, tais como cilindros, cubos, esferas e outros, que são combinadas através de operações booleanas como união, diferença e interseção para construir um objeto 3D. A Figura 2.2 ilustra os passos necessários para a construção de um modelo de objeto. A estrutura de dados é armazenada como uma árvore binária, construída de baixo para cima, onde os

nós terminais representam as instâncias de primitivas geométricas (volumétricas) e as ramificações representam as operações booleanas realizadas bem como as informações sobre o posicionamento das partes [Bcs85]. A árvore CSG define o volume do objeto e a área da superfície construída, podendo até não ser única para cada modelo (mais de uma árvore pode produzir o mesmo modelo).

Uma grande vantagem da representação CSG é que ela é capaz de representar objetos complexos com pequena quantidade de dados. É uma representação muito utilizada na construção de modelos de objetos por sistemas CAD/CAM, muito embora ainda não construa com facilidades modelos com superfícies esculturais, isto é, não lisas, como a face humana, por exemplo.

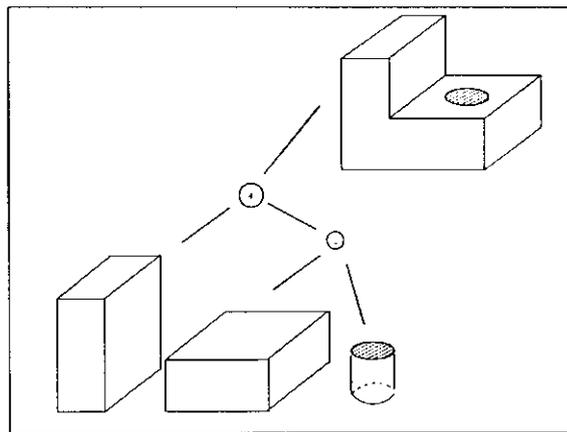


Figura 2.2: Árvore de CSG

2.2.4 Representação de Imagem Gaussiana Estendida

O método da esfera Gaussiana é utilizado para representar objetos sólidos através da orientação das superfícies do objeto. Se um sólido possui um conjunto de vetores normais unitários associado com cada uma de suas faces tal que cada vetor normal possa ser transladado para uma origem comum, retendo a sua direção original, então as extremidades destes vetores definem uma esfera Gaussiana e o lugar destas extremidades chama-se Imagem Gaussiana do sólido. A maior dificuldade em utilizar a imagem

Gaussiana é quando os sólidos têm superfícies não convexas porque dois sólidos não convexos poderão ter a mesma imagem Gaussiana. Além disto, algumas informações da superfície do sólido são perdidas, uma vez que apenas a informação sobre a direção da normal é preservada.

Um meio de adicionar mais informação à representação da imagem Gaussiana é tratar os pontos da imagem Gaussiana como pontos de massa calculados por alguma função peso [Roa87]. O mapeamento de normais das superfícies para dentro da esfera unitária Gaussiana, tendo a área da superfície como peso é chamado de Imagem Gaussiana Estendida (EGI, de "*Extended Gaussian Image*"). Nos poliedros, a EGI pode ser obtida colocando-se um peso em cada ponto da superfície Gaussiana igual à soma da área das faces na orientação correspondente. Isto pode ser facilmente calculado e pode ser usado para encontrar a orientação do objeto. A EGI portanto, representa a área total da superfície do objeto, mapeada para dentro da esfera unitária da superfície Gaussiana. Esta técnica é utilizada na análise de dados mas não na síntese de contornos de superfície e independe da posição do objeto.

Bhanu [Bha87] propõe uma abordagem para aplicações envolvendo o reconhecimento e manipulação de objetos 3D, para ambientes industriais. O sistema de Bhanu desenvolve algumas representações, a partir de modelos CAD, que poderão ser utilizadas em diversas aplicações. A técnica de EGI é utilizada para analisar o modelo CAD, extraindo-se as informações necessárias para construir a representação de curvaturas das superfícies do objeto. Outras informações sobre o uso da EGI podem ser encontradas no trabalho de Horn [Hor84].

2.2.5 Representação de Varredura

A representação de varredura de um objeto é descrita como sendo o volume do objeto varrido por um conjunto de seções transversais ao longo de um eixo, obedecendo uma regra de varredura. Em outras palavras, um espaço tri-dimensional é descrito por uma função bi-dimensional. A classe amplamente usada nesta representação é a do cilindro generalizado (CG ou GC, de "*Generalized Cylinder*"), introduzida na área de Visão por Computador por Nevatia-Binford [Nev77] num sistema de reconhecimento

de objetos 3D de superfícies curvas. A idéia do CG é que um corpo cilíndrico pode ser descrito por um círculo de raio definido por uma função, a qual representa o contorno de uma seção transversal em torno de uma linha reta representando o eixo de simetria do corpo cilíndrico. Se o raio é descrito por uma função linear ao longo do eixo, então a varredura irá gerar figuras que, juntas, assemelham-se a cones¹.

Esta representação é adequada para modelar objetos que possuam eixo de simetria definido. Embora seja inambígua para uma dada função de varredura, não há meios eficientes para descrever o contorno dos modelos nas suas extremidades. Além de poder descrever objetos curvos (3D) em formato 2D, esta representação tem a vantagem de que é invariante à rotação e translação dos objetos em relação ao plano $X-Y$. Os CGs são poucos utilizados em sistemas CAD. Outro trabalho que utilizou esta representação num sistema automático de modelagem é o de Brooks [Bro79], conhecido por *ACRÔNIMO* (“*ACRONYM*”). Outro exemplo mais recente é o sistema de reconhecimento de contornos de objetos 3D desenvolvido por Chung [Chu94].

Segundo Requicha [Req80], esta técnica é utilizada em alguns sistemas CAD, cuja representação de sólidos é considerada um padrão ANSI (de “*American Society of Mechanical Engineers*”), aplicado a projetos mecânicos.

2.2.6 Outras Representações

Alguns sistemas CAD fazem uso de esquemas de representação híbrida. Estes são projetados pela combinação das abordagens discutidas anteriormente. Podem ser citados como exemplo, as representações híbridas CSG/B-rep e CSG/varredura. Nestes casos, as combinações representam estruturas de árvore CSG compostas por representações B-rep ou de varredura, respectivamente. O sistema deve ser capaz de tratar cada representação em separado e quando combinadas - isto é um problema um tanto complicado de resolver. Alguns sistemas preferem as representações híbridas em separado visando utilizar programas mais simples.

O Padrão Internacional de Especificação de Troca Gráfica Inicial (*IGES*) é um conjunto de normas e padrões industriais desenvolvido pelo “*National Institute of Stan-*

¹Esta representação também é conhecida como cone generalizado.

dards and Technology" ("U.S. Dept. of Commerce 1988") para permitir a transferência ou troca de modelos CAD entre os vários sistemas CAD comerciais. O IGES define as entidades de projeto padrão para desenhos de objetos em 2D e 3D, bem como os diagramas elétricos e outros elementos industriais. Muitos sistemas CAD são capazes de converter arquivos de suas representações internas para representação IGES e vice-versa.

No formato IGES a representação de sólidos 3D consiste de numerosas entidades tais como plano, curva/superfície spline paramétrica, curva/superfície spline racional, superfície de revolução, linhas, arcos circulares, curvas NURBS, seções cônicas e outras. A última revisão do IGES permitiu descrições de objetos 3D na forma de árvore CSG [Fly91]. Vários sistemas CAD são compatíveis com o IGES, estes possuem programas adequados tanto para ler quanto para escrever arquivos neste formato. O maior obstáculo em usar o formato IGES é a lacuna de informações topológicas do modelo na outra representação, isto é, as relações de várias entidades requeridas em um sistema devem ser deduzidas usando apenas as informações disponíveis no outro sistema. Tal problema ocorre devido as diferenças existentes entre as primitivas das duas representações.

2.2.7 Conclusões

A classificação dos tipos de representações de sólidos por CAD não é rigorosa nem definitiva, varia ao longo dos anos e de autor para autor. A classificação apresentada neste trabalho difere da classificação proposta por Requicha (1980), que inclui as representações de primitivas puras, ocupância espacial e decomposição de células e não inclui as representações de superquádricas e de B-rep. Difere de Besl (1985) e Flynn (1991), que não consideram a representação de CSG como sendo uma representação de volume e difere também de Armann (1993), que considera a representação de varredura (CG) como sendo uma representação volumétrica. A classificação apresentada é mais parecida com a de Bhanu [Bha87] (1987), sendo que as representações de primitivas puras, ocupância espacial e decomposição de células, por não serem mais utilizadas nos sistemas CAD atuais, foram substituídas pelas representações de superquádricas e

octree. Além disto, a classificação apresentada inclui as representações de superfícies quádricas e paramétricas (B-rep), não citadas por Bhanu.

Na classificação apresentada, a representação de CSG é considerada uma representação de volume porque é constituída de primitivas geométricas de volume (cubos, esferas, cilindros, etc), concordando assim com a definição desta representação. A representação de varredura não é considerada uma representação volumétrica porque apesar de tratar com objetos de superfícies cilíndricas (3D), esta descreve o objeto em forma bi-dimensional (2D), através da função matemática do círculo representando a seção transversal do cilindro, e portanto *não* volumétrica.

2.3 Extração de Características da Cena

Ao iniciar o projeto de um sistema de reconhecimento de objetos, deve-se planejar como será feita a aquisição da imagem do objeto, quais as características necessárias para serem extraídas das imagens e como representá-las adequadamente no computador, antes de trabalhar no algoritmo de reconhecimento. Tais representações são utilizadas para descrever não apenas o objeto na cena, como também para descrever os modelos, pois elas ditam as estratégias do casamento. A seguir será feita uma abordagem sobre cada uma das etapas citadas.

2.3.1 Aquisição de imagens do objeto

As técnicas de reconhecimento de objetos 3D através de arranjo de imagens ou de múltiplas vistas empregam métodos baseados na estrutura geométrica formada pela câmera, projetor de luz (se houver) e objeto. As técnicas mais comuns são: triangulação ("*triangulation*") e retorno de feixe ("*time-of-flight*"). Na triangulação, a estrutura física forma um triângulo em plano epipolar, em cujos vértices estão a câmera, o projetor de luz ou outra câmera e o objeto (cena). A triangulação pode ser passiva (estéreo) ou ativa. A triangulação estéreo (estereoscopia) envolve duas ou mais câmeras (com um ou mais triângulos) enquanto a ativa envolve projetor de luz em um dos cantos do triângulo [Gon93]. A Figura 2.3 ilustra esta estrutura.

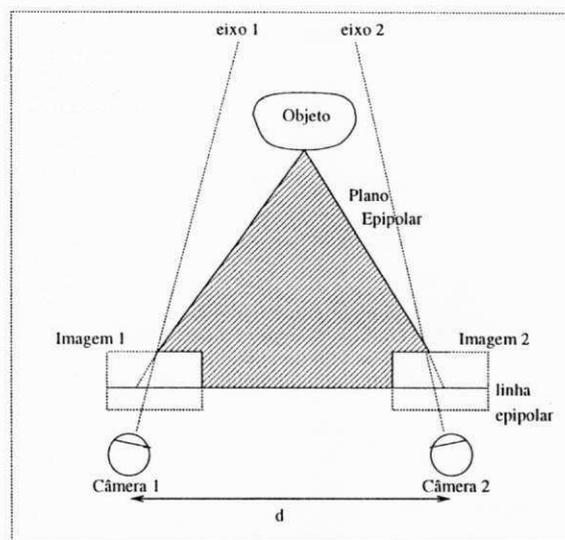


Figura 2.3: Estrutura para técnica de triangulação.

A interseção das linhas laterais do triângulo com o plano das imagens das câmeras (plano perpendicular aos eixos) denominam-se de linhas epipolares. Tais linhas são muito utilizadas como referência para cálculos de correspondência entre imagens estéreo. Ito [Ito86], por exemplo, utiliza o cálculo de coeficientes de correlação da distribuição do brilho nas linhas epipolares contidas em três vistas do objeto, para encontrar os pontos de interseção entre as vistas. Ayache [Aya94] também faz pesquisa de pontos nas linhas epipolares de três vistas, comparando algumas características como comprimento, gradiente e ângulo para encontrar a correspondência entre as imagens.

A técnica de retorno de feixe consiste de um transmissor de sinal direcionado ao objeto e um receptor (sensor) eletrônico que capta o sinal refletido pelo objeto e mede o tempo gasto pelo retorno do sinal (reflexão) e sua intensidade. Este tempo é proporcional à distância e às dimensões do objeto. Na prática, são usados transmissores de ultra-som ou de luz laser. Esta técnica é mais sofisticada e tem resolução pobre devido às dificuldades em gerar feixes acústicos estreitos [Nit88].

A estereoscopia requer montagens de estruturas mais simples e é mais confiável,

uma vez que já estão sendo comercializadas câmeras digitais com boa resolução.

Após a obtenção das imagens da cena (descrição de baixo nível), estas passam por processos de tratamento digital de sinais, visando extrair as informações úteis. O tratamento digital de imagens inclui as etapas comuns de pré-processamento (realce, redução de ruídos), e segmentação para serem retiradas das imagens as características úteis ao reconhecimento.

2.3.2 Descrição do objeto na cena

Os tipos de características que serão utilizadas na fase de casamento de atributos, levam a definir com antecedência qual é a representação que contém aquelas características. Por exemplo, se for definido que a área é o atributo que vai ser utilizado no casamento, é mais conveniente utilizar a representação de superfícies para que seja conseguido tal atributo da representação do objeto.

As principais representações já foram mencionadas na Seção 2.2, na modelagem. Tais representações podem também ser utilizadas nesta etapa de descrição da cena.

Na representação de superfície pode ser incluído o método de “crescimento de região”, descrito por Srihari [Sri81], onde as superfícies são caracterizadas a partir de um conjunto de três pontos não co-lineares. Os referidos pontos *pixels*, pertencentes a diferentes superfícies do objeto, são determinados e adicionados com os *pixels* vizinhos verificando a similaridade entre eles. Onde há similaridade, aumenta-se o tamanho da região em cada dimensão até encontrar o seu limite, caracterizando-se assim três superfícies do objeto. Quando o crescimento encontra o seu limite, o processo pára. Escolhe-se outro conjunto de três pontos que não pertença a uma região já investigada e o procedimento reinicia. O processo continua até que todos os *pixels* de todos os planos do objeto tenham sido investigados. Para uma maior eficiência do método, a técnica pode ser aplicada simultaneamente em várias regiões, levando sempre em consideração as similaridades com as regiões vizinhas no processo de crescimento. O limite de similaridade necessário nesta abordagem é expresso pela probabilidade do *pixel* ser semelhante ao seu vizinho. Esta informação é obtida da mudança do gradiente local no contorno da superfície.

2.3.3 Tipos e representações de características

As características mais importantes (primitivas) na análise da imagem digital estão agrupadas em [Bes85]:

1. *características globais* - que estão relacionadas com as propriedades geométricas do objeto: perímetro, centróide, área, momento, normal e outras;
2. *características locais* - que estão relacionadas com as bordas do objeto: segmentos de linha, comprimento de arcos, vértices e outras;
3. *características relacionais* - que descrevem as relações que existem entre as características locais: ângulo entre duas arestas ou duas normais, arestas entre superfícies, distâncias entre centróides de superfícies e outras.

As representações das características mencionadas e que são mais utilizadas são [Chi86]:

1. *mapas de profundidade* - tabela com valores das três dimensões: (x, y, z) ;
2. *estruturas de árvores hierárquicas* - composta de níveis e subníveis, representando partes e subpartes do objeto;
3. *estruturas de grafos relacionais* - composta de arcos e nós, representando partes e ligações.

Em geral, extraem-se dois tipos de características do objeto, característica local e relacional ou global e local ou global e relacional, etc, de tal modo que uma delas tenha um nível de detalhamento da característica maior do que outro. Este detalhe fará a distinção de um modelo do objeto entre os demais. Por exemplo, numa determinada superfície as características locais podem ser: arestas e vértices, enquanto as características globais podem ser: comprimento da aresta, área, número de vértice. Mesmo que dois modelos coincidam com as características locais, eles podem diferenciar nas características globais.

O tipo de representação das características utilizadas deve ser adequado com as próprias características. Por exemplo, mapas de profundidade são adequados para representar: vértices, normal, centróide e outras.

2.4 Casamento de Características - Reconhecimento de Objetos

O reconhecimento de objetos está relacionado diretamente com o problema de casamento de dados ou seja, correspondências de características. Grimson [Gri90] classifica os sistemas de reconhecimento baseado nas características globais (objeto inteiro) e nas características locais (partes do objeto). O casamento de características globais inclui parâmetros geométricos como área, perímetro, número de Euler, momentos de inércia, transformadas discretas, etc. Podem ser citados os trabalhos de Dudani [Dud77], Wallace [Wal81], Engebretch [Eng88] e Wang [Wan84].

Na maioria dos trabalhos, a etapa de reconhecimento é realizada em dois passos: a geração de hipóteses e a verificação das hipóteses. Na geração de hipóteses, é estabelecida uma correspondência entre dois conjuntos de descrições, um conjunto obtido da imagem do objeto na cena e o outro conjunto do modelo armazenado. Nesta etapa, o algoritmo de casamento faz uma varredura em todos os modelos armazenados procurando quais os que casam com o conjunto de descrições do objeto, obedecendo a algum critério de erro que selecione os prováveis modelos. O algoritmo deve ser capaz de alcançar este passo usando uma descrição parcial do objeto e uma descrição completa do modelo. Podem surgir, nesta etapa, vários modelos candidatos ao reconhecimento mas é no próximo passo que apenas um deles é escolhido.

A verificação de hipóteses, é estabelecida através de um novo casamento, agora apenas entre os modelos selecionados no primeiro passo com um segundo conjunto de descrições mais detalhadas do objeto. Neste casamento, apenas um modelo deve ser selecionado dentre aqueles candidatos ou então nenhum, significando que o sistema não reconhece aquele objeto. O modelo selecionado é o modelo procurado e o reconhecimento do objeto é concluído com a descrição completa do modelo encontrado.

O casamento está relacionado com o tipo de representação de dados utilizados nas características extraídas do objeto e usados na modelagem. Por exemplo, modelos que utilizam a representação de cilindros generalizados para representar o objeto na cena podem usar a representação de grafo relacional para representar aquele cilindro. Associado ao grafo relacional pode ser produzida uma tabela característica contendo

dados específicos (dimensões, raio, área, etc) de uma região do objeto. Assim, na etapa de geração de hipóteses, o grafo relacional do objeto é comparado ao grafo relacional de cada modelo da base de dados e serão selecionados os modelos com grafos casados. Em seguida, na etapa de verificação de hipóteses, o modelo é selecionado quando a tabela do objeto é comparada com a tabela dos modelos selecionados no primeiro passo e portanto, apenas um modelo é capaz de coincidir todos os dados e ser escolhido.

O estabelecimento do casamento de correspondências entre os modelos e a cena requer que as características encontradas sejam arbitrariamente ordenadas e a primeira característica da cena seja hipotetizada para fazer a correspondência com cada característica dos N modelos. Numa pesquisa de árvore, por exemplo, isto representa o primeiro nível da árvore, obtendo-se cada hipótese ilustrada na Figura 2.4.

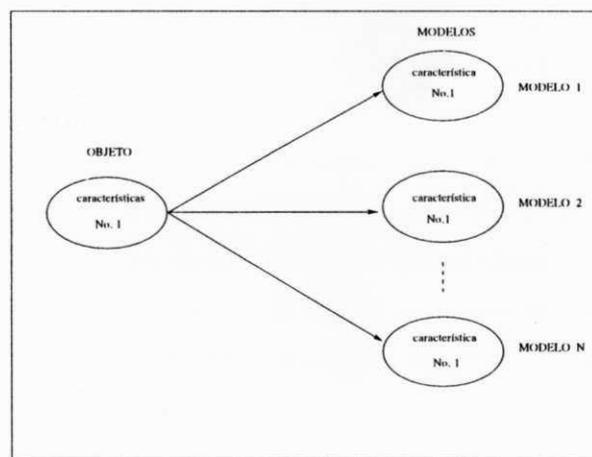


Figura 2.4: Casamento de características 1.

A segunda característica do objeto é hipotetizada para fazer a correspondência com cada segunda característica dos modelos. Isto representa o segundo nível da árvore, adicionado na pesquisa de árvore. Cada caminho nesta árvore, da raiz para as ramificações, representa uma solução possível para o problema de casamento de correspondências. O objetivo é investigar um caminho tal que uma transformação rígida do objeto para o modelo possa ser encontrada. Esta pesquisa, no entanto, não é muito utilizada com estruturas grandes de árvore, pois o processo torna-se computacionalmente exaustivo.

Em estruturas de grafos, onde os nós representam tipos de superfície e os arcos representam as relações entre elas, o casamento de correspondências é abordado como um problema de isomorfismo. A pesquisa de correspondências para estruturas de grafo relacional tem procedimentos semelhantes a pesquisa de árvore.

A maioria dos métodos de reconhecimento, além de identificar o modelo do objeto da cena, ainda fornecem informações adicionais como a localização e posição exatas do objeto na cena ou mesmo a localização de uma superfície onde será feita alguma operação. Este procedimento requer o conhecimento de transformações geométricas, uma vez que o modelo descrito em um sistema de coordenadas deve ser reconstruído em outro sistema de coordenadas.

2.5 Análise de Alguns Métodos

O método proposto por Nevatia e Binford [Nev77] foi um dos primeiros a considerar o reconhecimento de objetos 3D, onde foi introduzido o conceito de cilindros (ou cones) generalizados para representar a superfície tanto do objeto quanto do modelo. Esta representação enfatiza a análise de cenas com objetos curvos. As características extraídas das vistas 2D do objeto são colecionadas em conjuntos de arestas chamados *ribbons*, que são representadas por uma estrutura de grafos relacionais associados aos conjuntos de CG's correspondentes às subpartes do objeto. Os nós do grafo relacional representam o encontro (juntas) dos *ribbons* e os arcos aos *ribbons*. Em seguida, um conjunto de descritores é construído associado com o grafo de cada objeto, contendo características das tiras. Para cada peça distinguível extraída de um objeto, um código binário é gerado para descrever e facilitar a organização e a pesquisa na biblioteca de modelos. O reconhecimento ocorre primeiramente com o casamento entre o conjunto de descritores do objeto e o conjunto de descritores de todos os modelos; depois com o casamento entre o grafo relacional obtido da cena com o grafo relacional de cada modelo selecionado no primeiro casamento.

O método de Oshima e Shirai [Osh83] foi um dos primeiros a tratar de uma variedade de classes de objetos tri-dimensionais, o qual é capaz de reconhecer uma cena contendo múltiplos objetos em qualquer direção. Os objetos podem ter superfícies

planas, curvas ou ambas. A descrição da cena e as características dos objetos são conseguidas através da estereoscopia ativa em triangulação. Os modelos são construídos das imagens do próprio objeto a ser reconhecido, o qual é mostrado com antecedência ao sistema em posições diferentes. Os autores utilizaram a representação de volume para descrever modelos e objetos. A descrição é feita usando técnicas de segmentação de superfícies planas e curvas das imagens individuais. As propriedades geométricas encontradas durante a segmentação são representadas num grafo relacional onde os nós representam as superfícies extraídas do mapa de profundidade da cena e os arcos representam as relações entre as superfícies adjacentes. As regiões ou superfícies extraídas do(s) objeto(s) na cena que possuem maior potencial de casamento com o modelo são chamadas de *kernels* ou núcleos. O procedimento de reconhecimento é realizado comparando-se o grafo relacional das descrições do objeto na cena com o grafo relacional de cada modelo na biblioteca de modelos, através de dois passos. Primeiro, é feita uma pesquisa de grafos em todos os modelos, selecionando os modelos que casam as superfícies de *kernels*. Segundo, o sistema faz um novo casamento entre as demais superfícies da cena com as superfícies vizinhas do *kernel* de cada modelo candidato encontrado no primeiro passo, visando casar todas as regiões do objeto na cena.

O sistema 3DPO (de "A 3D-Part Orientation") é um sistema de visão para robô, desenvolvido por Bolles et al [Bol86] com o objetivo de reconhecer, localizar e agarrar peças industriais moderadamente complexas. A estratégia abordada no 3DPO para o reconhecimento e localização do objeto na cena é tal que o sistema deve identificar inicialmente uma característica-chave (distinguível) do objeto, classificando-as em alguns grupos de modelos. Em seguida, esta primeira característica irá sugerir qual a segunda característica a procurar. Estas duas características juntas fazem a previsão da terceira característica a procurar, classificando subgrupos de modelos dentro dos grupos e assim por diante, até que o modelo do objeto possa ser identificado e localizado confiável e precisamente. As características foram extraídas num dispositivo "scanner" em montagem de triangulação. Uma luz laser ilumina a cena contendo os objetos numa mesa de trabalho e o sensor do dispositivo mede as componentes x , y e z dos objetos bem como a intensidade luminosa dos pontos que varrem toda a cena. As arestas são localizadas através das discontinuidades nas imagens obtidas, em seguida

são classificadas em grupos para formar uma rede de características do objeto. Para checar cada uma das hipóteses, o sistema constrói o modelo de remendos, baseado nos dados encontrados na geração de hipóteses. O modelo estimado representa o que o sensor teria visto se o objeto estivesse na posição hipotetizada. Este é comparado com a imagem do objeto na cena e então toma decisões de acordo com a correlação entre as duas imagens. Se a soma dos erros for grande, o programa pode rejeitar as hipóteses geradas. Caso contrário, o programa faz uma avaliação mais estrutural do casamento (refinamento de parâmetros). Quando dois ou mais objetos apresentam mesma posição, significa que eles podem estar superpostos (empilhados). O programa então, prossegue a análise para identificar o objeto que está no topo da pilha (desde que os outros estejam parcialmente ocultos).

Brooks et al [Bro79] desenvolveram um sistema de Visão Artificial chamado ACRÔNIMO, para reconhecer aviões nas pistas de um aeroporto através de fotografias aéreas tiradas de cima dos mesmos. Também utilizaram a técnica de cilindro generalizado e representação de objetos através de tiras e elipses do cilindro generalizado, cujas características são representadas por grafos relacionais. Cada objeto a ser reconhecido ou modelado tem uma estrutura de árvore que apresenta uma descrição hierárquica do objeto, onde os nós mais altos correspondem às partes mais significativas do objeto. Em seguida, o sistema gera um conjunto de regras capaz de identificar o contorno das superfícies e prever a aparência do objeto em termos de tiras e elipses observados na imagem. Como resultado deste processo, o sistema constrói o grafo de predição, onde os nós correspondem às características invariantes ou quase invariantes da imagem e os arcos correspondem a quão provável o par de características que se unem a eles pode ocorrer no objeto, através de uma das ocorrências: “deve ser”, “deveria ser” ou “excluído”. O reconhecimento é realizado em duas etapas. Primeiro, há um casamento de classes do objeto definida pelo nó raiz dos grafos de restrição. Segundo, os modelos selecionados na primeira etapa são checados para verificação de consistência global, ou seja, é realizado outro casamento entre os grafos de restrições dos modelos selecionados com o grafo de restrição do objeto na cena. Na verificação de hipóteses, o casamento é realizado entre as subclasses dos modelos selecionados e do objeto na cena.

Bhanu [Bha84] apresenta um sistema para reconhecimento de objetos tri-dimen-

sionais através de casamento de contornos do objeto. O método combina as vistas individuais de um objeto contra um modelo em 3D. A aquisição da imagem é baseada na estereoscopia ativa em triangulação. O objeto é colocado numa mesa giratória que faz um movimento para cima e para baixo (eixo- y) e um movimento de rotação (eixo- x). O feixe é refletido no objeto e a distância z é calculada a partir da localização da resposta máxima em um banco de detetores. Extraem-se as vistas do objeto obtidas em cada ângulo igualmente espaçado do movimento de rotação. A representação utilizada para o objeto e os modelos é a de superfície, determinada através do método de crescimento de região seguida de aproximação poligonal. Em cada vista, as faces encontradas são rotuladas (numeradas) enquanto é confeccionada uma tabela referente às suas propriedades. O casamento consiste em identificar cada face rotulada do objeto desconhecido com uma face de mesmas propriedades entre os modelos. O processo de calcular todas as probabilidades que uma face do objeto desconhecido tem de pertencer a uma face de um modelo é denominado rotulagem estocástica de face. As propriedades geométricas das faces (ou superfícies) são usadas para computar as probabilidades iniciais da rotulagem estocástica de face. A estrutura de características é dividida em níveis de hierarquia de modo que o nível mais alto da estrutura contém mais informação contextual do objeto. A compatibilidade entre uma face da vista desconhecida com uma face do modelo é obtida por transformações matemáticas obedecendo critérios de erros nas medidas encontradas. Estas transformações se baseiam nos valores de escala, translação, orientação e rotação das faces envolvidas. O algoritmo de casamento é realizado de modo iterativo procurando maximizar as medidas de compatibilidade dos níveis mais baixos da estrutura, de acordo com a probabilidade inicial. Em seguida, utilizam-se estes resultados para acelerar e melhorar a precisão dos resultados nos níveis mais altos da hierarquia. Com este procedimento, o algoritmo vai eliminando os modelos que tenham resultados inconsistentes ou ambíguos (duvidosos) após cada iteração. Quando o algoritmo chega ao topo da estrutura apenas um modelo é consistente, o que significa que o procedimento concluiu a tarefa de casamento. No final, as informações de rotação e translação do objeto também estão disponíveis.

Majumdar et al [Maj89] desenvolveram um sistema de visão para robô que identifica e localiza a orientação de uma superfície-chave, onde é feita uma operação específica,

em um objeto que encontra-se na mesa de trabalho do sistema. A superfície-chave referida é a superfície do objeto que possui características especiais adequadas para o seu reconhecimento e, na qual o robô irá realizar uma operação. A descrição da cena é produzida através de três imagens do objeto fotografadas sobre três planos: plano da mesa de trabalho $X-Y$, plano de elevação $X-Z$ e plano de fundo (profundidade) $Y-Z$. As características geométricas do objeto são extraídas das fotos do objeto e processadas através de técnicas comuns de pré-processamento digital de imagens. As características utilizadas para o casamento são aquelas que são invariantes a posição de rotação do objeto. Os modelos do objeto são construídos no sistema CAD. O algoritmo de reconhecimento é dividido em dois módulos: o módulo *teach-in*, ou seja de auto-aprendizagem, e o módulo de visão. O módulo *teach-in* atua nas características geradas do modelo CAD, preparando-as para a etapa do casamento. O módulo de visão faz o casamento das características extraídas do objeto com as do módulo *teach-in*. Inicialmente, o módulo de visão pesquisa na base de dados do módulo *teach-in* quais os modelos cuja base de dados casam com os parâmetros característicos extraídos das imagens fotografadas do objeto. Isto determina as posições estáveis do objeto. Após a identificação de qualquer posição estável e seguindo um conjunto de regras de visibilidade do modelo encontrado, é possível identificar não só a localização como também a orientação da superfície-chave na posição presente do objeto.

Wong [Won92] descreve um método de reconhecimento de objetos tri-dimensionais utilizando técnicas de casamento de subgrafos. O objeto 3D é modelado por grafos de atributos; a imagem das partes visíveis do objeto é obtida através de câmeras de TV interligadas ao microcomputador e processada de modo a extrair seu desenho de linhas em 2D. O desenho de linhas é representado por grafos de projeção, os quais serão utilizados na etapa do reconhecimento para o casamento com os grafos do modelo. Para o casamento, é necessário que um grafo de projeção da imagem da cena seja um subgrafo do grafo do modelo armazenado. Um nó no grafo de projeção corresponde a uma junção de linhas (projeção de um vértice do objeto 3D); um arco corresponde a uma linha no desenho de linhas (projeção de uma aresta do objeto 3D). Para representar as linhas conectadas às junções ocultas do objeto 3D, utilizam-se arcos abertos, ou seja, que não se conectam a qualquer nó no grafo de projeção. O modelo é construído a partir do

próprio objeto, utilizando o mesmo esquema para aquisição de imagens. Os nós do grafo do modelo representam um vértice do objeto 3D e, os arcos entre dois nós representam as arestas do objeto 3D. Para o grafo representar o modelo completamente, são obtidas mais de uma imagem do objeto. Portanto, os grafos capturam a topologia representada pelos vértices conectados às arestas do objeto. Tanto o grafo do modelo, quanto o grafo de projeção é associado com uma tabela que possui uma lista ordenada dos nós com seu tipo de junção e com o tipo de junção dos seus três nós vizinhos. O procedimento de reconhecimento de objetos ocorre primeiramente pela geração de hipóteses através de isomorfismo de subgrafos, em seguida, pelo casamento de atributos contidos nas tabelas. Os atributos dos modelos casados também estabelecem a correspondência de pares de pontos entre a imagem 2D (desenho de linhas) e o modelo 3D para estimar os valores da transformada da câmera. Esta transformada é usada para projetar os modelos geométricos 3D, obtidos na geração de hipóteses, para dentro do plano da imagem, isto é, para a comparação da imagem projetada com a imagem da cena. Esta etapa é chamada de verificação final de hipóteses, cujo algoritmo é capaz de distinguir diferenças entre as dimensões ou detalhes geométricos que o casamento de subgrafos não faz, selecionando apenas um modelo correspondente ao objeto na cena.

O trabalho de Park et al [Par93] apresenta uma nova idéia para descrever e reconhecer objetos tri-dimensionais. Os autores desenvolveram um método que se baseia nas imagens extraídas do objeto da posição ortogonal à normal de cada superfície. O conjunto de imagens extraído do objeto na posição da normal à superfície é chamado pelos autores de Imagens da Normal de Superfícies (SNI, de "*Surface Normal Images*"). A partir da fotografia do objeto na cena, extraem-se as superfícies por técnicas de segmentação. Para se obter as vistas ortogonais do objeto na cena, rotaciona-se a imagem do objeto de tal modo que cada superfície do objeto se posicione de frente ao observador, ortogonal à normal de uma de suas superfícies. Em outras palavras, rotaciona-se de tal modo que o eixo-Z coincida com a normal da superfície escolhida. Nesta posição, têm-se as vistas da superfície-base e das demais que consegue visualizar na posição do observador. Assim, são obtidas várias imagens denominadas de imagens de entrada rotacionadas (RII, de "*Rotated Input Images*"). Os modelos são construídos com base nas SNI's extraídas de cada objeto. Cada SNI é composta pela superfície-

base (ortogonal à normal) e suas superfícies vizinhas visíveis. Neste método, a vista da posição da normal da superfície apresenta uma referência direta da transformada entre o modelo e o objeto na cena, reduzindo bastante o tempo requerido no reconhecimento, em relação ao tempo requerido nos métodos que extraem as características da cena para análise. O reconhecimento se dá quando é feito o casamento das RII's com as SNI's em uma sequência de testes, determinada de acordo com os atributos de cada superfície-base, por exemplo: primeiro as superfícies planas seguidas das curvas, ou então, as superfícies maiores seguidas das menores. Inicialmente, o programa procura as SNI's que casam com a primeira RII da sequência de testes, selecionando alguns modelos e eliminando outros. Em seguida, a segunda RII da sequência de testes é casada com as SNI's dos modelos selecionados na primeira iteração e novamente o programa vai selecionar alguns modelos e eliminar outros, diminuindo a quantidade de modelos selecionados. O procedimento de casamento continua com as iterações na sequência até que um único candidato seja encontrado.

Borges [Bor96] apresentou recentemente um sistema de reconhecimento de objetos que identifica e localiza objetos curvos complexos com articulações, a partir de um arranjo de imagem. O sistema proposto é caracterizado por apresentar técnica de segmentação para representações parametrizadas de superquadráticas deformadas correspondentes às partes do objeto e um conjunto de primitivas volumétricas (geons) de superquadráticas para descrever modelos e objetos na cena. A fase de interpretação é composta por três etapas: classificação das partes, casamento e estimação de posição. A classificação dos contornos dos superquadráticas em um dos geons é feita por uma função neural. Na etapa do casamento o algoritmo faz uma pesquisa na biblioteca de modelos tentando o casamento dos geons classificados. As hipóteses estabelecidas são analisadas numa árvore de interpretação para avaliar as ligações entre elas. As hipóteses aceitáveis são rotuladas e limiarizadas, de acordo com uma função nebulosa ("fuzzy") de similaridade. Em seguida são ordenadas e a melhor hipótese é utilizada para calcular as posições dos objetos articulados usando um algoritmo baseado no Filtro de Kalman.

A orientação da normal de superfície é um atributo geométrico utilizado em vários trabalhos de reconhecimento e de modelagem (ou reconstrução) de objetos 3D. Um dos

primeiros trabalhos nessa linha é o de Grimson-Perez[Gri84], onde a técnica de geração de hipóteses consiste em determinar possíveis pares de conjuntos de atributos relativos à posição e orientação, determinados por pontos extraídos da superfície do objeto na cena e encontrados nos modelos. As possíveis interpretações são obtidas usando algumas restrições geométricas, para controlar o potencial combinatorial do espaço de pesquisa. Cada uma das interpretações são testadas (verificação de hipóteses) através das equações de superfície que descrevem os modelos do objeto, os quais determinam o local onde o ponto do objeto deveria estar, através de cálculos de rotação e translação na superfície do modelo. Se o ponto não estiver em nenhuma superfície, a interpretação é considerada inconsistente, e portanto descartada. As restrições geométricas locais que geram o conjunto de interpretações possíveis são formadas pela distância entre dois pontos obtidos na imagem do objeto que devem ser casados com dois pontos no modelo, e o ângulo entre as normais das superfícies dos pontos considerados. Dois pontos encontrados na imagem podem gerar quatro restrições de ângulos obtidos dos produtos escalares: $n_1 \cdot n_2$, $d \cdot n_1$, $d \cdot n_2$, $d \cdot n_{1,2}$, onde d é o vetor que separa os dois pontos cujas normais são n_1 e n_2 e $n_{1,2} = n_1 \times n_2$, o produto vetorial entre os dois vetores normais. Apesar do sucesso encontrado nos testes realizados, o sistema possui alguns pontos negativos: 1) a representação das superfícies dos modelos através de equações matemáticas se torna difícil para superfícies mais complexas (ex. polígono com várias arestas); 2) a geração de hipóteses baseada em pares de pontos, implica em um número elevado de buscas na biblioteca de modelos e por sua vez em um elevado número de combinatorios (interpretações) do espaço de pesquisa; 3) o reconhecimento realizado apenas para um número reduzido (esparso) de pontos na imagem, para evitar extração de ruídos e erros maiores nas interpretações pode deixar de detectar características importantes ou discriminantes do objeto.

Outro trabalho que utiliza quase a mesma estratégia de geração de hipóteses de Grimson-Perez é o de Murray [Mur87]. Ele desenvolveu um sistema de reconhecimento 3D que utiliza um número maior de restrições geométricas como o cálculo do vetor normal das superfícies do modelo obtida dos pontos extraídos do contorno. Isto reduz um pouco o número de interpretações possíveis mas não elimina todos os pontos negativos encontrados no trabalho de Grimson, como a complexidade dos cálculos ou o número

de buscas na biblioteca de modelos.

Yuan [Yua93] apresenta um sistema de modelagem automática de objetos 3D, através da análise de “Cadeias de Vetor de Massa” (MVC, de “*Mass Vector Chains*”) do objeto e suas relações com o modelo do objeto. As orientações dos vetores normais às superfícies são utilizadas para investigar a visibilidade das mesmas. A MVC é uma série de vetores ponderados no qual cada vetor representa uma superfície de remendos, que aponta para a direção normal média da superfície inteira e cujo peso é proporcional ao tamanho da região visualizada (projetada) num plano perpendicular à superfície de remendos. Utilizando conceitos de massa Gaussiana de um objeto, é encontrada uma relação entre a massa e as superfícies. Com isto, consegue-se estimar superfícies e ajustar vistas que não foram processadas, necessárias na modelagem.

Vaillant et al [Vai92] apresentam um método de modelagem de objetos 3D com superfícies curvas através das suas bordas de extremidades (arestas do objeto com o plano de fundo), que não são normalmente descritas por representações de superfícies. É mostrado que o vetor normal unitário de cada ponto na aresta de extremidade pode ser associado com a esfera unitária Gaussiana. Cálculos comprovam a necessidade de, no mínimo, três câmeras em montagem estéreo para determinar a aresta de extremidade e fornecer uma boa estimativa das superfícies na vizinhança.

Klemt et al [Kle95] apresentam um método de reconstrução de superfícies da anatomia humana necessário para identificação de estruturas patológicas, através do uso das normais das superfícies. A reconstrução ocorre em duas unidades, a unidade de produção que faz as tarefas de pré-processamento (segmentação e determinação das faces visíveis nas imagens 2D, Raio-X, CT, etc) e a unidade do usuário, que gera a imagem 3D a partir das informações da primeira unidade e estabelece o ângulo de visualização do objeto (parte humana) pelo profissional de saúde. A visibilidade de uma face plana do objeto na posição de um observador é determinada pelo ângulo (produto escalar) entre o vetor normal unitário da referida face com o ponto onde se encontra o observador. A face é visível ao observador se o produto escalar for maior que zero. São estabelecidas unidades espaciais formadas por superfícies planas (ou curvas, aproximadamente planas) que passam na origem (centro do objeto) e que interseccionam uma esfera unitária. O vetor normal de cada unidade espacial é classificado como visível,

invisível ou duvidoso, o que é suficiente para reconstruir o objeto em 3D, considerando apenas as faces visíveis.

2.6 Conclusões

Muito embora não tenha sido desenvolvido ainda um sistema de reconhecimento para objetos 3D de propósito geral, vários fatores têm contribuído para grandes progressos nas pesquisas realizadas ao longo dos anos. Podem ser citados como causa destes progressos os equipamentos de aquisição e digitalização de imagens que aumentaram a precisão e a velocidade dos dados coletados nos objetos; uso de novos esquemas de representação dos objetos, que usam técnicas de maiores precisões e o uso dos sistemas CAD em 3D para a modelagem, com importantes recursos gráficos.

Com relação aos métodos apresentados podem-se concluir:

1. O sistema de Nevatia e Binford e o de Brooks apresentam maior ênfase no reconhecimento de objetos de superfícies curvas, uma vez que representam os objetos por cilindros generalizados. Estes podem ser úteis, por exemplo para o reconhecimento em aplicações médicas e imagens de satélite. Em termos de velocidade de processamento, estes sistemas devem apresentar baixa velocidade, pois a análise requer vários passos de procedimentos. Em termos de precisão, não conseguem ótimos resultados porque os detalhes são eliminados pela representação utilizada, o que reduz um pouco seu campo de aplicação.
2. O sistema de Oshima e o de Bhanu trabalham com o contorno das superfícies do objeto, extraíndo destes as propriedades para os procedimentos de análise e interpretações. Este enfoque vai interferir na precisão das características do objeto na cena, no que diz respeito às saliências encontradas sobre a superfície do objeto, como por exemplo, cavidades ou rugosidades. Além disto, o método de Bhanu, por fazer rotulagem estocástica de faces, deixa o sistema computacionalmente complexo, o que poderá torná-lo lento. Estes dois métodos, provavelmente não dariam bons resultados em reconhecimento de imagens de satélite, ou imagens médicas. Em algumas aplicações industriais estes sistemas produziram bons

resultados, se não for necessária alta velocidade.

3. O sistema 3DPO, que atua nas arestas das imagens dos objetos, faz segmentação de um modo muito eficiente, tornando-o um método muito preciso e de grande aplicação. No entanto, tal procedimento interfere na velocidade de operação, por possuir procedimentos longos. Também produz bons resultados em aplicações industriais. Seu maior mérito se refere ao fato de fazer reconhecimento de objetos mesmo que estejam empilhados.
4. O sistema de Borges atua em objetos de superfícies curvas com articulações, o qual pode ser muito útil no reconhecimento de partes do corpo humano ou de animais, por exemplo. A maior vantagem é o uso de funções nebulosas para distinguir peças de alto grau de similaridade, o que torna o sistema muito confiável.
5. Os trabalhos que utilizam a modelagem por sistemas CAD conseguem maior precisão no reconhecimento dos objetos. Isto se justifica pelo fato dos sistemas CAD possuírem mais recursos de extração de características do modelo que a segmentação convencional. Os sistemas de Majundar, de Wong e o de Park são os mais rápidos entre os apresentados, porque possuem procedimentos de análise menores, apesar do método de Majundar apresentar ainda o tempo para pesquisa de árvore na etapa de geração de hipóteses. O método de Wong é um pouco mais rápido e preciso mas ainda requer tempo da operação para verificação de isomorfismo de subgrafos. Os três têm aplicações restritas ao reconhecimento de objetos com superfícies planas.
6. O sistema de Park utiliza sua metodologia de geração de dados para reconhecimento de objetos, semelhante à utilizada por seres humanos. Se o sistema não consegue reconhecer o objeto com uma imagem do mesmo, ele faz uma rotação no ponto de visão até conseguir identificar uma superfície. Apesar do artigo não esclarecer como é feita esta rotação com a imagem, se é manual, isto é, o usuário que procura o melhor ângulo, ou se é automática, isto é, através de programação. A maior desvantagem é a necessidade de um grande número de vistas para cada objeto, o que implica em necessidade maior de memória, além

de exigir um tempo de busca maior para pesquisa de casamento. Este tempo é compensado pela redução de cálculos, uma vez que o casamento ocorre entre as próprias vistas.

Os trabalhos citados diferenciam-se na aplicação, no custo, na precisão dos resultados, na velocidade de operação e na flexibilidade do sistema. A situação ideal, que reúne o estado ótimo em todos os itens num único sistema, ainda está longe de ser alcançada. A velocidade de processamento, desde a aquisição da imagem do objeto até a análise e o reconhecimento, deve ser comparável com a da execução da tarefa a ser realizada. Frequentemente, a taxa de "tempo real" é menos do que frações de segundo por operação.

A taxa de reconhecimento dos objetos na cena, associada com a precisão na localização e orientação dos mesmos, deve ser alta para garantir o sucesso do sistema. A flexibilidade do sistema de visão deve ser suficiente para acomodar variações nas dimensões físicas bem como as incertezas na colocação da peça no ambiente da cena. Além disto, o sistema deve distinguir seguramente o objeto na cena das imagens indesejáveis (sujeiras).

Finalmente, o sistema de visão deve ser capaz de lidar com peças complexas, capturar e localizar saliências para poder realizar as operações desejadas.

O capítulo seguinte descreverá a metodologia utilizada para a elaboração da biblioteca de modelos, a primeira etapa do sistema de reconhecimento proposto.

Capítulo 3

Modelagem de Objetos 3D

3.1 Introdução

Este capítulo apresenta a metodologia para construção de modelos de objetos 3D do sistema de reconhecimento desenvolvido. Devido às vantagens que os sistemas CAD oferecem para elaboração de modelos em sistemas de reconhecimento de objetos para aplicações industriais (v. Seção 2.2.1), os modelos dos objetos desta tese foram elaborados num sistema CAD.

A estrutura básica de cada modelo da biblioteca de modelos do sistema é composta de duas partes: *1) a forma 3D e 2) os atributos geométricos 3D dos objetos*. A forma e os atributos geométricos 3D foram elaborados no sistema AutoCAD R12, com base na metodologia apresentada no trabalho de Farias [Far97a]. Os atributos geométricos são extraídos do modelo projetado para complementar o conjunto de informações. Os atributos utilizados pertencem a categoria de atributos de superfícies. A estrutura de um modelo, portanto, pode ser representada pelo diagrama da Figura 3.1.

Foram construídos 10 (dez) modelos de formato poliedral, representando formas comumente encontradas em análise de sistemas de reconhecimento. Para se conseguir a estrutura de modelagem proposta, foram realizadas duas etapas:

1. Elaboração das operações necessárias para o AutoCAD construir as formas do objeto 3D;

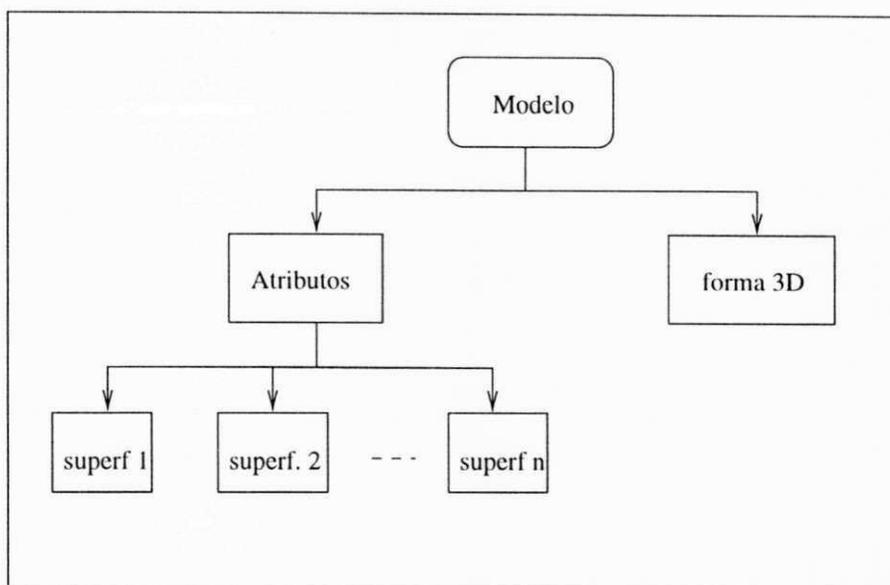


Figura 3.1: Estrutura básica de um modelo.

2. Elaboração de programa com rotinas para extrair os atributos geométricos de cada modelo.

3.2 Construção das Formas 3D

Os modelos foram elaborados levando em consideração que os objetos repousam sobre uma mesa de trabalho, na qual o canto inferior esquerdo do plano da mesa é considerado como sendo a origem do sistema de coordenadas 3D. Foram desenvolvidos no programa chamado Extensão Avançada de Geração de Modelos (AME, *Advanced Modeling Extension*) do AutoCAD R12. No AME, as formas 3D elaboradas em desenhos de linhas (*wire-frames*) podem ser transformadas em sólidos, além de se ter acesso aos atributos geométricos. O AME utiliza como técnicas de modelagem a representação de Geometria de Sólidos Construtivos - CSG (de *Constructive Solid Geometry*) e a representação de bordas - B-rep (de *Boundary*) [Adv92].

Na técnica CSG os modelos compostos podem ser criados a partir de outros modelos compostos ou de primitivas geométricas, combinando-os através de operações booleanas

de união, subtração ou interseção. As primitivas são funções que permitem a geração de sólidos com formas simples, tais como: cubos, cones, cilindros, esferas, etc. Na B-rep um sólido pode ser descrito através de um conjunto de superfícies, as quais são descritas por um conjunto de curvas espaciais cuja conectividade é descrita por grafos relacionais. Esta última representação é muito utilizada quando se descreve o sólido a partir das informações do contorno de suas superfícies [Fly91].

Os comandos para a construção dos sólidos consistem de três etapas:

1. Preparação do ambiente gráfico;
2. Preparação do ambiente do sólido (AME);
3. Construção dos modelos;

As etapas 1 e 2 são iguais para todos os modelos. Elas preparam o ambiente do AutoCAD para a construção do modelo desejado. No ambiente gráfico utilizou-se os comandos para definição dos limites e escalas de visualização. O ambiente do sólido permite estabelecer os parâmetros e variáveis necessários para o desenvolvimento dos modelos, tais como tipo de visualização (*wire*), densidade de malha, sistema de coordenadas (*world*), etc.

A Figura 3.2 mostra a vista 3D dos modelos construídos¹. O modelo 9, por exemplo, foi elaborado com representação B-rep, já o modelo 2 com a representação CSG.

3.3 Atributos Geométricos

Em geral, os sistemas de reconhecimento de objetos requerem estratégias de casamento capazes de estabelecer correspondências entre a descrição parcial do objeto na cena com a descrição completa dos modelos armazenados. Estas correspondências ocorrem, na sua maioria, através de casamento de atributos. As estratégias utilizadas durante a fase do casamento de atributos estão relacionadas ao tipo de atributo disponível e ao tipo de representação deste atributo. Os atributos geométricos são os mais empregados nesta situação.

¹A numeração em cada modelo refere-se ao número da superfície.

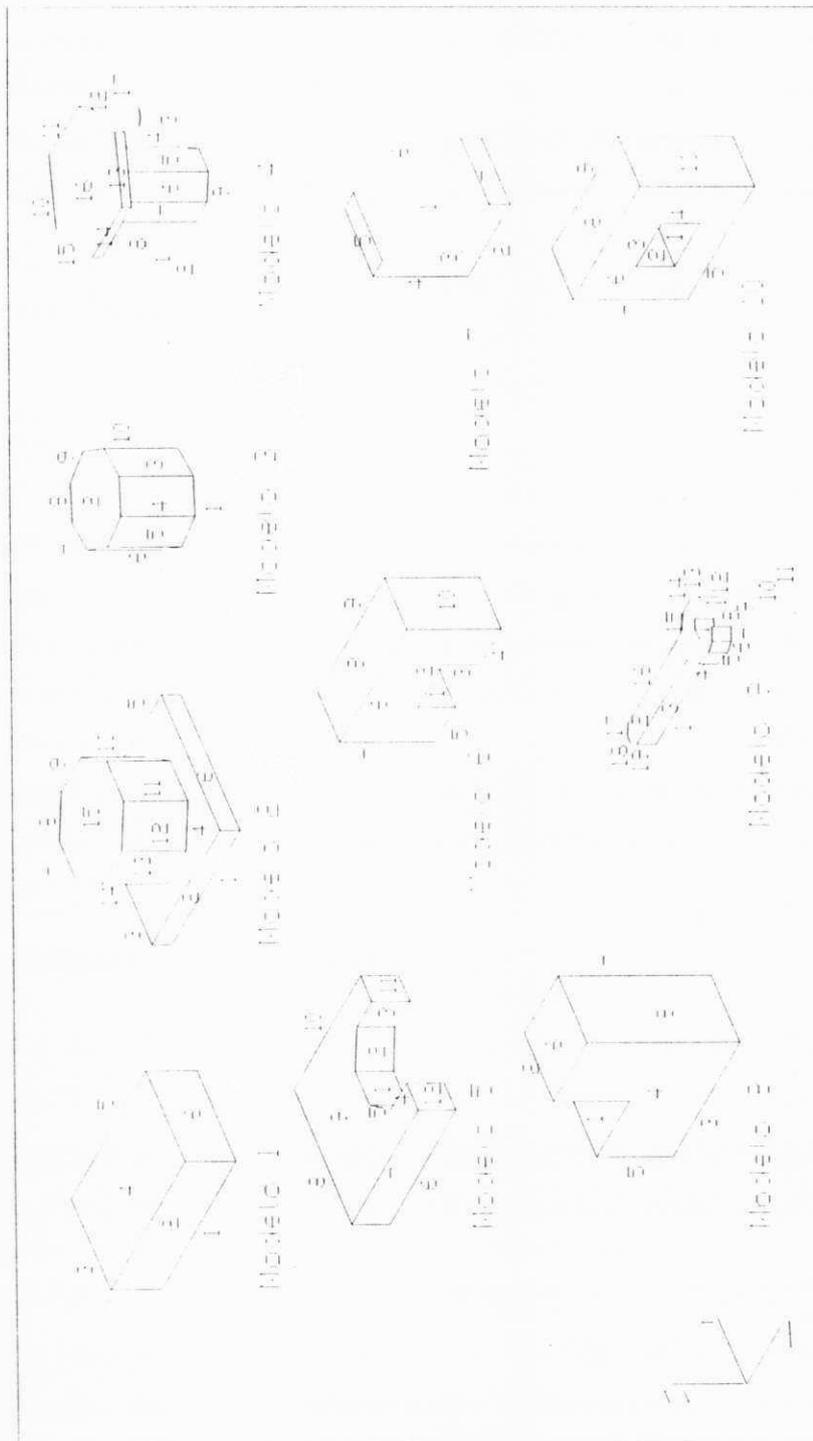


Figura 3.2: Vista 3D dos modelos.

Nesta tese, os atributos geométricos utilizados são os seguintes atributos das superfícies: *perímetro*, *normal*, *centróide*, *número de arestas*, *vértices* e *área*. A partir destes atributos foram calculados alguns atributos do sólido tais como: *área mínima*, *área máxima*, *número de arestas mínimo* e *número de arestas máximo*.

Todos os atributos foram obtidos através de um programa elaborado em linguagem AutoLISP, que roda no AutoCAD. Com o modelo 3D elaborado, o programa extrai ou calcula os atributos citados, confeccionando um banco de dados relativo à cada modelo.

3.4 Conclusões

A biblioteca de modelos conseguiu atingir seus objetivos com êxito. Os modelos contêm informações topológicas e geométricas (forma 3D e atributos), visando sua utilização em sistemas de reconhecimento de objetos 3D baseado em casamento de modelos.

A avaliação da metodologia utilizada é feita com base na avaliação de esquemas de modelagem geométrica indicada por Brown [Bro81], que considera como requisitos para a representação utilizada na modelagem geométrica as propriedades formais e informais. As propriedades formais são: *domínio*, *validade*, *inambigüidade*, *unicidade* e *consistência*; e as propriedades informais: *concisão*, *facilidade de criação* e *eficiência*.

Considerando as representações de CSG e B-rep para os modelos construídos, podem-se concluir, com relação às propriedades formais, que:

1. São elementos do domínio: as primitivas e as operações (CSG), bem como os tipos de faces e de arestas (B-rep).
2. As representações dos modelos são inambíguas pois os elementos do domínio são inambíguos [Bro81].
3. Os modelos B-rep são *únicos* mas os modelos CSG não são únicos. No entanto, isto não caracteriza um problema para sistemas de reconhecimento.
4. Os modelos são válidos e consistentes pois correspondem a objetos reais.

Quanto às propriedades informais pode-se concluir que os modelos são fáceis de serem construídos, modificados, armazenados e transmitidos, além de serem eficien-

tes fontes de dados para visão computacional de peças industriais. Dependendo da aplicação, as formas ou os dados geométricos ou a quantidade de modelos do objeto podem ser ampliados, reduzidos ou modificados.

Foram elaborados apenas modelos de superfícies planas por causa da aplicação imediata, o sistema de reconhecimento de objetos poliedrais, já mencionado. Não há dificuldades na elaboração de modelos ou na extração de atributos para objetos de superfícies curvas, sejam elas esféricas, parabólicas, cilíndricas, etc.

O capítulo que segue apresenta a análise da cena, em técnica multi-vista, desde a aquisição das imagens até a extração de dados para o reconhecimento.

Capítulo 4

Aquisição e Extração de Dados da Cena

4.1 Introdução

Neste capítulo, apresenta-se uma metodologia de aquisição de imagens e extração de dados 3D para uso em sistemas de visão estéreo, a partir de três imagens 2D da cena - Técnica Multi-vista.

Uma possível montagem para utilização desta estratégia é ilustrada na Figura 4.1, onde o objeto é colocado numa mesa de trabalho de onde são tiradas três fotografias de diferentes posições, tal que captem dados correspondentes ao comprimento, largura e altura do objeto. As câmeras devem ser colocadas à mesma distância do centro da mesa e devem ter mesma distância focal para manter a mesma unidade de comprimento com as imagens do objeto [Mar92].

A mesa serve como referência para o sistema de coordenadas da cena. O canto inferior esquerdo é considerado a origem do sistema de coordenadas. As câmeras devem ser colocadas na cena em planos ortogonais, ou seja, uma câmera é colocada em paralelo ao plano da mesa e acima dela de modo a captar a vista superior do objeto, o que corresponde ao plano $X-Y$ (largura-comprimento) da cena. A segunda câmera é colocada ao lado da mesa de modo a obter a vista lateral do objeto, o que corresponde ao plano $X-Z$ (largura-altura) da cena. A terceira câmera é colocada no outro lado da mesa de

modo a captar a vista frontal correspondendo ao plano $Y-Z$ (comprimento-altura) da cena.

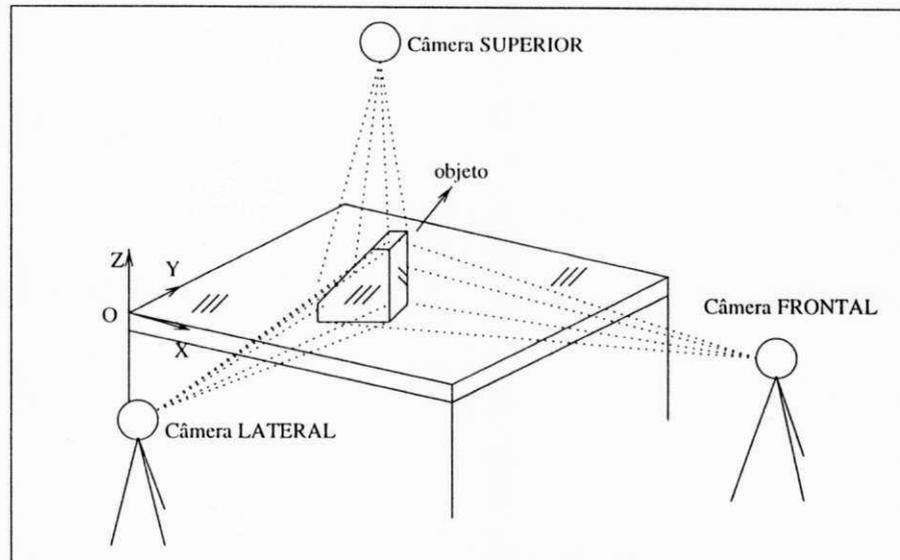


Figura 4.1: Estrutura de aquisição das imagens.

Na Figura 4.2 apresenta-se o diagrama em blocos do método proposto. A imagem captada por cada câmera é enviada para a etapa de digitalização que inclui o pré-processamento e a segmentação de imagens. Nesta etapa é realizada a extração de dados (vértices) de superfícies das vistas 2D das cenas. Os dados são reunidos, formando um banco de dados 2D, para a análise das regiões de cada vista. Através de técnica multivista, que faz cruzamento (casamento) de coordenadas dos vértices das superfícies de cada vista, identificam-se as coordenadas dos vértices 3D das superfícies do objeto ou objetos na cena. Em seguida, com os dados 3D encontrados, calculam-se os demais atributos de superfície que sejam necessários ao processo de reconhecimento. As superfícies que não puderem ser descritas em 3D são descartadas.

No final, faz-se uma análise dos atributos das superfícies visíveis das vistas eliminando as duplicidades, ou seja, eliminando as superfícies visualizadas por mais de uma vista. O relatório de saída fornecido pelo sistema, contém, portanto, os vértices e os atributos 3D das superfícies visíveis.

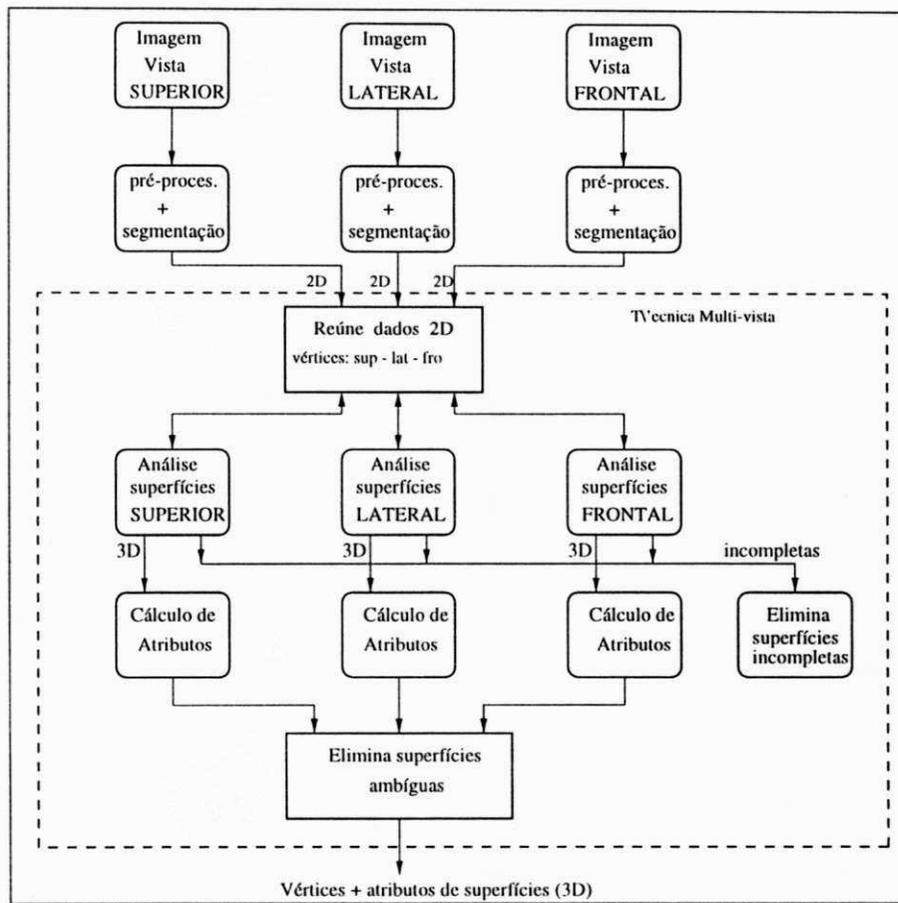


Figura 4.2: Diagrama em blocos de análise da cena.

Para que as imagens sejam processadas e analisadas adequadamente no computador alguns aspectos da captura da imagem e calibração da câmera devem ser levados em consideração. Estes aspectos dizem respeito às características da câmera utilizada na obtenção da imagem e na sua digitalização [Mar92]. A Seção 4.2 descreverá as transformações geométricas das imagens no sistema de visão proposto. As seções seguintes descreverão as etapas de pré-processamento, segmentação e cruzamento de dados.

4.2 Transformações Geométricas da Câmera

Os procedimentos utilizados para encontrar os parâmetros reais presentes nas imagens do objeto são conhecidos como transformações geométricas da câmera. Tais transformações envolvem as características da câmera utilizada para a captura e digitalização da imagem. A escolha destes elementos depende da aplicação específica do sistema de visão [Mar92, Liu90a, Che89, Dho89, Chi92a].

Em geral, as características da câmera envolvem parâmetros internos do aparelho. Os parâmetros internos mais utilizados em câmeras comuns ou digitais são o comprimento focal e a distorção de lentes. A relação entre as coordenadas reais do objeto e da imagem formada no interior da câmera é considerada como parâmetro externo.

A distorção causada pelo uso de lentes pode deteriorar a qualidade da imagem e portanto, realizar medidas errôneas do objeto. Tal distorção ocorre devido ao raio de luz não emergir das lentes nas direções desejadas. É um parâmetro importante quando o sistema procura calcular a medida de distância do objeto ao centro da projeção da câmera. O comprimento focal indica a distância entre o plano da imagem e o centro de projeção da câmera.

A digitalização da imagem envolve a amostragem da intensidade da luz na imagem, medida através de grade espaçada regularmente nas direções horizontal e vertical.

Considerando apenas câmeras fotográficas comuns, cujas fotografias são digitalizadas em equipamentos *scanners*, tem-se:

f = comprimento focal da câmera,

d = distância do objeto à câmera ,

x_n = comprimento no plano de projeção (negativa),

X_f = distância medida na fotografia,

X_r = distância real medida na cena,

A = fator de ampliação entre a fotografia e a negativa,

$t = np/[comp]$ = taxa de amostragem da fotografia no *scanner*,

np = número de *pixels*,

$[comp]$ = unidade de comprimento da fotografia.

A relação entre a distância real e a distância entre um certo número de *pixels* na imagem processada, é obtida através dos seguintes passos:

1. A relação entre o comprimento do objeto e o comprimento da imagem negativa é obtida através da análise geométrica na Figura 4.3, onde:

$$\frac{f}{d} = \frac{x_n}{x_n + X_r}$$

$$\Rightarrow X_r = \left(\frac{d-f}{f} \right) x_n. \quad (4.1)$$

2. A relação entre o comprimento na negativa e o comprimento na fotografia depende do fator de ampliação, ou seja:

$$x_n = \frac{X_f}{A}. \quad (4.2)$$

3. A relação entre o comprimento na fotografia e o comprimento na imagem digitalizada depende da taxa de amostragem, a qual representa o número de *pixels* por unidade de comprimento da fotografia, ou seja:

$$X_f = \frac{np}{t}. \quad (4.3)$$

Portanto, o comprimento na imagem real pode ser obtido substituindo as equações 4.3 em 4.2 e 4.2 em 4.1, ou seja:

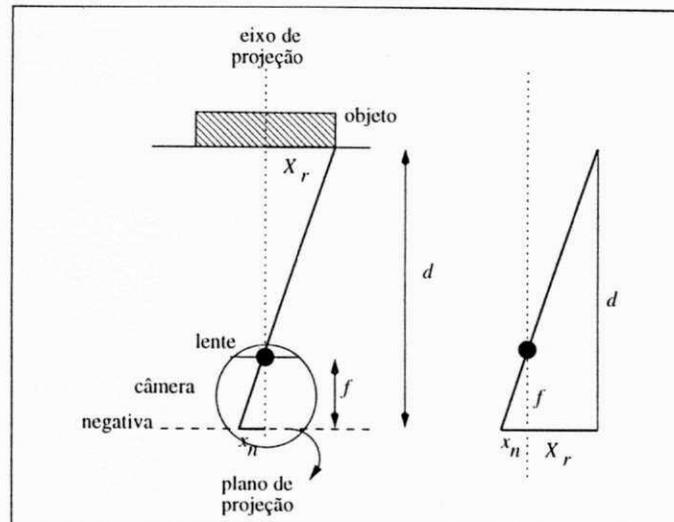


Figura 4.3: Relação entre comprimento real e projetado na negativa

$$X_r = \left(\frac{d - f}{fAt} \right) np. \quad (4.4)$$

Em resumo, sabendo-se o número de *pixels* da imagem processada, tem-se o comprimento real correspondente. Pode-se dizer que a transformada geométrica é o termo Tg , dado pela equação 4.5:

$$Tg = \frac{d - f}{fAt}. \quad (4.5)$$

4.3 Pré-processamento e Segmentação das Imagens

Em cada imagem proveniente da cena são aplicadas técnicas comuns de pré-processamento e segmentação. A função do pré-processamento é reduzir cada uma das vistas a uma imagem de linhas, onde apenas os contornos das superfícies são visíveis. Estes contornos devem ser contínuos e bem definidos, de modo a garantir uma boa extração de regiões (superfícies) na etapa de segmentação [Gon93]. A segmentação fornece uma lista de vértices 2D que definem o contorno das superfícies de cada imagem.

Nas etapas de pré-processamento e segmentação foram utilizados algoritmos já desenvolvidos por alunos e professores do LAPS/UFPB¹ [Mel96a, Mel97, Mel96b], como parte de um projeto na área de Processamento Digital de Imagens. Este projeto visa a implementação em *hardware* para processamento de imagens em tempo real. Tais algoritmos foram aplicados com sucessos às imagens dos testes realizados neste trabalho, com exceção do algoritmo para extração de vértices (segmentação) que ainda necessita alguns ajustes para incluir os padrões de vértices formados por três arestas [Vil97].

O algoritmo de detecção de bordas citado consiste de três etapas:

- obtenção da imagem gradiente - *Spline*,
- cálculo do máximo local e obtenção do primeiro contorno do objeto - Caminho I,
- obtenção do segundo e último contorno - Caminho II.

A imagem gradiente é obtida através da aplicação da algoritmo *Spline*, cuja função é definida por quatro máscaras 5×5 . Os parâmetros destas máscaras correspondem aos coeficientes da derivada das *splines* que interpolam o *pixel* central nas direções horizontal, vertical e diagonais durante convolução com a imagem de entrada. A média dos valores obtidos em cada uma destas direções corresponde ao gradiente da imagem na posição do *pixel* central. Este procedimento garante um operador muito resistente à ruídos na imagem. O uso desta função é justificado porque a interpolação produzida por ela ajusta-se suavemente ao contorno da imagem, reproduzindo sem erros os pontos de amostragem originais [Mel96a, Mel97].

Uma função ponderada é utilizada para definir os caminhos mais prováveis do contorno. Esta função baseia-se numa análise do *pixel* considerando uma janela 5×5 que calcula o peso (nível de cinza) do *pixel* central em relação aos *pixels* vizinhos menores ou iguais a este. Em seguida, os caminhos encontrados são checados por janelas 3×3 para verificação de possíveis caminhos padrões visando garantir a continuidade da borda.

O segundo contorno é utilizado baseado numa soma ponderada dos caminhos do primeiro contorno o qual garante um caminho contínuo formado por apenas um *pixel*.

¹Laboratório de Processamento de Sinais da Universidade Federal da Paraíba.

O algoritmo de segmentação consiste na obtenção dos vértices das regiões fechadas (superfícies) das imagens. O algoritmo desenvolvido por Vilar [Vil97] determina os pontos de vértices, a partir da imagem de contorno (Caminho II), comparando-se cada janela (5×5) de contorno com janelas de vértices previamente padronizados. Se a comparação for positiva, analisa-se a qualidade deste vértice baseado na variação angular total dos segmentos que formam o vértice encontrado.

4.4 Técnica multi-vista

A técnica multi-vista tem a finalidade de fazer o cruzamento dos pontos de vértices 2D nas três imagens e determinar sua descrição em 3D. Inicialmente, é necessária a seguinte definição:

Definição 4.4.1 *Sejam k_s, k_l e k_f o número de superfícies das vistas superior, lateral e frontal respectivamente.*

Sejam W_{ls}, W_{ml}, W_{nf} o número total de vértices da l -ésima superfície da vista superior, m -ésima superfície da vista lateral e n -ésima superfície da vista frontal, respectivamente.

Os vértices das superfícies de cada vista são definidos por:

$$V_{lp} = (x_{lp}, y_{lp}), \quad (4.6)$$

$$V_{mq} = (x_{mq}, z_{mq}), \quad (4.7)$$

$$V_{nr} = (y_{nr}, z_{nr}), \quad (4.8)$$

onde,

V_{lp} = p -ésimo vértice da l -ésima superfície da vista superior,

V_{mq} = q -ésimo vértice da m -ésima superfície da vista lateral,

V_{nr} = r -ésimo vértice da n -ésima superfície da vista frontal

e,

$$1 < p \leq W_{ls}, \quad 1 < q \leq W_{ml}, \quad 1 < r \leq W_{nf}$$

$$1 < l \leq k_s, \quad 1 < m \leq k_l, \quad 1 < n \leq k_f.$$

4.4.1 Análise das Vistas

Considere a situação ilustrada na Figura 4.4 - um vértice visível nas três vistas. Na vista superior tem-se o vértice V_{lp} com coordenadas (x_{lp}, y_{lp}) e espera-se encontrar a coordenada z_{lp} nos vértices correspondentes nas demais vistas: V_{mq} e V_{nr} . Para isto, faz-se uma pesquisa nas vistas lateral e frontal, tentando casar x_{lp} com x_{mq} e y_{lp} com y_{nr} . Ou seja, espera-se encontrar um vértice que seja visível na vista superior e numa das vistas lateral ou frontal.

A Figura 4.5 mostra um diagrama em blocos desta análise. O casamento entre as coordenadas nas pesquisas às vistas lateral e frontal ocorre, respectivamente, quando são satisfeitos os seguintes critérios de erro:

$$|x_{lp} - x_{mq}| \leq Erro \quad (4.9)$$

$$|y_{lp} - y_{nr}| \leq Erro. \quad (4.10)$$

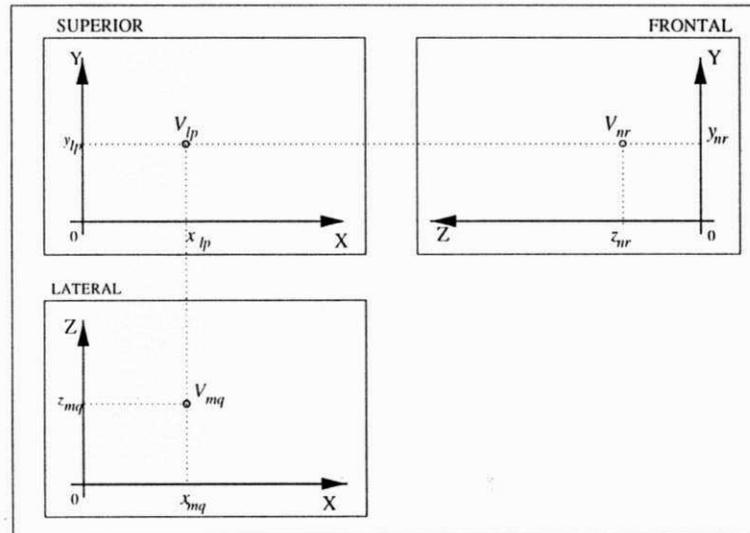


Figura 4.4: Análise Multi-vista.

Para cada casamento encontrado, armazena-se o valor do z correspondente, z_{mq} e/ou z_{nr} . É possível encontrar mais de uma coordenada z . Como a vista superior

capta os pontos de maior altura do objeto, escolhe-se a coordenada de maior valor entre elas. Ou seja,

$$z_{lp} = \text{Maior} < z_{mq}, z_{nr} > .$$

Observe que se a pesquisa na vista lateral não tiver sucesso, existe ainda a possibilidade de sucesso na pesquisa da vista frontal. Caso as duas não tenham sucessos, as superfícies são consideradas incompletas e serão rejeitadas.

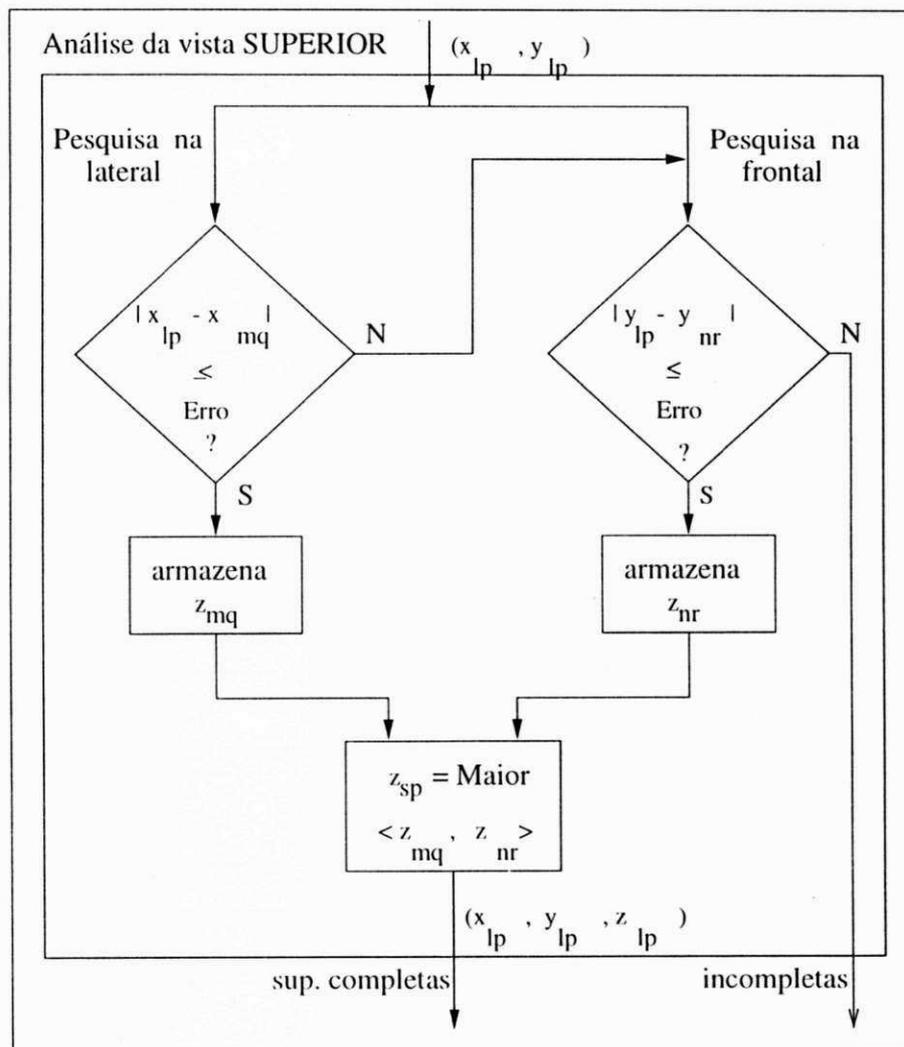


Figura 4.5: Diagrama da análise da vista superior.

Na análise da vista lateral, faz-se apenas a pesquisa na vista superior (Figura 4.6). Tem-se o vértice V_{mq} com coordenadas (z_{mq}, x_{mq}) e deseja-se encontrar a coordenada y_{mq} . Através da pesquisa na vista superior, investiga-se o casamento entre as coordenadas x_{mq} e x_{lp} , considerando o critério de erro:

$$|x_{mq} - x_{lp}| \leq Erro. \quad (4.11)$$

Armazena-se o valor da coordenada y_{lp} correspondente, para cada casamento encontrado. É possível encontrar mais de uma coordenada y na correspondência. Neste caso, escolhe-se o menor valor entre eles porque estará mais próximo da câmera e, portanto, com maior possibilidade de estar visível. Ou seja,

$$y_{mq} = Menor < y_{lp} > .$$

Do mesmo modo como na análise da vista superior, as superfícies com vértices incompletos são rejeitadas.

Na análise da vista frontal, também faz-se a pesquisa apenas na vista superior (Figura 4.7). Tem-se o vértice V_{nr} com coordenadas (z_{nr}, y_{nr}) e espera-se encontrar a coordenada x_{nr} . Através da pesquisa, verifica-se o casamento entre as coordenadas y_{nr} e y_{lp} , de modo a satisfazer o critério de erro:

$$|y_{nr} - y_{lp}| \leq Erro. \quad (4.12)$$

Armazena-se o valor de x_{nr} correspondente, para cada casamento encontrado. Ocorrendo mais de um casamento na correspondência, escolhe-se o maior valor de x , ou seja,

$$x_{nr} = Maior < x_{lp} > .$$

São rejeitadas as superfícies com vértices incompletos.

Estes procedimentos são repetidos até que todos os vértices de cada superfície visível sejam investigados.

O limiar de erro *Erro*, nas equações 4.12, 4.11, 4.9 e 4.10, corresponde a um erro fixo medido em unidades de comprimento é empírico e serve para deixar uma margem de segurança aos erros que podem ocorrer no alinhamento das imagens e extração dos

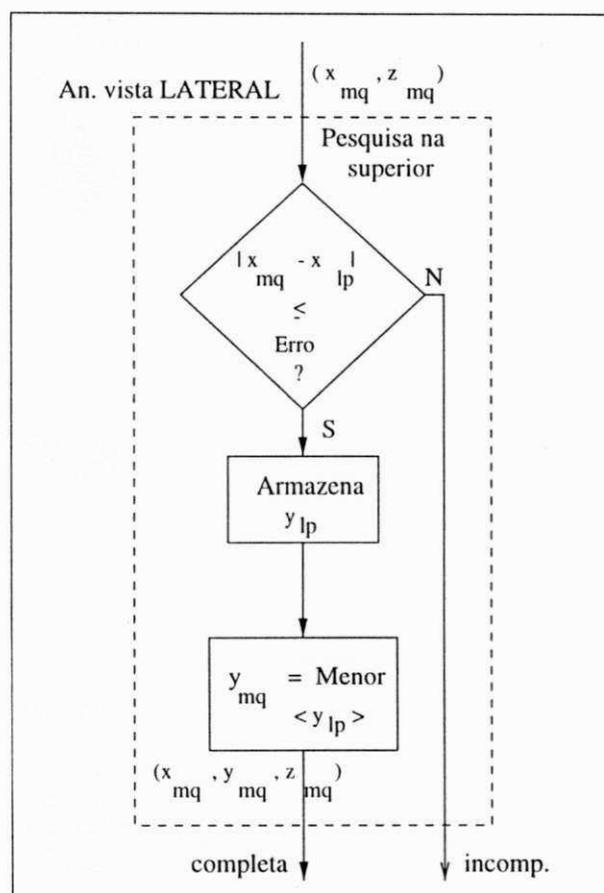


Figura 4.6: Diagrama da análise da vista lateral.

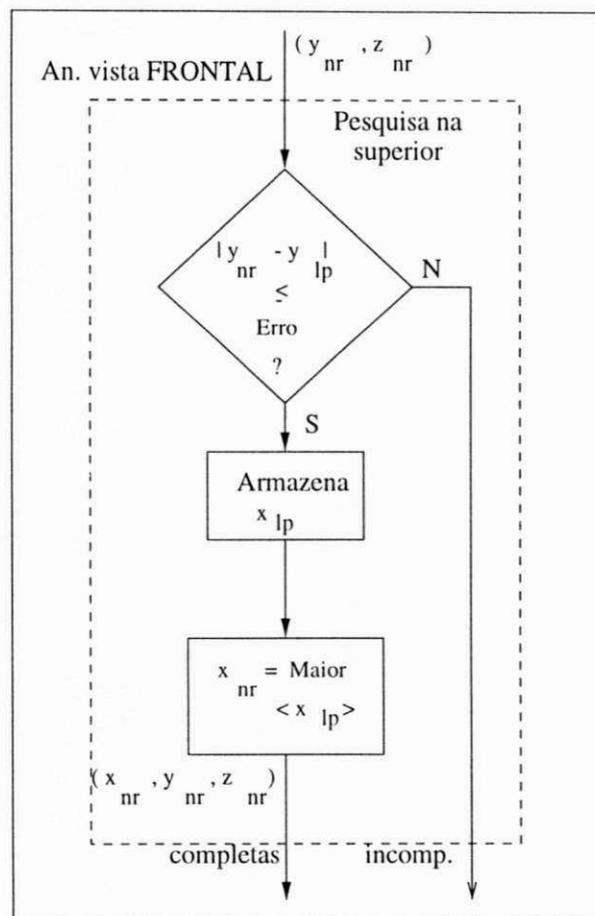


Figura 4.7: Diagrama da análise da vista frontal.

vértices. Este valor é encontrado baseado na menor distância entre duas arestas no objeto ou na menor distância entre dois objetos.

As superfícies consideradas incompletas em 3D na vista superior são aquelas que não são visíveis nas vistas lateral ou frontal, ou seja, não obedecem aos critérios de erro citados (equações 4.9 e 4.10) e portanto, não encontram suas correspondências. Isto pode ocorrer tanto por causa da própria posição que o objeto ocupa na cena que oculta o vértice correspondente, quanto a particularidades existentes no próprio objeto. É o caso, por exemplo, de objetos com superfícies contendo um buraco cujos vértices não são visualizados na vista lateral ou frontal. Ou ainda, de superfícies com alguma saliência que oculta alguns vértices de sua parte posterior, dependendo do ângulo de visão e se o objeto não estiver inclinado. Podem ocorrer também casos com objetos superpostos onde um vértice de um dos objetos na vista superior pode alinhar-se acidentalmente com um vértice de outro objeto na vista lateral e/ou frontal. Estas situações também podem ocorrer nas vistas lateral ou frontal com relação a vista superior. Por exemplo, um buraco em uma superfície visto na lateral ou frontal não irá ser visualizado na vista superior. Portanto, o método deve ser aplicado a objetos com superfícies planas, especificamente, poliedros.

Observa-se nesta técnica que a vista superior tem predominância sobre as demais. Isto significa que é a vista superior que orienta a descrição em 3D das superfícies. Se um vértice for visualizado na lateral ou frontal e não tiver sido visualizado na superior, esta superfície não poderá ser descrita em 3D.

4.4.2 Cálculos dos atributos de superfícies

Com as superfícies completamente descritas por seus vértices 3D, é possível calcular alguns atributos geométricos de superfícies, necessários ao processo de reconhecimento de objetos.

Os atributos a serem utilizados são: *área*, *perímetro*, *normal*, *número de vértices* e *centróide*. Todos estes foram calculados a partir dos vértices 3D de superfície encontrados. Procurou-se utilizar atributos extraídos dos vértices das superfícies em virtude de serem comumente utilizados em trabalhos de reconhecimento de objetos polidrais.

Para o cálculo dos atributos, formulados pelos lemas 4.4.1 a 4.4.5, é necessário fazer a seguinte definição:

Definição 4.4.2 *Seja um polígono (superfície k do objeto) de n vértices, p_1, p_2, \dots, p_n , com $n \geq 3$, ordenados no sentido anti-horário e definidos no \mathbb{R}^3 , tal que $p_i = (x_i, y_i, z_i)$.*

Lema 4.4.1 *A área a da k -ésima superfície é dada pela metade da norma do produto vetorial dos seus vértices² [Fig91]. Ou seja,*

$$a_k = \frac{1}{2} \left\| \sum_{j=1}^n \vec{p}_j \times \vec{p}_{j+1} \right\| \quad (4.13)$$

desenvolvendo os produtos vetoriais³, tem-se:

$$a_k = \frac{1}{2} \sqrt{x_k^2 + y_k^2 + z_k^2} \quad (4.14)$$

onde,

$$\begin{aligned} x_k &= \sum_{j=1}^n (y_j z_{j+1} - y_{j+1} z_j) \\ y_k &= \sum_{j=1}^n (z_{j+1} x_j - z_j x_{j+1}) \\ z_k &= \sum_{j=1}^n (x_j y_{j+1} - x_{j+1} y_j). \end{aligned}$$

Lema 4.4.2 *Seja d_i o comprimento da aresta ligando o i -ésimo e o $i+1$ -ésimo vértices da superfície k . O perímetro Pm_k é dado pela soma de todos os comprimentos das arestas da superfície, ou seja,*

$$p_k = \sum_{i=1}^n d_i \quad (4.15)$$

onde,

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}. \quad (4.16)$$

²O vértice-vetor, \vec{p}_i , é considerado como sendo um vetor que parte da origem até o ponto p_i .

³Quando $j = n \Rightarrow j + 1 = 1$.

Lema 4.4.3 *Sejam três vértices, p_1, p_2, p_3 da k -ésima superfície. A normal unitária \vec{n}_k é dada pela razão entre a normal \vec{n}_v da superfície e sua norma. A normal \vec{n}_v é dada pelo produto vetorial das arestas formadas pelos vértices com coincidência em p_2 . Ou seja,*

$$\vec{n}_v = (\vec{p}_3 - \vec{p}_2) \times (\vec{p}_1 - \vec{p}_2) \quad (4.17)$$

e,

$$\vec{n}_k = \frac{\vec{n}_v}{\|\vec{n}_v\|}. \quad (4.18)$$

Lema 4.4.4 *O número de arestas da k -ésima superfície, nl_k , é igual ao número de vértices n . Ou seja,*

$$nl_k = n. \quad (4.19)$$

Lema 4.4.5 *O centróide da k -ésima superfície, \vec{c}_k , é dado pela média aritmética dos vértices. Ou seja,*

$$\vec{c}_k = \frac{\sum_{j=1}^n \vec{p}_j}{Nl_k}. \quad (4.20)$$

4.4.3 Eliminação das superfícies duplicadas

Nos sistemas multi-vistas é importante identificar quais as superfícies visíveis por mais de uma câmera para evitar ambigüidades na sua resposta. Deve-se, portanto, elaborar uma maneira de identificar estas superfícies e eliminá-las do resultado final do processo.

A eliminação das superfícies duplicadas, isto é, que já foram descritas por outra vista, ocorre através do casamento de atributos entre as superfícies. Para isto, é necessário analisá-las numa certa ordem de prioridade: primeiro são analisadas as superfícies da vista superior, depois as da vista lateral e por último as da vista frontal. São eliminadas as superfícies da vista lateral que já tiverem sido descritas na vista superior. São eliminadas as superfícies na vista frontal que já tiverem sido descritas na vista superior ou na vista lateral ou em ambas. Portanto, as superfícies da vista superior não são eliminadas por duplicidades. O lema 4.4.6 propõe os critérios de erro para identificar as duplicidades.

Lema 4.4.6 *Seja a k -ésima superfície de uma vista (lateral ou frontal) a ser analisada para detecção de duplicidade.*

Seja a t -ésima superfície já descrita por uma vista de maior prioridade. Elimina-se a k -ésima superfície se ocorrerem simultaneamente os seguintes critérios de erro:

$$|a_k - a_t| \leq 0.20a_k, \quad (4.21)$$

$$|p_k - p_t| \leq 0.20p_k, \quad (4.22)$$

$$|nl_k - nl_t| \leq 0.20nl_k, \quad (4.23)$$

$$|\vec{n}_k - \vec{n}_t| \leq 0.20\vec{n}_k, \quad (4.24)$$

$$|\vec{c}_k - \vec{c}_t| \leq 0.20\vec{c}_k. \quad (4.25)$$

O limiar de 0.20 permite uma margem de 20% entre a diferença dos referidos atributos com relação ao atributo da superfície em análise. Se todos os vértices de ambas superfícies forem corretamente descritos em 3D, a superfície em análise será corretamente eliminada.

Após a aquisição das imagens da cena, os atributos extraídos devem ser utilizados na identificação, conforme será relatado no capítulo seguinte.

Capítulo 5

Método de Identificação e Localização

5.1 Introdução

O método de reconhecimento consiste de duas etapas:

1. *identificação de objetos,*
2. *localização do objeto identificado.*

Na identificação, distinguem-se duas fases: *geração de hipóteses* e *verificação de hipóteses*. As hipóteses são geradas independentemente uma das outras, através do casamento entre atributos geométricos de superfícies, tais como: área, perímetro e outros. A verificação das hipóteses é realizada em duas etapas. Na primeira etapa são selecionados, a partir das hipóteses geradas, pares de superfícies do modelo tal que o ângulo entre suas normais e a distância entre seus centróides casem com o ângulo entre as normais e distância entre centróides de pares de superfícies descritas em 3D da cena, respectivamente. Na segunda fase, selecionam-se modelos que englobem conjuntos de pares de superfícies combinados e constrói-se uma árvore de pesquisa por modelo para selecionar a melhor combinação dentre as superfícies de cada modelo [Far97b].

Na localização, a superfície identificada é reconstruída no ambiente da cena através de transformações geométricas espaciais.

Para reduzir o espaço de pesquisa na geração de hipóteses, cada modelo foi agrupado de acordo com as faixas de variações da área e do número de lados (arestas) de suas superfícies. Assim, um modelo só será pesquisado se a área e o número de lados da face da cena se enquadrarem nestas variações. Este procedimento foi adotado na tese de Cortez [Cor96] na identificação de objetos em 2D. Outro agrupamento é realizado na primeira fase da verificação de hipóteses, os pares de superfícies são selecionados apenas dentro de cada modelo.

Cada casamento é realizado através de critérios de erro pré-estabelecidos para cada situação. Para a geração de hipóteses, pode-se dizer que o casamento ocorre basicamente entre atributos geométricos de superfícies e na verificação, o casamento ocorre entre os atributos relacionais de superfícies. Ou seja, a primeira fase identifica a superfície e a segunda identifica a sua posição em relação às outras.

Uma vez que os sistemas de coordenadas da cena e dos modelos são diferentes, as faces do modelo identificado devem sofrer transformações geométricas em 3D para que possam ser devidamente localizadas na cena.

Para os testes realizados foram sintetizadas várias cenas com auxílio do AutoCAD, que forneceu as imagens de linhas correspondentes às vistas superior, frontal e lateral do objeto ou objetos na cena. Um programa em AutoLISP aplicado às cenas forneceu a lista dos vértices 3D das superfícies visíveis em mais de uma vista. A partir dos vértices 3D, calculou-se os demais atributos de superfícies da cena, necessários ao casamento: *área, perímetro, quantidade de vértices (ou arestas), comprimento das arestas, centróide e normal unitária.*

Testes com imagens reais também foram realizados utilizando câmeras fotográficas comuns e digitalizadas em equipamento *scanner*, bem como uma câmera digital.

As cenas foram elaboradas com os objetos em posições diferentes das utilizadas para a construção dos modelos correspondentes.

5.2 Características dos Objetos

As características (atributos) dos modelos e das cenas são formuladas através das seguintes definições:

Definição 5.2.1 *Seja o conjunto de modelos*

$$\mathcal{M} = \{M_1, M_2, \dots, M_m\}.$$

Os atributos do modelo M_k , onde $M_k \in \mathcal{M}$, são:

- *sólido k : quantidade de superfícies (faces) (NF_k), área mínima ($AMIN_k$) e área máxima ($AMAX_k$) das faces, número mínimo de arestas ($NLMIN_k$) e número máximo de arestas ($NLMAX_k$) das faces.*
- *superfície S : área (A_{kS}), perímetro (P_{kS}), normal (\vec{N}_{kS}), centróide (C_{kS}), número de arestas (NL_{kS}) e o conjunto de vértices (V_{kS}).*

Definição 5.2.2 *Os atributos da cena com nf faces visíveis são:*

- *superfície s : área (a_s), perímetro (p_s) normal (\vec{n}_s), centróide (c_s), número de arestas (nl_s) e o conjunto de todos os vértices (VC_s).*

5.3 Identificação

5.3.1 Geração de hipóteses

A geração de hipóteses é inicializada após a constatação do seguinte lema:

Lema 5.3.1 *O modelo k é analisado se a área e o número de arestas da i -ésima superfície da cena estiverem contidos nas faixas de variações da área e do número de arestas do modelo, ou seja:*

$$AMIN_k \leq a_i \leq AMAX_k \quad (5.1)$$

$$NLMIN_k \leq nl_i \leq NLMAX_k \quad (5.2)$$

Este lema é importante porque evita a pesquisa à modelos com atributos fora destas faixas de valores, ou seja, reduz o tempo de pesquisa. Caso seja satisfeito, cada superfície do modelo k é analisada. Assim, a I -ésima superfície do modelo k é selecionada se houver simultaneamente o casamento entre:

1. a área da i -ésima superfície da cena (a_i) com a área da I -ésima superfície do modelo k (A_{kI}), com $1 \leq i \leq nf$ e $1 \leq I \leq NF_k$, através do seguinte critério de erro ea :

$$ea = \frac{|a_i - A_{kI}|}{A_{kI}} \leq 0.20; \quad (5.3)$$

2. o perímetro da i -ésima superfície da cena (p_i) com o perímetro da I -ésima superfície do modelo k (P_{kI}), através do seguinte critério de erro ep :

$$ep = \frac{|p_i - P_{kI}|}{P_{kI}} \leq 0.20; \quad (5.4)$$

3. o número de arestas da i -ésima superfície da cena (nl_i) com o número de arestas da I -ésima superfície do modelo k (NL_{kI}), através do seguinte critério de erro enl :

$$enl = \frac{|nl_i - NL_{kI}|}{NL_{kI}} \leq 0.20; \quad (5.5)$$

4. o comprimento de uma aresta da i -ésima superfície da cena ($comp_{c_i}$) com o comprimento de alguma aresta da I -ésima superfície do modelo k ($COMP_{kI}$), através do critério de erro el :

$$el = \frac{|comp_{c_i} - COMP_{kI}|}{COMP_{kI}} \leq 0.20. \quad (5.6)$$

Estas quatro restrições somadas às iniciais (eqs 5.1 e 5.2) geram todas as hipóteses da identificação. O espaço de pesquisa, comparado ao espaço total, é bem reduzido e bastante confiável. O limiar fixo de 0.20 nos critérios de erro é definido empiricamente e corresponde a um erro (ou diferença) máximo entre os atributos da cena e modelo de vinte por cento (20%) do atributo do modelo. Ao final desta fase, para cada superfície i da cena é selecionado um conjunto contendo os modelos k_q e as respectivas superfícies I_{k_q} casadas, ou seja:

$$V_i = \{(k_q; I_{k_q S_1}, I_{k_q S_2}, \dots)\}, q = 1, 2, \dots \quad (5.7)$$

A geração de hipóteses encontra as superfícies casadas isoladamente, ou seja, realiza o casamento de atributos locais e globais das superfícies independentemente uma

das outras. A fase seguinte, a verificação de hipóteses caracteriza-se por realizar casamento com atributos relacionais das superfícies, visando selecioná-las de acordo com sua posição espacial no sólido. A verificação procura agrupar as superfícies hipotetizadas, tal que as relações entre elas sejam as mesmas relações encontradas entre as superfícies da cena.

Os atributos geométricos mais indicados para determinar as relações entre as superfícies de sólidos são *a normal* [Gri84, Mur87, Kle95] e *o centróide* [Osh83]. Estes valores são definidos em relação ao sistema de coordenadas onde encontra-se o objeto. Como os sistemas de coordenadas da cena e do modelo são diferentes, utilizam-se como atributos relacionais *o ângulo entre as normais e a distância entre os centróides* de duas superfícies do objeto. Na realidade, o ângulo entre as duas normais representa o ângulo entre as duas superfícies correspondentes e a distância entre dois centróides representa a distância média entre as duas superfícies correspondentes.

5.3.2 Verificação de hipóteses

A verificação das hipóteses ocorre em duas fases. Na primeira fase as superfícies de cada modelo hipotetizado são agrupadas em pares, tal que o ângulo entre as normais e a distância entre os centróides das componentes de cada par casem com o ângulo entre as normais e a distância entre os centróides das componentes de um par de superfícies na cena. Considerando que a cena pode conter um ou mais objetos, são feitas todas as combinações possíveis de pares de superfícies para esta análise, com exceção da combinação que envolve a superfície com ela própria. Os pares de superfícies do modelo são selecionados de acordo com os critérios de erro estabelecidos nos lemas 5.3.2 e 5.3.3.

Lema 5.3.2 *Sejam θ_{ij} e θ_{kIJ} o ângulo entre as normais de um par de superfícies da cena (\vec{n}_i, \vec{n}_j) e um par do modelo k ($\vec{N}_{kI}, \vec{N}_{kJ}$), respectivamente. Seja $e\theta$ o erro angular dado pela diferença absoluta entre os cossenos destes ângulos (produtos escalares). O critério de erro será dado por*

$$e\theta = |\cos \theta_{ij} - \cos \theta_{kIJ}| \leq 0.28 \quad (5.8)$$

onde:

$$\begin{aligned}\cos \theta_{ij} &= \vec{n}_i \cdot \vec{n}_j \\ \cos \theta_{kIJ} &= \vec{N}_{kI} \cdot \vec{N}_{kJ}.\end{aligned}$$

Lema 5.3.3 *Sejam $dist_{ij}$ e $dist_{kIJ}$ as distâncias entre os centróides¹ de duas superfícies da cena (c_i, c_j) e duas do modelo (C_{kI}, C_{kJ}), respectivamente. Seja ed o erro da distância dado pela diferença absoluta entre as distâncias dos centróides relativa a $dist_{kIJ}$. O critério de erro será dado por*

$$ed = \frac{|dist_{ij} - dist_{kIJ}|}{dist_{kIJ}} \leq 0.10 \quad (5.9)$$

onde,

$$\begin{aligned}dist_{ij} &= \sqrt{(\bar{x}_i - \bar{x}_j)^2 + (\bar{y}_i - \bar{y}_j)^2 + (\bar{z}_i - \bar{z}_j)^2} \\ dist_{kIJ} &= \sqrt{(\bar{x}_{kI} - \bar{x}_{kJ})^2 + (\bar{y}_{kI} - \bar{y}_{kJ})^2 + (\bar{z}_{kI} - \bar{z}_{kJ})^2}\end{aligned}$$

O valor de limiar 0.28, definido empiricamente, na equação 5.8 corresponde a uma variação angular das normais aproximadamente de 15° , próximo de $\frac{\pi}{2}$. O valor de limiar 0.10 na equação 5.9, também definido empiricamente, é menor que o limiar da geração de hipóteses para permitir um maior refinamento nesta seleção.

Nesta fase, procura-se uma sequência de superfícies na cena e nos modelos que obedecem um dado critério de continuidade na formação dos pares. Por exemplo, se o par de superfícies da cena (a, b) tem como resposta o par de superfícies (v, w) em algum modelo, o próximo par de superfícies na cena a ser analisado deve ser algum que comece com b . Por sua vez, procura-se no mesmo modelo pares que começam com w , e assim por diante. Esta regra consegue agrupar um conjunto de superfícies no modelo que possui a mesma sequência de ângulos e distâncias que um dado grupo de superfícies na cena. Além disto, pode-se concluir no final desta análise qual ou quais os modelos que apresentam correspondência com um grupo de superfícies na cena.

Na segunda fase, a verificação final, os grupos de superfícies por modelo são analisados através de uma árvore de pesquisa contendo todos os pares selecionados. Os nós

¹O centróide é dado pela média aritmética dos vértices, onde $c_s = (\bar{x}_s, \bar{y}_s, \bar{z}_s)$ e $C_{ks} = (\bar{x}_{ks}, \bar{y}_{ks}, \bar{z}_{ks})$.

da árvore correspondem aos pares de superfícies, os arcos correspondem às ligações da continuidade e os níveis correspondem aos pares de superfícies da sequência da cena.

A melhor combinação de pares corresponde a um caminho em que não há repetição de faces, apenas um nó por nível seja visitado e o segundo elemento do último par coincide com o primeiro elemento do par que iniciou a pesquisa. As superfícies que compõem este caminho da árvore de pesquisa são as superfícies procuradas.

Elabora-se assim uma lista dos resultados encontrados contendo os modelos ordenados na ordem decrescente da quantidade de superfícies identificadas. Esta ordem indica os modelos mais prováveis e são eliminados os que têm superfícies correspondentes à superfícies da cena já identificadas por outro modelo de prioridade maior.

É importante enfatizar que este método de identificação consegue determinar não apenas o modelo como também as superfícies deste modelo que estão visíveis nas vistas e que foram descritas corretamente em 3D.

5.4 Localização

Determinar a localização do modelo identificado na cena é o mesmo que reconstruir o modelo no sistema de referência da cena. Este processo de reconstrução ocorre através de transformações geométricas que transportam o modelo elaborado no sistema de referência do AutoCAD para dentro do sistema de coordenadas da cena, o mundo real. Tais transformações envolvem operações de translações e rotações de superfícies no espaço 3D de acordo com os seguintes passos ilustrados na Figura 5.1:

1. Traçar a superfície do modelo no sistema de coordenadas da cena,
2. Transladar as superfícies da cena e modelo para origem tal que a origem coincida com os centróides de cada superfície,
3. Rotacionar a superfície do modelo tal que sua normal coincida com a normal da superfície da cena,
4. Rotacionar a superfície do modelo tal que uma de suas arestas coincida com uma aresta da superfície da cena,

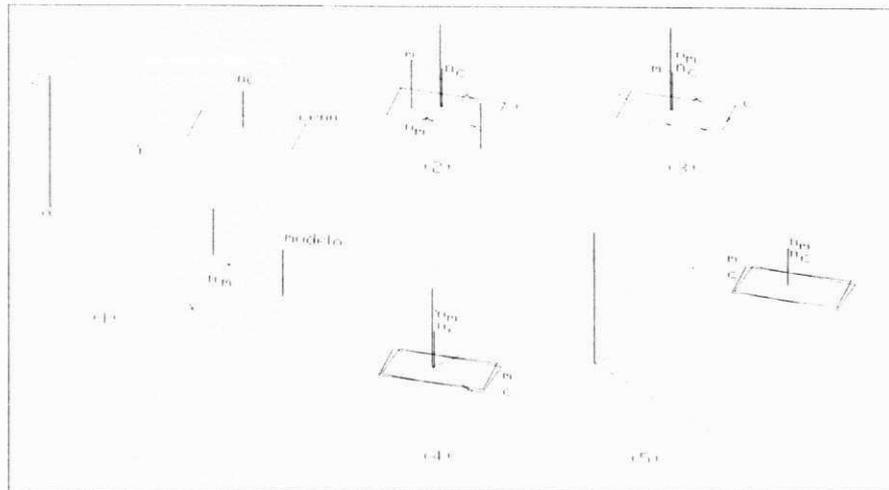


Figura 5.1: Passos para a localização.

5. Transladar as superfícies da cena e modelo até a posição original da superfície da cena.

A transformação geométrica necessária para transladar um plano de sua posição original até a origem do sistema de coordenadas e rotacionar em torno de um vetor de rotação que parte também da origem dos eixos, é definida pela matriz de transformação MT na equação 5.10 [Pav82],

$$MT = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -t_1 \\ r_{21} & r_{22} & r_{23} & -t_2 \\ r_{31} & r_{32} & r_{33} & -t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (5.10)$$

O vetor coluna $t = (-t_1, -t_2, -t_3)$ representa a translação do plano até a origem e a submatriz r representa a rotação de θ graus em torno de um vetor de rotação vr . As componentes da submatriz de rotação são definidas por:

$$\begin{aligned}
 r_{ij} &= \cos \phi_i \cos \phi_j (1 - \cos \theta) - \cos \phi_k \sin \theta (-1)^{i+j}, & \text{se } i < j \\
 r_{ij} &= \cos \phi_i \cos \phi_j (1 - \cos \theta) + \cos \phi_k \sin \theta (-1)^{i+j}, & \text{se } i > j \\
 r_{ij} &= \cos^2 \phi_i (1 - \cos \theta) + \cos \theta, & \text{se } i = j
 \end{aligned} \tag{5.11}$$

onde,

$$\begin{aligned}
 1 \leq i, j \leq 3 \text{ e } k \in \{1, 2, 3\} | k \neq i, j \\
 \cos \phi_1 = \frac{x_r}{d}, \quad \cos \phi_2 = \frac{y_r}{d}, \quad \cos \phi_3 = \frac{z_r}{d} \\
 d = \sqrt{x_r^2 + y_r^2 + z_r^2}, \text{ com } v_r = (x_r, y_r, z_r).
 \end{aligned}$$

Os passos de (1) a (5) traduzem a necessidade de apenas duas matrizes de Transformações: MT_1 e MT_2 . A matriz MT_1 é composta da translação do item (2) e da rotação do item (3). A matriz MT_2 é composta da rotação do item (4) e da translação de (5).

Uma maneira de realizar a translação de um plano para a origem do sistema de coordenadas é através de seu centróide, que representa a distância média do plano à origem. Aplicando-se esta translação (negativa) em todos os vértices, ter-se-á o plano centrado na origem, com o centróide coincidindo com a própria origem.

O ângulo e o eixo de rotação para a submatriz r da matriz MT_1 , aplicada ao modelo, são calculados com as normais das superfícies casadas (da cena \vec{n}_c e do modelo \vec{n}_m). O ângulo, θ_1 , é obtido através do produto escalar entre os dois vetores, ou seja,

$$\cos \theta_1 = \vec{n}_c \cdot \vec{n}_m \tag{5.12}$$

e o eixo de rotação pelo produto vetorial, conforme ilustrado na Figura 5.2,

$$\vec{v}_r = \vec{n}_m \times \vec{n}_c. \tag{5.13}$$

Para a submatriz de rotação de MT_2 , calcula-se inicialmente o ponto médio de duas arestas casadas (da cena pm_c e do modelo pm_m). Uma vez que ambas superfícies estão na origem, calcula-se o ângulo de rotação através dos vetores normalizados dos pontos médios, ou seja,

$$\cos \theta_2 = \frac{pm_c}{|pm_c|} \cdot \frac{pm_m}{|pm_m|}. \tag{5.14}$$

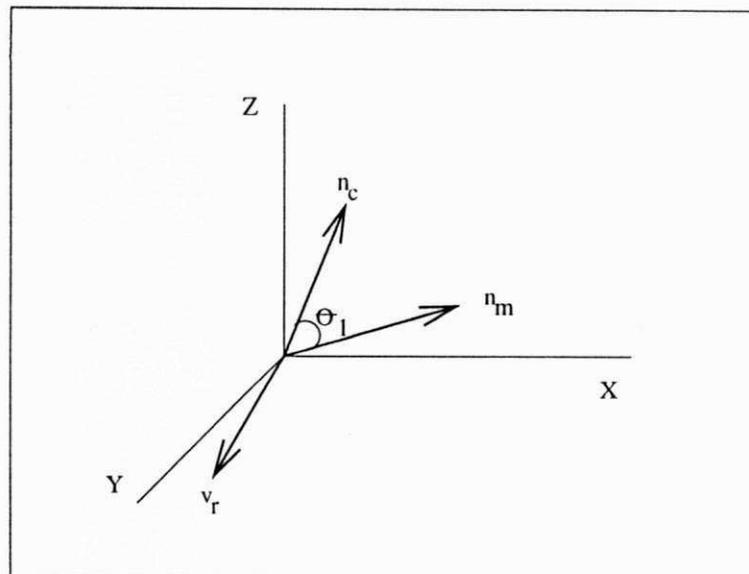


Figura 5.2: Ângulo e vetor de rotação.

O eixo de rotação desta matriz coincide com a própria normal da superfície da cena, conforme ilustrado na Figura 5.3.

O sentido desta segunda rotação é determinado comparando-se o vetor resultante do produto vetorial dos dois vetores ($pm_m \times pm_c$) com o vetor normal da cena. Se tiverem direções diferentes, significa que a aresta do modelo está adiantada da aresta da cena (Figura 5.3a). Se ambos tiverem mesma direção significa que a aresta do modelo está atrasada da aresta da cena (Figura 5.3b), considerando o sentido anti-horário. Desta análise resulta o ângulo $-\theta_2$ ou θ_2 , respectivamente.

A translação para a matriz MT_2 é a mesma que o plano da cena fez inicialmente, sendo que em sentido contrário, ou seja da origem para a posição original.

O capítulo seguinte apresenta os testes efetuados e os resultados obtidos com a aplicação do método de reconhecimento descrito, em cenas simuladas e reais. No Apêndice A, encontram-se os passos para formalização do algoritmo de identificação.

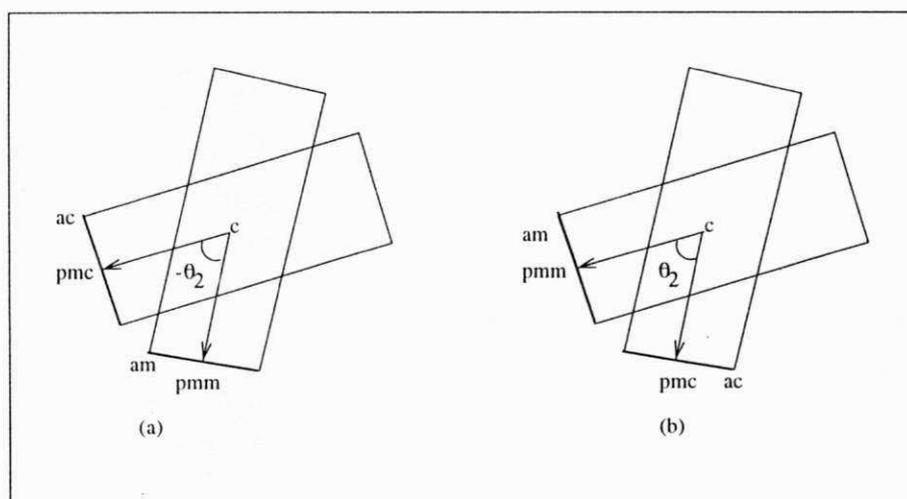


Figura 5.3: Ângulo entre as arestas (a) a_m adiantada (b) a_m atrasada.

Capítulo 6

Testes Efetuados e Resultados Obtidos

6.1 Introdução

Neste capítulo avalia-se a metodologia de reconhecimento de objetos proposta, através dos resultados obtidos experimentalmente. Os testes realizados estão divididos em duas partes, uma parte contendo imagens simuladas e outra com imagens reais. As cenas com imagens simuladas foram elaboradas no AutoCAD que gerou as três vistas 2D. Algumas cenas com imagens de objetos reais foram obtidas com câmera fotográfica comum e outras com câmera fotográfica digital. No caso da câmera fotográfica comum, as fotografias foram digitalizadas em *scanner optico* e em seguida convertidas em arquivos adequados para processamento.

Nas cenas simuladas no AutoCAD, não houve necessidade de realizar a fase de pré-processamento porque as imagens de cada vista são imagens de desenho de linhas (*wire-frames*) do objeto. A fase de segmentação, que inclui a extração dos vértices 2D foram implementadas em linguagem AutoLISP. A técnica multi-vista para elaboração do ambiente tri-dimensional e cálculo dos demais atributos da cena, foi implementada em linguagem C, máquina PENTIUM de 200MHz, assim como os métodos de identificação e localização.

Serão descritas neste trabalho treze cenas contendo objetos isolados, objetos super-

postos e objetos com oclusão parcial, sendo quatro com objetos reais.

Os parâmetros utilizados para avaliar o desempenho do sistema são: taxa de reconhecimento, taxa de rejeição, taxa de acerto e taxa de erro dos objetos e das superfícies, bem como o tempo de processamento e a distância média dos vértices da superfície reconstruída na cena com os da própria cena.

A taxa de reconhecimento do sistema indica quantos objetos (ou superfícies) foram reconhecidos dentre o total (descrito em 3D) da cena. A taxa de rejeição indica quantos não foram reconhecidos. A taxa de acerto, também conhecida como taxa de confiabilidade, indica quantos objetos (ou superfícies) foram identificados corretamente dentre o total reconhecido. A taxa de erro indica quantos foram reconhecidos errados.

O tempo de processamento indica o tempo gasto para o sistema fazer o reconhecimento da cena, foi medido apenas nas fases de identificação e localização.

Em tarefas de reconhecimento de objetos, observa-se que é preferível rejeitar uma descrição da cena a fazer uma identificação errada. Ou seja, a relação entre a taxa de rejeição e a taxa de erro deve ser maior ou igual a um, nunca menor.

As seções seguintes descreverão os resultados encontrados nos testes efetuados (cenas simuladas e cenas reais).

6.2 Cenas Simuladas

O AutoCAD gerou as imagens dos planos X - Y , X - Z e Y - Z das cenas simuladas, considerando que o objeto é colocado em posição estável sobre uma mesa de trabalho. A parte de segmentação da imagem, ou melhor, a extração de vértices em 2D foi simulada também no AutoCAD através de um programa em AutoLISP, o qual gerou arquivos de dados contendo os atributos 3D extraídos das imagens da cena, necessários na etapa de identificação e localização.

As sub-seções seguintes mostrarão os resultados com algumas cenas simuladas, com objetos isolados e objetos com ocultação parcial (espalhados ou empilhados).

6.2.1 Cenas com objetos isolados

As Figuras 6.1 a 6.5, mostram algumas cenas com objetos isolados. A Tabela 6.1 apresenta os números das superfícies totalmente descritas em 3D para cada cena, através da etapa de cruzamento de vértices. São estas que passarão para as demais etapas do sistema. As outras superfícies da cena não são analisadas.

A identificação consta das etapas de geração e verificação de hipóteses. Na verificação há duas fases, uma que identifica o modelo através da coincidência de pares de superfícies e a outra que identifica as superfícies através da análise da pesquisa de árvore contendo as combinações dos pares. A Tabela 6.2 mostra a etapa de geração de hipóteses para as cenas mencionadas. As Tabelas 6.3 a 6.6 mostram os resultados da etapa de verificação. Os "x"s nas referidas Tabelas indicam o grupo de superfícies com modelo coincidente. Por exemplo, na cena 1 - Tabela 6.3, as superfícies s1 (superior 1), l1 (lateral 1), l2 (lateral 2) e f2 (frontal 2) combinadas entre si, estão presentes em um ou em vários grupos de superfícies do modelo 1. Em seguida, os grupos são analisados para identificar as superfícies correspondentes. No caso, as superfícies s1, l1, l2 e s2 correspondem as superfícies 2, 3, 1 e 6, respectivamente.

Como pode ser observado, nas cenas com objetos isolados, o algoritmo de reconhecimento identificou corretamente todos os modelos, constatando-se uma eficiência de 100%. Este percentual também prevaleceu nas outras cinco cenas com objetos isolados analisadas.

Com relação a identificação das superfícies, pode-se verificar que nas cenas 1 e 4 todas as superfícies foram corretamente identificadas.

Uma superfície na cena 2 e uma na cena 5 não foi reconhecida, dentre todas as descritas em 3D, por não ter sido descrita corretamente. Nestas superfícies, houve um falso positivo, ou seja, um dos vértices foi confundido com outro na análise multi-vista. Na cena 3, as quatro superfícies descritas em 3D foram corretamente identificadas. As demais foram rejeitadas por não serem visíveis na vista superior. Nestas cenas, algumas superfícies estão duplicadas.

No total de 36 superfícies a serem reconhecidas nas cenas de objetos isolados aqui relatadas, o algoritmo da técnica multi-vista efetuou a descrição em 3D de 24 superfícies,

excluindo-se as duplicidades. Destas 24, o algoritmo identificou corretamente 22 superfícies, e não identificou duas, o que corresponde a uma taxa de acerto de 91.6%. Em nenhum caso, houve reconhecimento errado de uma superfície.

As taxas medidas na cenas simuladas e reais, com objetos isolados e com mais de um objeto, estão mostradas nas Tabelas 6.18 e 6.19, respectivamente. Foram analisadas 10 cenas simuladas (10 Cen i) de objetos isolados, num total de dez objetos com 54 superfícies descritas em 3D. Na identificação de objetos obteve-se 100% de reconhecimento e 100% de acerto. Na identificação de superfícies, obteve-se 74.1% de reconhecimento e 100% de acerto.

Com relação a etapa de localização, as superfícies identificadas puderam ser posicionadas na cena com um erro médio (distância média) entre os vértices da cena e da superfície do modelo correspondente menor que 0.5cm (ver Tabela 6.8). Este valor corresponde a uma posição bem próxima da cena, se for considerado que as dimensões dos objetos encontram-se em torno de 5 a 15 cm.

A Tabela 6.8 também apresenta o tempo de processamento do método de identificação e localização dos objetos de cada cena. Como pode ser observado, o tempo é menor que um segundo. Um valor bem compatível com processamento em tempo real.

Tabela 6.1: Superfícies 3D (cruzamento 2D).

vistas	cena 1	cena 2	cena 3	cena 4	cena 5
superior	1	1 2 3	1	1	2
lateral	1 2	1 2 4	1 2	1 2	2
frontal	2	4	2	2	1 5 6

6.2.2 Cenas com mais de um objeto

As Figuras 6.6 a 6.9 apresentam casos de cenas contendo mais de um objeto. As Tabelas 6.9 a 6.15 mostram os resultados obtidos.

Como pode ser verificado, dentre as cenas analisadas o algoritmo de reconhecimento não conseguiu identificar um dos objetos da cena 8 e identificou errado um dos objetos

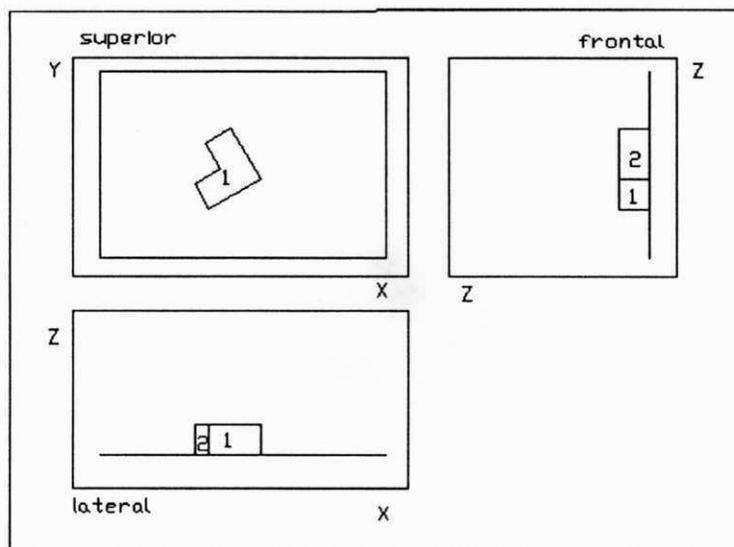


Figura 6.1: Cena 1.

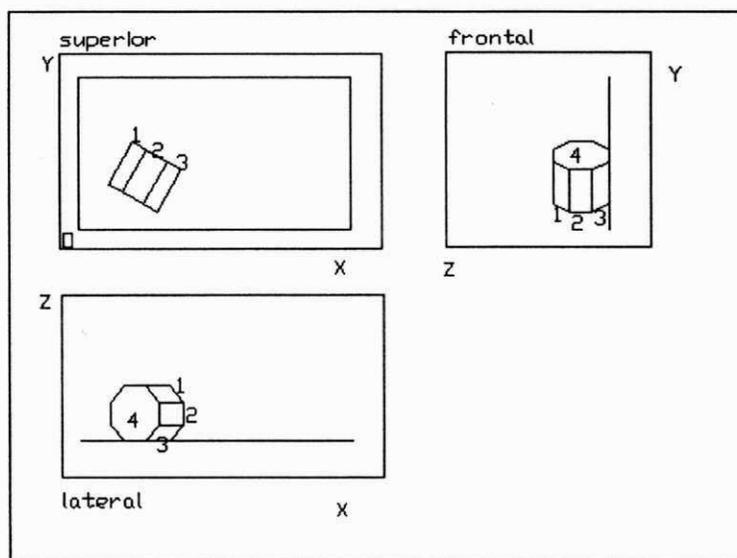


Figura 6.2: Cena 2.

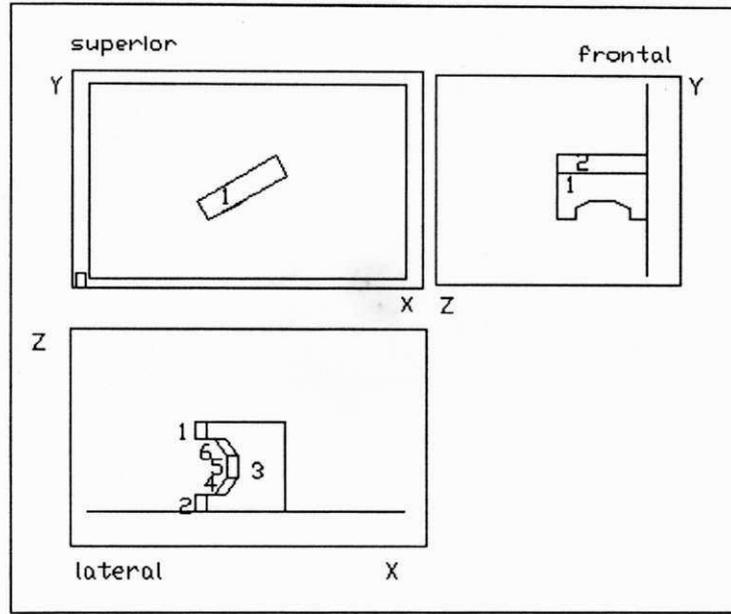


Figura 6.3: Cena 3.

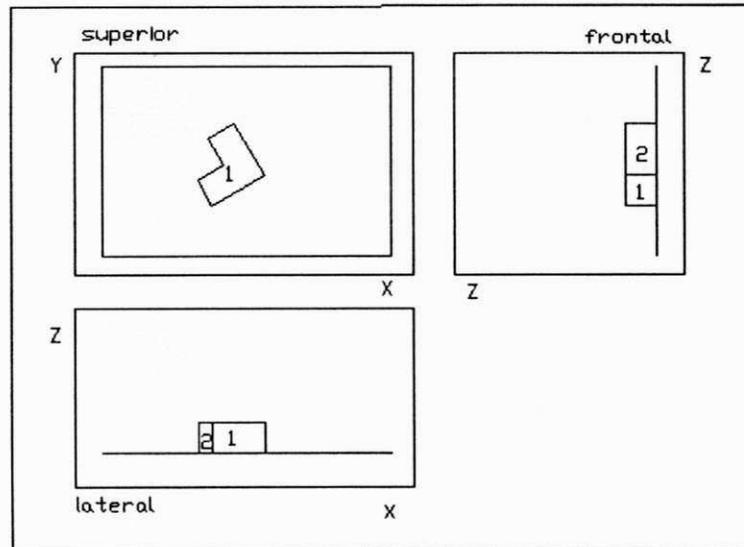


Figura 6.4: Cena 4.

Tabela 6.2: Geração de hipóteses.

Cena	vista	sup.cena	Vi=(modelo; superfícies casadas)
1	sup	1	(1; 2 5)
	lat	1	(1; 3 6) (6; 6 9) (10; 7 10)
		2	(1; 1 4)
	fro	2	(1; 3 6) (7; 3 6) (10; 7 10)
2	sup	1	(5; 11 12)
		2	(3; 3 4 5 6 7 8 9 10)
		3	(3; 3 4 5 6 7 8 9 10)
	lat	1	(3; 3 4 5 6 7 8 9 10)
		2	(3; 3 4 5 6 7 8 9 10)
		4	(3; 1 2) (4; 9)
	fro	4	(3; 1 2) (4; 9)
3	sup	1	(5; 7 8 10)
	lat	1	(5; 11 12)
		2	(5; 11 12)
	fro	2	(5; 7 8 10)
4	sup	1	(8; 4 7)
	lat	1	(6; 7 10) (7; 2 3 4 6) (8; 1 2 5 6)
		2	(1; 2 5) (8; 3 8)
	fro	2	(8; 3 8)
5	sup	2	(3; 1 2) (4; 9)
	lat	2	(4; 10 11 12 13 14 15)
	fro	1	(4; 10 11 12 13 14 15)
		5	(4; 1 2 3 4 5 6 7 8)
		6	(4; 10 11 12 13 14 15)

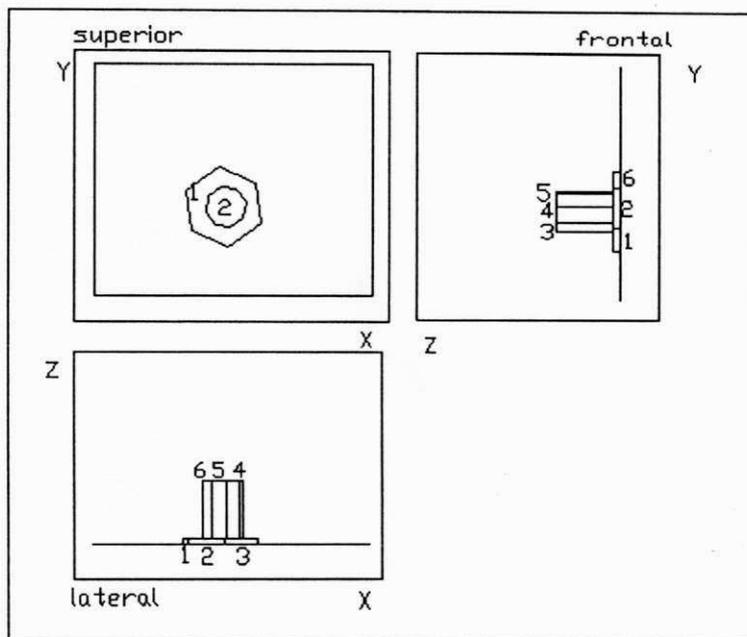


Figura 6.5: Cena 5.

Tabela 6.3: Verificação - Cena 1.

s1	l1	l2	f2	Mod
x	x	x	x	1
2	3	1	6	1

Tabela 6.4: Verificação - Cena 2.

s1	s2	s3	l1	l2	l4	f4	Mod
-	x	x	x	x	x	x	3
-	3	4	6	5	1	2	3

Tabela 6.5: Verificação - Cena 3.

s1	l1	l2	f2	Mod
x	x	x	x	5
7	11	12	8	5

Tabela 6.6: Verificação - Cena 4.

s1	l1	l2	f2	Mod
x	x	x	x	8
4	5	3	8	8

Tabela 6.7: Verificação - Cena 5.

s2	l2	f1	f5	f6	Mod
x	x	x	-	x	4
9	10	11	-	13	4

Tabela 6.8: Localização e tempo de processamento.

CENAS	1	2	3	4	5
\bar{d} (cm)	0.24	0.23	0.42	0.45	0.45
t (seg)	0.50	1.05	0.48	0.51	0.70

na cena 7, apesar de ter listado o modelo correto (modelo 10) como sendo um dos prováveis. Ou seja, houve uma falha na decisão final tomada pelo sistema. Nas demais cenas todos os objetos foram corretamente identificados. Observa-se, neste erro, que o número de superfícies 3D descrito foi mínimo, apenas duas. Isto gerou duas hipóteses com mesma probabilidade, acarretando numa decisão pela primeira da lista. Se a técnica multi-vista tivesse descrito mais uma superfície neste objeto, o sistema o teria identificado corretamente.

Na cena 8 há dois objetos superpostos e o objeto não reconhecido é o que está em baixo. Neste caso, houve falso alinhamento dos vértices na técnica multi-vista comprometendo a descrição 3D deste objeto.

Considerando os oito objetos contidos nas 4 cenas apresentadas, seis foram identificados corretamente, o que corresponde a uma taxa de 75% de acerto. No total de sete cenas analisadas contendo 15 objetos, 11 foram identificados, o que corresponde ao total de 73.3% de reconhecimento. Apenas um foi reconhecido erroneamente, o que corresponde a uma taxa de acerto de 90.9%. O fato de um objeto não ser reconhecido, indica um elevado grau de oclusão do mesmo naquela cena, na maioria dos casos de rejeição efetuada.

Com relação à identificação de superfícies, o percentual também é menor do que em cenas de objetos isolados. Das 48 regiões fechadas (nem todas superfícies completas) nas quatro cenas, 33 foram descritas em 3D, excluindo-se as duplicadas. Destas 33, 15 foram reconhecidas, o que corresponde a um percentual de 45.5%. Em princípio, este percentual parece baixo mas em apenas duas superfícies houve erro de reconhecimento. É o caso do objeto reconhecido erroneamente. A taxa de acerto correspondente é de 86.6%. Caso o modelo tivesse sido reconhecido corretamente, estas superfícies também as teriam sido.

Foram analisadas 7 cenas (7 Cen +), num total de 15 objetos contendo 51 superfícies em 3D, obtendo-se uma taxa de 73.3% de reconhecimento e 90.9% de acerto, conforme pode ser verificado na Tabela 6.18. É uma taxa menor do que nas cenas com objetos isolados, como esperado, em virtude do número de superfícies visíveis serem menores e as dificuldades do cruzamento de dados na técnica multi-vista maiores. O problema é mais acentuado com objetos superpostos uma vez que há mais obstáculos

no cruzamento de dados na vista superior, que serve de referência para outras vistas. Com objetos espalhados na mesa, é possível que haja ocultação parcial de algumas superfícies numa das vistas e na outra não, sendo possível portanto, sua identificação.

A taxa de rejeição para as superfícies é alta, ultrapassando 50%, como é o caso de cenas simuladas com mais de um objeto. Entretanto, a taxa de erro é pequena prevalecendo a máxima de que é preferível não identificar do que identificar errado.

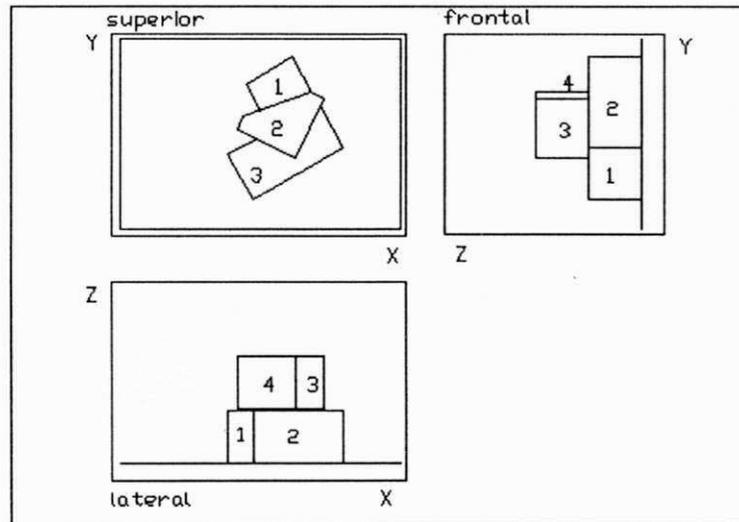


Figura 6.6: Cena 6.

Tabela 6.9: Superfícies 3D (cruzamento 2D).

vistas	cena 6	cena 7	cena 8	cena 9
superior	1 2	3	1 2 3 4 5	1 2 3
lateral	1 2 3 4	1 2 3 5	1 2 3 4 5 6 7	1 2 3 4 5 6
frontal	2 3 4	3 4	1 2 5	2 3 4 5 6 7

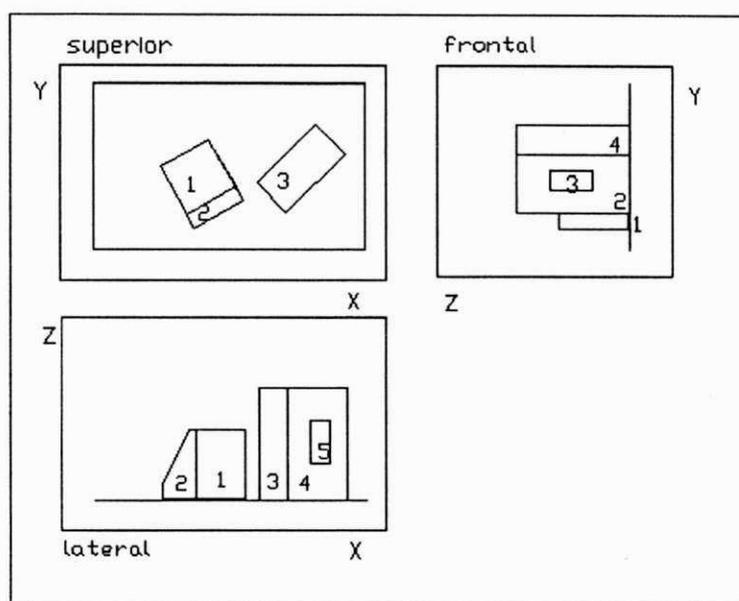


Figura 6.7: Cena 7.

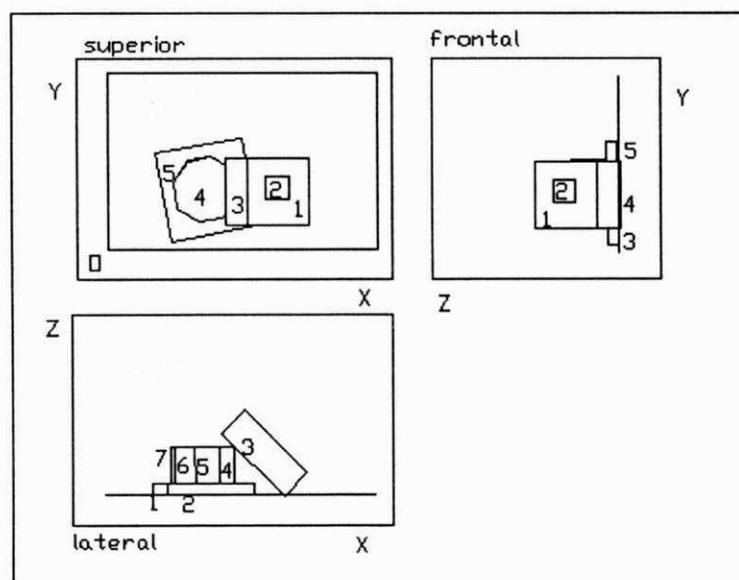


Figura 6.8: Cena 8.

Tabela 6.10: Geração de hipóteses.

Cena	vista	sup.cena	Vi=(modelo; superfícies casadas)
6	sup	1	(1; 3 6) (6; 7 10) (7; 3 6) (10; 7 10)
		2	(7; 3 6)
	lat	1	(8; 1 2 5 6)
		2	(1; 2 5) (8; 3 8)
		3	(7; 1 2 3 4 6) (8; 1 2 5 6)
		4	(6; 7 10) (7; 2 3 4 6) (8; 1 2 5 6)
	fro	2	(8; 3 8)
		3	(7; 1 2 3 4 6)
		4	(4; 10 11 12 13 15)
	7	sup	1
lat		1	(7; 1 2 4)
		2	(7; 3 6)
		3	(5; 7 8 10)
		5	(2; 7 8 9 10 11 12 13 14)
fro		3	(1; 3 6) (6; 7 10) (10; 7 10)
		4	(6; 8) (10; 5 8)
8	sup	1	(6; 6 9) (10; 6 9)
		2	(5; 1 2 3 4 5)
		3	(5; 7 8 10)
	lat	1	(2; 2 3 5 6)
		2	(2; 2 3 5 6)
		3	(6; 8) (7; 1) (10; 5 8)
		4	(1; 3 6) (7; 1) (10; 7 10)
		5	(2; 7 8 9 10 11 12 13 14) (6; 2 4 5) (10; 13)
		6	(2; 7 8 9 10 11 12 13 14) (10; 2 4)
		9	
9	sup	2	(3; 12) (4; 9)
		3	(6; 6 9) (10; 6 9)
	lat	1	(4; 10 11 12 13 14 15)
		2	(6; 7 10) (7; 3 6) (8; 6)
		3	(6; 8) (10; 5 8)
	fro	5	(4; 1 2 3 4 5 6 7 8)
		2	(6; 7 10)
		3	(6; 7 10)

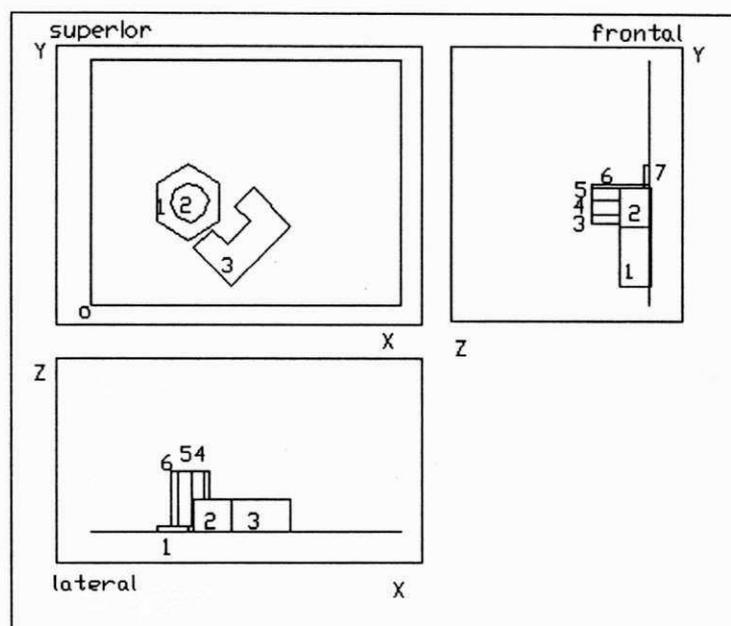


Figura 6.9: Cena 9.

Tabela 6.11: Verificação - Cena 6.

s1	s2	l1	l2	l3	l4	f2	f3	f4	Mod
-	-	x	x	-	-	x	-	-	8
-	x	-	-	-	x	-	-	-	7
-	-	5	3	-	-	8	-	-	8
-	3	-	-	-	2	-	-	-	7

Tabela 6.12: Verificação - Cena 7.

s3	l1	l2	l3	l5	f3	f4	Mod
x	-	-	-	-	-	x	6
-	x	x	-	-	-	-	7
x	-	-	-	-	-	x	10
7	-	-	-	-	-	8	6
-	2	3	-	-	-	-	7

Tabela 6.13: Verificação - Cena 8.

s1	s2	s3	l1	l2	l3	l4	l5	l6	Mod
-	-	-	x	x	-	-	-	x	2
-	-	-	2	3	-	-	-	14	2

Tabela 6.14: Verificação - Cena 9.

s2	s3	l1	l2	l3	l5	f2	f3	Mod
-	-	-	x	x	-	x	-	6
x	-	x	-	-	-	-	-	4
-	-	-	7	8	-	10	-	6
9	-	10	-	-	-	-	-	4

Tabela 6.15: Localização e tempo de processamento.

CENAS	6	7	8	9
\bar{d} (cm)	0.96	1.25	1.33	0.59
t (seg)	0.6	0.42	0.35	0.46

6.3 Cenas Reais

As Figuras 6.10 a 6.13 apresentam as cenas obtidas com objetos reais. Estas foram captadas num estúdio fotográfico, tentando reproduzir o ambiente proposto para a aquisição das imagens, apesar de ter sido utilizada uma única câmera fotográfica, para as três posições. A cena 12 foi captada com a câmera digital e as outras com a câmera comum.

As imagens foram digitalizadas em 256 níveis de cinza em taxa de amostragem de 75 *pixels/inch* e tamanho de 450×300 *pixels*. A distância média dos objetos à câmera foi de 60 cm, o comprimento focal das lentes de 100mm e o fator de ampliação foi de 4.16 (na cena com a câmera digital, o fator de ampliação foi considerado 1).

Após a aquisição, as imagens são convertidas para arquivos de *pixels* em ASCII, visando facilitar os procedimentos para seu processamento. Em seguida, passam pelas etapas de pré-processamento: obtenção da imagem gradiente, do primeiro contorno, do segundo contorno (um *pixel*) e das coordenadas dos vértices. As coordenadas dos vértices das superfícies (regiões fechadas) são obtidas das imagens de *pixels* e o algoritmo de identificação necessita dos vértices em unidades de comprimento real. A etapa seguinte é da transformação geométrica da câmera, para se obter as posições reais do objeto. Com as dimensões reais dos vértices das superfícies de cada imagem, obtém-se os vértices 3D, através do cruzamento de dados. Por fim, faz-se a etapa de identificação e localização dos objetos na cena.

As Figuras 6.14 a 6.25 mostram os resultados das etapas de pré-processamento realizado com as imagens reais: gradiente e contorno da cena. Devido a resolução da impressão, as imagens gradiente e de contorno são muito parecidas. As Tabelas 6.16, 6.17 e 6.19 mostram os resultados finais obtidos com as cenas reais.

Como pode ser verificado, o algoritmo de reconhecimento só não conseguiu identificar o objeto que está abaixo do outro na cena 12. As demais cenas tiveram resultados muito bons, comprovando a eficiência do sistema. Nestas cenas, o percentual de acerto bruto dos modelos foi de 83.3% e o percentual de acerto de superfícies dos modelos foi de 71.8%, que corresponde a 23 superfícies reconhecidas de um total de 32 descritas.

Em nenhum caso houve reconhecimento errôneo. Este resultado indica a confiabili-

dade das decisões do sistema, ou seja, em 100% dos casos em que houve reconhecimento este foi efetuado corretamente.

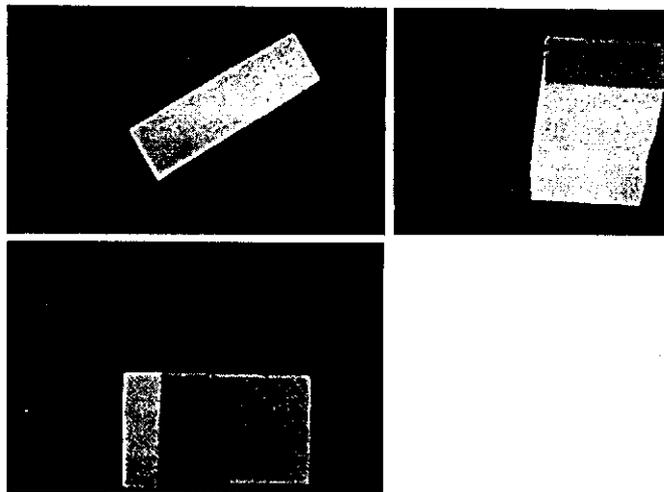


Figura 6.10: Cena 10 - superior, lateral e frontal.

Tabela 6.16: Superfícies 3D (cruzamento 2D).

vistas	cena 10	cena 11	cena 12	cena 13
superior	1	1	1 2	1 2 3
lateral	1 2	1 2	1 2 3 4 5 6	2 3
frontal	2	2	1 2 3 4 6	3 4 5 6 7 8

6.4 Pré-Processamento das Cenas Reais

As Figuras 6.14 a 6.25 mostram os resultados obtidos nas etapas de pré-processamento das cenas com imagens reais.

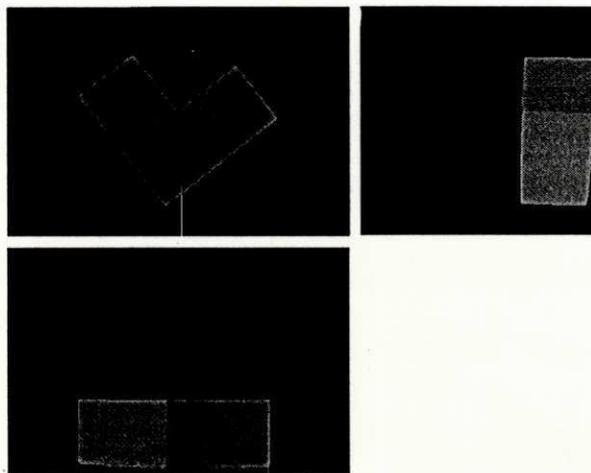


Figura 6.11: Cena 11 - superior, lateral e frontal.

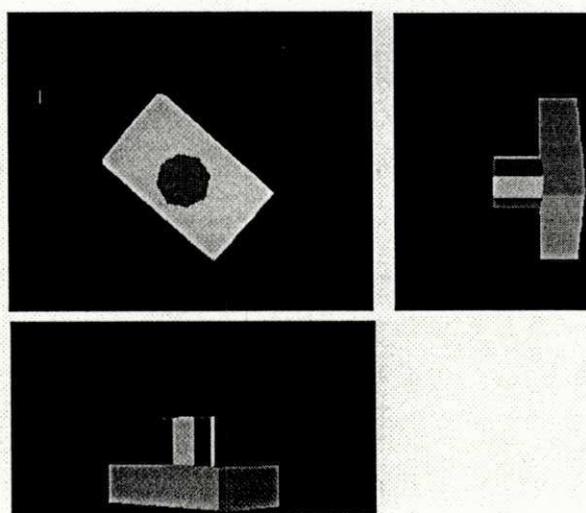


Figura 6.12: Cena 12 - superior, lateral e frontal.

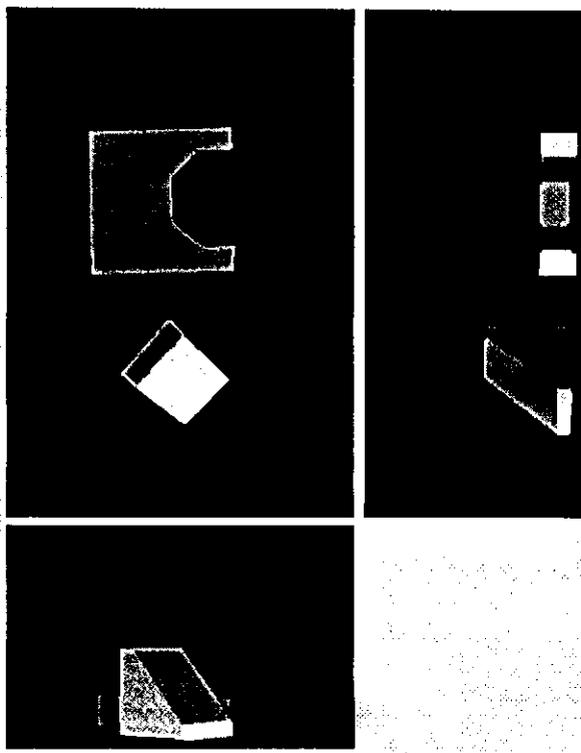


Figura 6.13: Cena 13 - superior e lateral.

Tabela 6.17: Reconhecimento das imagens reais.

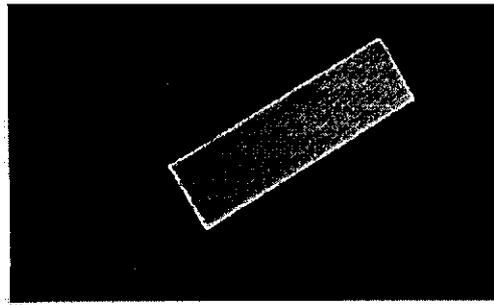
	cena 10	cena 11	cena 12	cena 13
modelo	1	8	3	5 7
superfícies	2 3 1 6	4 3 8 6	1 3 4 5 6 7 8	(5; 6 11 1 3) (7; 1 3 5 6)
tempo(s)	0.42	0.51	1.10	0.62
\bar{d}	0.28	0.39	1.1	0.82

Tabela 6.18: Cenas Simuladas.

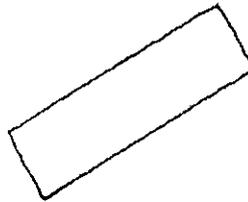
10 Cen i	Rec.	%	Rej.	%	Ac.	%	Er.	%
10 obj.	10	100	0	0	10	100	0	0
54 sup	40	74.1	14	26.9	40	100	0	0
7 Cen +	Rec.	%	Rej.	%	Ac.	%	Er.	%
15 obj	11	73.3	4	26.7	10	90.9	1	9.1
51 sup	24	47.1	27	52.9	22	91.7	2	8.3

Tabela 6.19: Cenas Reais.

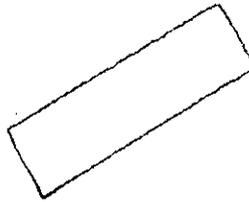
2 Cen i	Rec.	%	Rej.	%	Ac.	%	Er.	%
2 obj	2	100	0	0	2	100	0	0
4 sup	4	100	0	0	4	100	0	0
2 Cen +	Rec.	%	Rej.	%	Ac.	%	Er.	%
4 obj	3	75.0	1	25.0	3	100	0	0
24 sup	15	62.5	9	37.5	15	100	0	0



(a)



(b)



(c)

Figura 6.14: (a) Imagem superior (b) gradiente e (c) contorno final da cena 10.

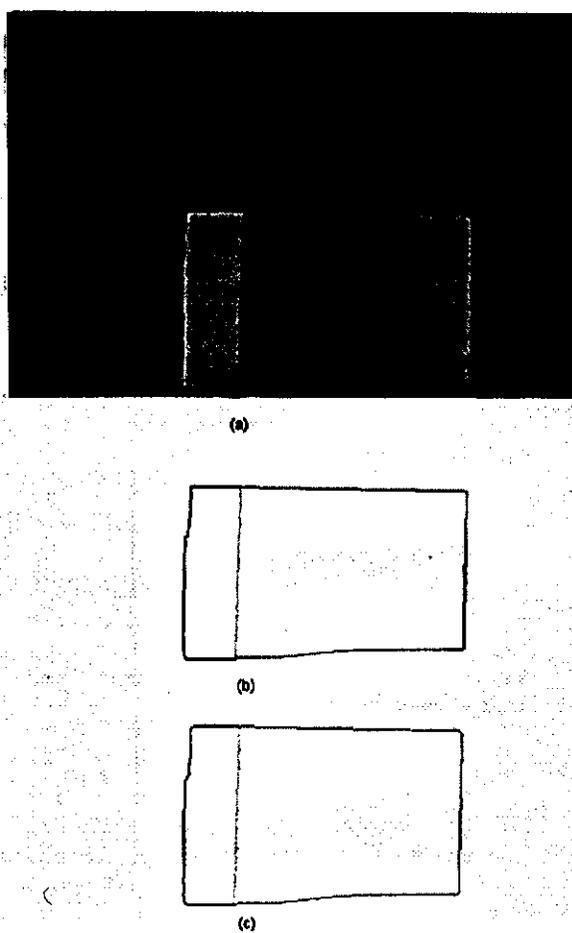
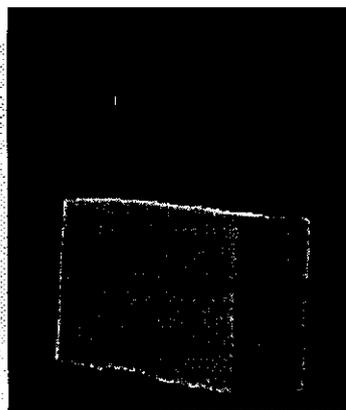
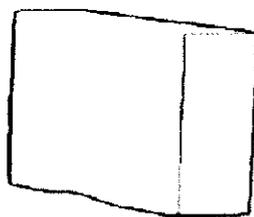


Figura 6.15: (a) Imagem lateral (b) gradiente e (c) contorno final da placa 10.



(a)

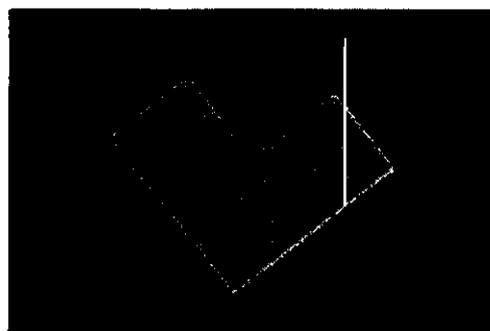


(b)

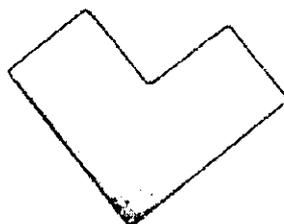


(c)

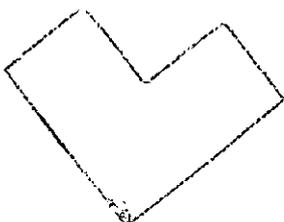
Figura 6.16: (a) Imagem frontal (b) gradiente e (c) contorno final da cena 10.



(a)

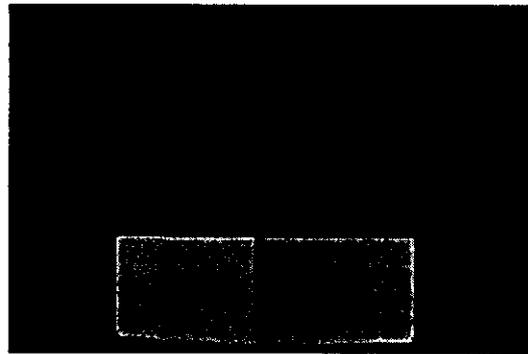


(b)



(c)

Figura 6.17: (a) Imagem superior (b) gradiente e (c) contorno final da cena 11.



(a)

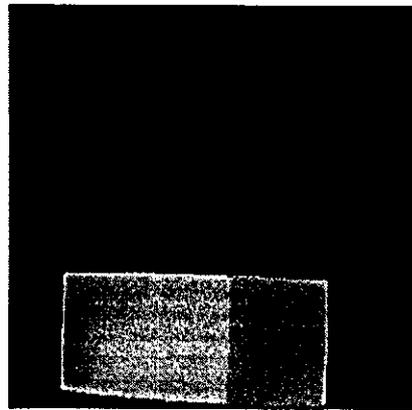


(b)

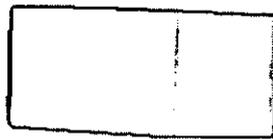


(c)

Figura 6.18: (a) Imagem lateral (b) gradiente e (c) contorno final da cena 11.



(a)

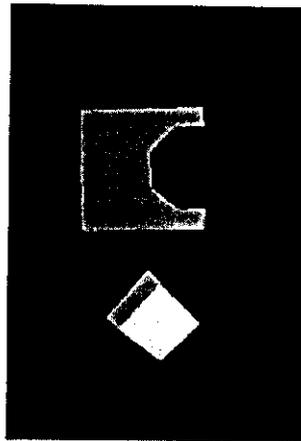


(b)

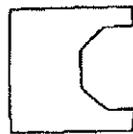


(c)

Figura 6.19: (a) Imagem frontal (b) gradiente e (c) contorno final da cena 11.



(a)

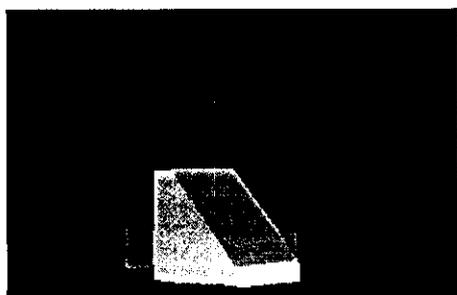


(b)



(c)

Figura 6.20: (a) Imagem superior (b) gradiente e (c) contorno final da cena 12.



(a)

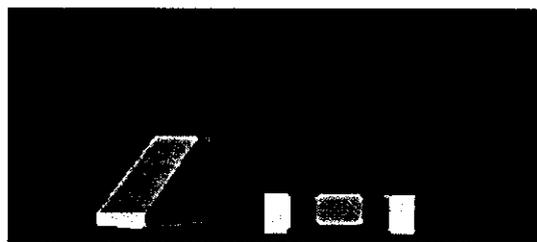


(b)



(c)

Figura 6.21: (a) Imagem lateral (b) gradiente e (c) contorno final da cena 12.



(a)

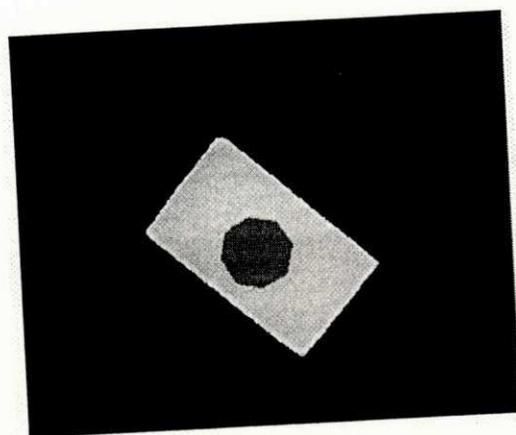


(b)

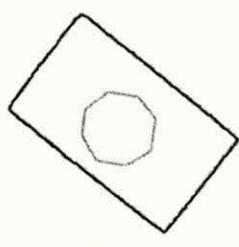


(c)

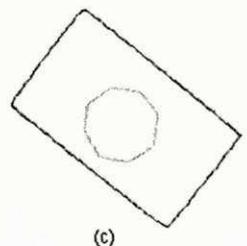
Figura 6.22: (a) Imagem frontal (b) gradiente e (c) contorno final da cena 12.



(a)

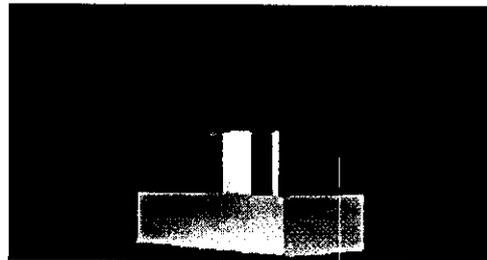


(b)

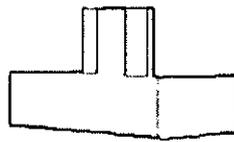


(c)

Figura 6.23: (a) Imagem superior (b) gradiente e (c) contorno final da cena 13.



(a)

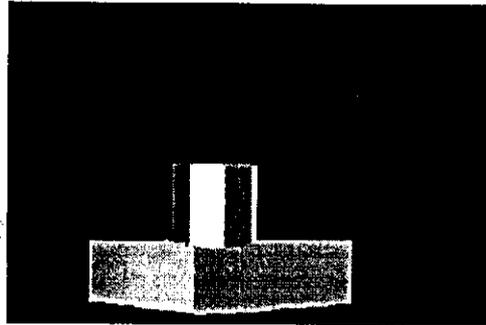


(b)

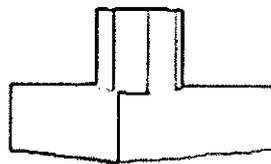


(c)

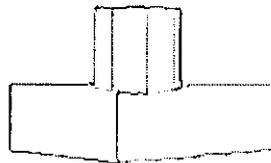
Figura 6.24: (a) Imagem lateral (b) gradiente e (c) contorno final da cena 13.



(a)



(b)



(c)

Figura 6.25: (a) Imagem frontal (b) gradiente e (c) contorno final da cena 13.

Capítulo 7

Conclusões e Contribuições

Esta tese apresenta um método para reconhecimento de objetos 3D de superfícies planas, baseado no casamento de atributos geométricos extraídos de imagens do objeto.

As principais características do método desenvolvido são:

- *eficiência na convergência para um modelo* - o fato de se analisar, através de agrupamentos por modelo, as ocorrências de pares de normais de superfícies casadas da cena e do modelo, resulta numa rápida convergência para um modelo quando se determina o modelo que contém todas as ocorrências. Esta eficiência é comprovada pelos baixos tempos de processamento obtidos.
- *precisão e confiabilidade* - as altas taxas de acerto e as baixas taxas de erro obtidas tanto para objetos isolados quanto para com oclusão parcial, atestam a precisão e a confiabilidade do sistema.
- *simplicidade de entendimento* - o uso de características geométricas comuns como a área e a normal de superfícies resulta em um sistema de operação extremamente simples, de fácil entendimento e aprendizagem, sem necessidade de cálculos superiores.
- *baixo esforço computacional e alta velocidade de operação* - a simplicidade da técnica empregada no casamento de atributos, o uso de poucas características

geométricas e as restrições iniciais para a pesquisa da geração de hipóteses contribuem para um baixo esforço computacional e conseqüentemente alta velocidade de operação.

- *pouco espaço de memória e baixo custo* - o uso de modelos construídos através do AutoCAD substitui as estruturas de montagem necessárias para aquisições de imagens para elaborar o modelo do objeto. Além disto, a representação dos objetos por um número pequeno de atributos geométricos contribuem para um baixo custo do projeto.
- *facilidade de construção e ampliação da biblioteca de modelos* - as facilidades do AutoCAD para elaboração, visualização e extração de atributos geométricos facilitam bastante a operacionalização do sistema.
- Possibilidade de operação conjunta com hardware de processamento de imagens, resultando em desempenho compatível com aplicações em tempo real.

Podem ser citados como contribuições originais resultantes do desenvolvimento desta tese:

1. A Técnica Multi-vista para descrição de vértices em 3D;
2. Construção da biblioteca de modelos através do AutoCAD;
3. Elaboração de programa em AutoLISP para criação de banco de dados geométricos dos modelos;
4. Algoritmo de Reconhecimento com geração de hipóteses através de casamento de características geométricas e verificação através de relações espaciais.
5. Localização do objeto no ambiente da cena realizada após a identificação do objeto e das superfícies visíveis, permitindo uma rápida transformação geométrica.

Grimson [Gri90] cita um conjunto de critérios para avaliação de sistemas de reconhecimento de objetos. Dentre estes critérios, destacam-se: *eficiência, exatidão, robustez e escopo*.

A eficiência deve ser medida em termos do tempo de operação do sistema, o qual está relacionado com a complexidade do método. A exatidão diz respeito às interpretações corretas dos dados a serem analisados, ou seja o sistema não deve acertar dados falsos nem errar dados verdadeiros. A robustez diz respeito a como o método processa os sinais desejáveis dentre os sinais indesejáveis (ruído). O escopo diz respeito ao conjunto de circunstâncias sobre o qual o método encontrará os critérios que tornarão o sistema tão amplo quanto possível. Um método realístico deve trabalhar sempre quando apresentado à cenas com objetos parcialmente ocultos, com dados espúrios e com informações relativamente esparsas.

Com relação ao método de reconhecimento apresentado, pode-se concluir em termos de eficiência que o tempo de operação em ambos testes realizados é pequeno, tipicamente algumas frações de segundos, excluindo o tempo de aquisição e pré processamento das imagens. Isto deve-se às características do sistema, já citadas. O tempo de processamento aumenta um pouco quando os objetos na cena contém superfícies semelhantes. No entanto, este aumento não afeta a eficiência do método uma vez que, comparado a outros trabalhos, este tempo ainda é pequeno (alguns poucos segundos).

Faugeras-Herbert [Fau86] encontram instâncias de objetos 3D na ordem de 5 a 30 segundos, excluindo o tempo de processamento das imagens.

Liu-Tsai [Liu90a] apresentam um método de reconhecimento de objetos 3D a partir de múltiplas vistas 2D obtidas, através de mesa giratória. O método baseia-se no casamento de silhuetas do objeto. O tempo de processamento não ultrapassa 5 segundos (máquina IBM PC/XT).

Hung et al [Hun89] apresentam um método de reconstrução de poliedros através de 3 vistas do objeto cuja análise baseia-se no casamento de desenhos de linhas (arestas e vértices) obtidos das imagens do objeto. O método é aplicado apenas para objetos isolados. O tempo de processamento é de 10 a 40 segundos.

O parâmetro *tempo de processamento* de um sistema é muito relativo, depende da máquina, da aplicação, do tamanho das imagens, etc. O ideal seria fazer uma comparação deste tempo com todos os sistemas rodando na mesma máquina e analisando as mesmas cenas. O que pode ser dito a respeito do sistema proposto com relação ao tempo de operação é que este é comparável a outros métodos, conseguindo valores até

inferiores, sem levar em consideração que o método desta tese faz identificação também das superfícies do objeto.

Com relação a exatidão dos resultados encontrados, os testes efetuados em ambos casos, cenas reais ou simuladas, quase sempre encontram soluções corretas. Como esperado, o método apresentou melhores resultados em cenas com objetos isolados sem buracos ou saliências em suas superfícies. Nos testes com mais de um objeto, ocasionalmente, o método anunciou um falso negativo, ou seja, perdeu uma instância (superfície) do objeto. Tal fato ocorre quando a superfície não é completamente visível em pelo menos, duas vistas.

Sobre a robustez, pode-se dizer que os sinais indesejáveis (ruídos) são bem tolerados nos procedimentos de pré-processamento das imagens. Como exemplo, pode-se verificar que as imagens reais (cena 10 e 11) contêm alguns ruídos de luminosidade, que foram reduzidos na conversão para arquivos de *pixels* (ASCII) e eliminados no cálculo do gradiente. Entretanto, observa-se que se houver ruído (sombra) nos pontos de vértices, será prejudicada a descrição 3D das superfícies e por sua vez a sua identificação.

Com relação ao escopo, pode-se dizer que o método de reconhecimento tem aplicações genéricas em qualquer situação em que seja possível representar o objeto por desenhos de linhas e descrever suas superfícies através de vértices 3D. Este método é adequado, por exemplo, à tarefas industriais em que o objeto é colocado numa mesa de trabalho para ser identificado, inspecionado ou para ser realizada alguma tarefa em alguma de suas superfícies.

Baseado nos resultados obtidos, o método é confiável na identificação do modelo e das superfícies visíveis, principalmente para aplicações com objetos isolados.

Como sugestões para trabalhos futuros nesta linha de pesquisa, tem-se:

1. Melhorar a técnica multi-vista para que possa descrever mais superfícies 3D, utilizando, por exemplo, mais câmeras ou mesa giratória que consiga captar mais imagens da cena.
2. Desenvolver algoritmo para que o método também possa ser aplicado a objetos de superfícies curvas.

3. Ajustar algoritmo de extração de contornos e vértices para que as arestas internas sejam contínuas nas coincidências de três arestas.
4. Ajustar algoritmo de reconhecimento para implementações em VLSI de acordo com metodologia codesign (software/hardware).

A fase de pré-processamento e segmentação das imagens é muito específica da aplicação do sistema. Fatores como calibração da câmera, luminosidade do ambiente e taxas de amostragem devem ser ajustados de acordo com a estrutura da cena. Nesta tese, procurou-se dar mais ênfase ao método de identificação e localização dos objetos.

Espera-se, neste trabalho, ter contribuído para o estado da arte e ao mesmo tempo, que sirva de incentivos para que outros, nesta área, possam ser desenvolvidos.

Bibliografia

- [Adv92] Advanced Modeling Extension. RELEASE 2.1 - Reference Manual. *AUTO-DESK*, July 1992.
- [Arm93] Arman, F. and Aggarwal, R. Model-Based Object Recognition in Dense-Range Images - A Review. *ACM Computing Surveys*, 25(1):5-43, March 1993.
- [Aya94] Ayache, N. and Lustman, F. Trinocular Stereovision: Recent Results. *The Third International Conference on Automation, Robotics and Computer Vision - Singapore*, pages 826-828, November 1994.
- [Bes85] Besl, P. J. and Jain, R. C. Three-Dimensional Object Recognition. *ACM Computing Surveys*, 17(1):75-145, March 1985.
- [Bes92] Besl, Paul J. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239-256, February 1992.
- [Bha84] Bhanu, B. Representation and Shape Matching of 3-D Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(3):349-351, May 1984.
- [Bha87] Bhanu, B. and Ho, C. CAD-Based 3D Object Representation for Robot Vision. *Computer*, pages 19-35, August 1987.
- [Bol86] Bolles, R. C. and Horaud, P. 3DPO:A Three-Dimensional Part Orientation System. *The International Journal of Robotics Research*, 5(3):3-25, 1986.

- [Bol91] Bolle, R. M. and Vemuri, B. C. On Three-Dimensional Surface Reconstruction Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):1-13, January 1991.
- [Bor96] Borges, DÍbio L. 3D Recognition by Parts: A Complete Solution using Parameterized Volumetric Models. *Anais do IX SIBGRAPI - MG*, pages 111-118, September 1996.
- [Bro79] Brooks, R. A. , Griener, R. and Binford, T.O. The Acronym Model-Based Vision System. In *Proceedings of 6th International Joint Conference on Artificial Intelligence*, pages 105-113, August 1979.
- [Bro81] Brown, Christopher M. Some Mathematical and Representation Aspects of Solid Modeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 3(4), July 1981.
- [Bro83] Brooks, R. A. Model-Based Three-Dimensional Interpretations of Two-Dimensional Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):140-150, March 1983.
- [Che89] Chen, C. H. and Kak, A. C. A Robot Vision System for Recognizing 3-D Objects in Low-Order Polynomial Time. *IEEE Trans. Systems, Man, and Cybernetics*, 19(6):1535-1563, November 1989.
- [Chi86] Chin, Roland T. and Dyer, Charles R. Model-Based Recognition in Robot Vision. *ACM Computing Surveys*, 18(1):67-108, March 1986.
- [Chi92a] Chiou, R. , Hung, K. , Kweiguó, J. , Chen, C. , Fan, T. and Lee, J. Y. . Polyhedron recognition from multiple 2d camera views. *Pattern Recognition*, (1):1-16, 1992.
- [Chi92b] Chiou, R., Hung, K., Guo, J., Chen, C., Fan, T. and Lee, J. Polyhedron Recognition Using Three-View Analysis. *Pattern Recognition*, 1(25):1-16, 1992.

- [Chu94] Chung, J. and Nagata, T. Stereo Matching by Spheres: Recognizing 3-D Shapes of Circular GCs from a Pair of Contours. *The Third International Conference on Automation, Robotics and Computer Vision, Singapore*, WA2.1:20-24, November 1994.
- [Cor96] Cortez, Paulo C. Reconhecimento de Formas 2D Usando uma Técnica Sequencial Integrada e Modelos Poligonais. *Tese de Doutorado - COPELE/UFPB*, Julho 1996.
- [Dho89] Dhond, Umesh R. and Aggarwal, J. K. Structure from Stereo; A Review. *IEEE Trans. Systems, Man, and Cybernetics*, 19(6):1489-1534, November 1989.
- [Dud77] Dudani, S. S., Breeding, K. F. and McGhee, R. G. Aircraft Identification by Moment Invariants. *IEEE Transactions on Computers*, 26(1):39-45, 1977.
- [Eng88] Engebrecht, J. R. and Wall, F. M. Polyhedral Object Recognition Using Hough-Space Features. *Pattern Recognition*, 21(2):155-167, 1988.
- [Far97a] Farias, M. de Fátima S. e Carvalho, João M. Modelagem Geométrica por Sistemas CAD para Visão Computacional. *XII Congreso Chileno de Ingeniería Eléctrica*, II:579-584, Novembro 1997.
- [Far97b] Farias, M. F. S. e Carvalho, J. M. Reconhecimento de Objetos 3D Utilizando Atributos de Superfícies de Modelos Poliedrais. *Anais do 3º Simpósio Brasileiro de Automação Inteligente*, pages 52-57, Setembro 1997.
- [Fau86] Faugeras, O. D. and Hebert, M. The representation Recognition and Locating of 3-D Objects. *Int. Journal of Robotics Resourch*, 5(3):27-52, 1986.
- [Fig91] Figueiredo, L. H. and Carvalho, P. C. P. Introdução à Geometria Computacional. *18º Colóquio Brasileiro de Matemática, CNPQ, IMPA*, 1991.
- [Fly91] Flynn, P. J. and Jain, A. K. CAD-Based Computer Vision: From CAD Models to Relational Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):114-132, February 1991.

- [Gon93] Gonzalez, R. C. and Woods, R. E. Digital Image Processing. September 1993.
- [Gri84] Grimson, W. L. and L.-Pérez, T. Model-Based Recognition and Localization from Sparse Range or Tactile Data. *The IJRR*, 3(3):3-35, 1984.
- [Gri90] Grimson, Eric L. Object Recognition Constraints. *The MIT Press*, 1990.
- [Gun87] Gunnarsson, K. T. and Prinz, F. B. CAD Model-Based Localization of Parts in Manufacturing. *Computer*, pages 58-65, August 1987.
- [Hof89] Hoffman, R., Keshavan, H. R. and Towfiq, F. CAD Driven Machine Vision. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1477-1488, November 1989.
- [Hor84] Horn, B. K. P. Extended Gaussian Images. In *Proceedings of the IEEE*, volume 72, pages 1671-1686, December 1984.
- [Hun89] Hung, King-C., Chiou, Rong-N., Shyl, Chia-N., Lee, Jan-Y. and Chen, Chin-H. Polyhedron Reconstruction Using Three-View Analysis. *Pattern Recognition*, 22(3):231-246, 1989.
- [Ito86] Ito, Minoru and Ishii, A. Three-View Stereo Analysis. *IEEE Transactions on Pattern Analysis Machine Intelligent*, 8(4):524-532, July 1986.
- [Kle95] Klemt, A. and Infantosi, A. F. C. Classification of Surface Normals in a Tutorial Microc.-Based System for Anatomy. *Proc. of 38th. MWSCAS - Brasil*, 2:1373-1376, 1995.
- [Leh70] Lehmann, Charles H. Geometria Analítica. *Editora Globo S.A.*, 1970.
- [Liu90a] Liu, C. and Tsai, W. 3D Curved Object Recognition from Multiple 2D Camera Views. *Computer Vision, Graphics and Image Processing*, (50):177-187, 1990.

- [Liu90b] Liu, Cheng-H. and Tsai, Wen-H. 3D Curved Object Recognition from Multiple 2D Camera Views. *Computer Vision, Graphics, and Images Processing*, (50):177-190, 1990.
- [Low87] Lowe, D. G. Three-Dimensional Object Recognition from Single Two-Dimensional Images. *Artificial Intelligence*, 31:355-395, 1987.
- [Maj89] Majumdar, J., Levi, P. and Rembold, U. 3-D Model Based Robot Vision by Matching Scene Description with the Object Model from a CAD Modeller. *North-Holland Robotics and Autonomous System*, (5):69-83, 1989.
- [Mar92] Marshall, A. D. and Martin, R. R. *Computer Vision, Models and Inspection*. World Scientific, 4, 1992.
- [Mel96a] Melcher, E., Carvalho, J. M., Barros, M., Naviner, L., Naviner, J. F., et al. A Novel Gradient Operator Suited for VLSI Implementation of 2D Shape Recognition. *Proceedings of IX SBCCI - Simpósio Brasileiro de Concepção de Circuitos Integrados*, pages 321-332, March 1996.
- [Mel96b] Melcher, E., Carvalho, J. M., Cortez, P. C., Naviner, L., Naviner, J. F. A Vertex Detection Algorithm for VLSI Implementation. *Proceedings of IX SIBGRAPI*, pages 367-368, October 1996.
- [Mel97] Melcher, E., Naviner, L., Carvalho, J. M., Naviner, J. F., Moreira, R. A. S., Villar, Y. M. and Morais, M. VLSI Implementation of Contour Extraction from Real Sequences. *VLSI'97 - IFIP International Conference on Very Large Scale Integration*, August 1997.
- [Mur87] Murray, D. W. Model-Based Recognition Using 3D Shape Alone. *CVGIP*, (40):250-266, 1987.
- [Nev77] Nevatia, R. and Binford, T. O. Description and Recognition of Curved Objects. *Artificial Intelligence*, (8):77-98, 1977.

- [Nit88] Nitzan, D. . Three-Dimensional Vision Structure for Robot Applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10:291–309, May 1988.
- [Osh83] Oshima, M. and Shirai, Y. Object Recognition Using Three-Dimensional Information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 5(4):353–361, May 1983.
- [Par93] Park, J. H., Chang, T. G. and Choi, J. S. Three-Dimensional Objects Representation Based on Surface Normal Images. *Pattern Recognition*, 26(6):913–920, 1993.
- [Pav82] Pavlidis, Theo. Algorithms for Graphics and Image Processing. *Computer Science Press*, 1982.
- [Req80] Requicha, A. A. G. Representation for Rigid Solids: Theory, Methods and Systems. *ACM Computing Surveys*, 12(4):437–464, December 1980.
- [Roa87] Roach, J. W., Paripati, P. K. and Wright, J. S. A CAD System Based on Spherical Dual Representations. *Computer*, pages 37–44, August 1987.
- [Ros88] Rosenfeld, A. Computer Vision: Basic Principles. *Proceedings of IEEE*, 76(8):863–868, August 1988.
- [Sha96] Shashua, Amon and Navab, Nassir. Relative Affine Structure; Canonical Model for 3D From 2D Geometry and Applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):873–883, September 1996.
- [She87] Shenneier, M. O., Lumia, R. and Herman, M. Prediction-Based Vision for Robot Control. *Computer*, pages 46–55, August 1987.
- [Sol90] Solina, F. and Bajcsy, R. Recovery of Parametric Models from Range Images. The Case for Superquadrics with Deformations. *IEEE Transactions on Pattern Analysis Machine Intelligent*, 12(2):131–147, February 1990.

- [Sri81] Srihari, S. N. Representation of Three-Dimensional Digital Images. *Computing Surveys*, 13(4):399-424, December 1981.
- [Til83] Tiller, W. Rational B-splines for Curve and Surface Representation. *IEEE Computing Graphics Applications*, 3(6):61-69, September 1983.
- [Tsu90] Tsui, H. and al. Orientation Estimation of 3D Surfaces Patches. *CVGJP*, (50):112-124, 1990.
- [Udu90] Udupa, J. K. and Ajjanagadde, V. G. Boundary and Object Labelling in Three-Dimensional Images. *Computer Vision, Graphics, and Image Processing*, 51:355-369, 1990.
- [Vai92] Vaillant, R. and Faugeras, O. D. Using Extremal Boudaries for 3-D Object Modelling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):157-172, February 1992.
- [Vil97] Villar, Yuri M. Projeto de um ASIC Para Extração de Vértices de Imagens em Tempo Real. *Dissertação de Mestrado em Engenharia Elétrica - COPELE/UFPB*, Dezembro 1997.
- [Wal81] Wallace, T. P. , Mitchell, O. R. and Fukunaga, K. . Three-Dimensional Shape Analysis Using Local Shape Descriptores. *IEEE Transactions on Pattern Analysis Machine Intelligent*, (3):310-323, 1981.
- [Wan84] Wang, Y. F., Magee, M.J. and Aggarwall, J. K. Matching Three-Dimensional Objects Using Silhouttes. *IEEE Transactions on Pattern Analysis Machine Intelligent*, 6(4):513-518, 1984.
- [Wen93] Weng, J. J., Ahuja, N. and Huang, T. S. Learning, Recognition and Segmentation of 3-D Objects from 2-D Images. *IV ICCV - Berlim, IEEE*, May 1993.
- [Won92] Wong, E. K. Model Matching in Robot Vision by Subgraph Isomorphism. *Pattern Recognition*, 25(3):287-303, 1992.

- [Xie94] Xie, Ming. 3D Reconstruction of Quadratics Curves by Stereovision. *The Third International Conference on Automation, Robotics and Vision, Singapore*, wp2.2:187–191, November 1994.
- [Yua93] Yuan, X. A Mechanism of Automatic 3D Object Modelling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):307–311, March 1993.
- [Zha93] Zhang, S., Sullivan, G. D. and Baker, K. D. The Automatic Construction of a View-Independent Relational Model for 3-D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):531–543, June 1993.

Apêndice A

Passos do Algoritmo de Identificação

I - PRINCIPAL

Início:

Passo 1: Le_modelos().

{le atributos dos modelos}

Passo 2: Le_gera().

{le atributos da cena, gera hipoteses}

Passo 3: Combina_normal().

{organiza pares de sup. da cena}

Passo 4: Analisa_normal().

{analisa pares de sup./cena, procurando sequencia continuada}

Passo 5: Resultados(Iu,Iv).

{analisa pares (Iu,Iv) dos modelos, com mesma sequencia da cena,

Iu e Iv sao pares casados com (u,v)}

Fim.

II - Le_modelos()

Início:

Passo 1: Le e armazena atributos_modelos(k).

{k varia de 1 ate numero total de modelos}

Fim.

III - Le_gera()

Início:

Passo 1: Le atributos_cena(i).

{i varia de 1 ao numero total de superficies 3D/cena}

Passo 2: Para cada i, pesquisa atributos modelos(k), faca

Passo 3: Se $AREA_MINIMA(k) \leq area(i) \leq AREA_MAXIMA(k)$ E
 $NUM_AREST_MIN(k) \leq num_arest(i) \leq NUM_AREST_MAX(k)$

Entao

pesquisa faces(j) de (k), calcule ea, ep, enl, el

{j varia de 1 a quantidade de faces de k,

ea, ep, enl, el sao criterios de erro - eqs 5.3 a 5.6}

Passo 4: Se $ea \leq 0.20$ E $ep \leq 0.20$ E $enl \leq 0.20$ E $el \leq 0.20$

Entao

Passo 5: armazene $V_i=k;j$

Termine Se.

Termine Se.

Fim.

IV - Combina_normal()

Inicio:

Passo 1: Pesquisa atributo_cena (i,j).

{i e j sao sup./cena, i e j variam de 1 ate num total
de sup/cena, i diferente de j}

Passo 2: Calcula eteta(i,j), ed(i,j).

{eqs. 5.8 e 5.9}

Passo 3: Se eteta \leq 0.28 E ed \leq 0.10

Entao

Passo 4: armazena (IkI,IkJ)

{IkI,IkJ) faces do modelo k casadas com (i,j), p/ todo i e j}

Termine Se.

Fim.

V - Analisa_normal()

Inicio:

Passo 1: Agrupa combinacoes de pares da cena por modelo coincidente

Passo 2: Armazena pares da cena numa estrutura arvore_cena.

Passo 3: Pesquisa arvore_cena para encontrar sequencia com continuidade
(sequencia_cena).

Fim

VI - Resultados()

Início:

Passo 1: Armazena pares de faces dos modelos correspondentes aos níveis da sequencia_cena (arvore_modelo).

Passo 2: Pesquisa arvore_modelo para encontrar sequencia_modelo equivalente a sequencia_cena.

Passo 3: Imprime relatorio de saida.

Fim