

ANNA HELENA REALI COSTA RILLO

**RECTRI: UM SISTEMA DE RECONHECIMENTO
TRIDIMENSIONAL A PARTIR DE UMA ÚNICA IMAGEM DE
INTENSIDADE LUMINOSA**

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Engenharia.

São Paulo

1994

ANNA HELENA REALI COSTA RILLO

**RECTRI: UM SISTEMA DE RECONHECIMENTO
TRIDIMENSIONAL A PARTIR DE UMA ÚNICA IMAGEM DE
INTENSIDADE LUMINOSA**

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Engenharia.

Área de Concentração:

Engenharia Elétrica

Orientador:

Prof. Dr. João Antonio Zuffo

Para Márcio, Regina e
Marina.

AGRADECIMENTOS

Ao Prof. Dr. João Antonio Zuffo, pela orientação e apoio constantes no decorrer do trabalho.

Ao Prof. Dr.-Ing. U. Rembold, diretor do "Institut für Prozessrechentchnik" (IPR) da Universidade de Karlsruhe, Alemanha, pelas diretrizes seguras e discussões fundamentais para a definição desta tese ocorridas durante minha estadia sob sua orientação, no período de janeiro de 1.991 a julho de 1.992.

Ao Prof. Dr. Márcio Rillo (EPUSP) pelas valiosas sugestões, críticas e observações no decorrer de todo o trabalho.

Ao Prof. Dr. Siang W. Song (IME-USP) pela compreensão, disponibilidade para animadas discussões e incentivo em momentos fundamentais.

À Profa. Dra. Selma S. S. Melnikoff pelo encorajamento, apoio e valiosas sugestões ao trabalho final.

Aos colegas Prof. Dr. Junior Barrera e Prof. Dr. Flávio C. da Silva (IME -USP) pelos estudos conjuntos, despertar de novos horizontes e interessantes discussões nas áreas de visão computacional e inteligência artificial.

Ao Luiz A. R. Costa (Tuca), Reinaldo Bianchi e Patrícia N. S. Rodrigues (Paty) pelo auxílio direto dado em diferentes fases desta pesquisa.

Aos colegas da DAIA / LSI pela amizade, incentivo e apoio no desenvolvimento de todo o trabalho, especialmente João

Kogler, Martin Tacoshi, Cláudio Katsumi, Lúcia Franco, Leliane N. Barros, Tuca, Bianchi e Paty.

Ao programa RHAÉ / CNPq pelo suporte durante minha estadia no IPR, Alemanha.

À CAPES pelo suporte dado durante o desenvolvimento do trabalho no Brasil.

Gostaria ainda de externar o meu mais profundo agradecimento aos meus familiares, que sempre estiveram ao meu lado nesta jornada, fornecendo todo o carinho, ajuda, companhia, incentivo e apoio para que o trabalho fosse realizado.

Finalmente, não sei o que seria de mim sem o permanente estímulo, amor, carinho, compreensão e paciência de meu marido Márcio e minhas queridas filhas Regina e Marina.

SUMÁRIO

Página

Lista de figuras	
Resumo	
"Abstract"	
1. INTRODUÇÃO	1
1.1. Visão Computacional	1
1.2. O Problema do Reconhecimento de Objetos	4
1.3. Objetivos, Histórico e Justificativas	10
1.4. Contribuições	16
1.5. Organização do Trabalho	17
2. RECONHECIMENTO BASEADO EM MODELOS	19
2.1. Definições	19
2.2. Processos Envolvidos	23
2.3. Modelamento	25
2.3.1. Representação por Estruturas Aramadas.....	29
2.3.2. Geometria de Sólidos Construtivos.....	30
2.3.3. Representação por Borda (B-rep).....	32
2.3.4. Representação por Varredura.....	33
2.3.5. Imagem Gaussiana Estendida (EGI).....	34
2.3.6. Geons.....	35
2.3.7. Representação Multivista.....	37
2.4. Formação da Imagem	38
2.4.1. Tipos de Imagens para Reconhecimento.....	39
2.4.2. Métodos de Aquisição de Imagens.....	40
2.5. Definição e Extração de Atributos	44
2.5.1. Atributos Extraíveis de uma Imagem de Intensidade Luminosa em Níveis de Cinza.....	44

2.6. Geração de Hipóteses	48
2.6.1. Classificação da Correspondência de Acordo com o Tipo de Atributos Usados.....	49
2.6.2. Classificação da Busca pelo Tipo de Controle Efetuado.....	50
2.7. Verificação	51
3. O ESTADO DA ARTE EM RECONHECIMENTO E UMA NOVA PROPOSTA	54
3.1. O Estado da Arte em Reconhecimento 3D Baseado em Modelos	54
3.1.1. Abordagens Pioneiras.....	55
3.1.2. Abordagens que Utilizam Imagens de Profundidades.....	60
3.1.3. Abordagens que Utilizam Imagem de Intensidades.....	69
3.1.4. Resumo e Comentários.....	84
3.2. Proposta de um Sistema de Reconhecimento 3D a Partir de uma Única Imagem 2D	87
3.2.1. Motivação.....	87
3.2.2. Proposta do Sistema RECTRI.....	89
4. SISTEMA RECTRI	91
4.1. Introdução	91
4.2. Fase de Pré-processamento	92
4.2.1. Esquema de Representação para a Informação Tridimensional.....	93
4.2.2. Modelos do Processo de Imageamento.....	95
4.2.3. Uso de Vistas Parciais.....	97
4.2.4. Construção da Hierarquia de Estratégias.....	99

4.3. Fase de Reconhecimento	107
4.3.1. Escolha dos Atributos.....	107
4.3.2. Extração dos Atributos.....	108
4.3.3. Agrupamento dos Atributos.....	109
4.3.4. Busca por Correspondência entre Imagem e Modelo.....	110
4.3.5. O Processo de Verificação.....	111
5. IMPLEMENTAÇÃO E RESULTADOS DO SISTEMA RECTRI	114
5.1. Implementação	114
5.2. Domínio Escolhido	116
5.3. Geração do Conjunto de Aparências	118
5.4. Construção da Hierarquia de Estratégias	122
5.5. Aquisição e Segmentação da Imagem	133
5.6. Extração e Agrupamento de Primitivas	138
5.7. Busca por Correspondência de Atributos	143
5.8. Apresentação e Análise dos Resultados	147
5.8.1. Metodologia da Geração de Dados Simulados.....	147
5.8.2. Apresentação dos Resultados.....	148
5.8.3. Resultados.....	149
5.8.4. Análise dos Resultados.....	162
6. CONCLUSÕES E TRABALHOS FUTUROS	164
6.1. Conclusões e Contribuições	164
6.2. Perspectivas Futuras	167
REFERÊNCIAS BIBLIOGRÁFICAS	171
Apêndice I - Dicionário	

LISTA DE FIGURAS

	Página
Fig.1.1	Áreas envolvidas na Visão Computacional.....3
Fig.1.2	Processo de reconhecimento de objetos.....5
Fig.1.3	Exemplos de métodos de formação da imagem e dimensionalidade de atributos extraídos.....6
Fig.1.4	Decomposição funcional típica dos sistemas de Reconhecimento baseados em Modelos.....7
Fig.1.5	A Célula de Montagem do LSI.....12
Fig.2.1	Definições envolvidas no processo de Modelamento.....26
Fig.2.2	Representação por estruturas aramadas.....29
Fig.2.3	Representação CSG.....30
Fig.2.4	Representação por Decomposição Celular.....31
Fig.2.5	Representação por Borda, com aproximação poligonal das faces.....32
Fig.2.6	Representação por Varredura.....33
Fig.2.7	Cilindros Generalizados.....34
Fig.2.8	Exemplo de EGI (simplificação para melhor visualização: desenhos das projeções das faces na esfera gaussiana, em vez dos vetores correspondentes às respectivas áreas).....35
Fig.2.9	Exemplo de primitivas, chamadas geons.....36
Fig.2.10	Exemplo de Representação Multivista.....38
Fig.2.11	Exemplo de aquisição de Imagem Binária.....40

Fig.2.12	Exemplo de aquisição de Imagem de Intensidade Luminosa, em Níveis de Cinza.....	41
Fig.2.13	Estéreo Binocular.....	42
Fig.2.14	Exemplo de aquisição por luz estruturada.....	43
Fig.2.15	Exemplo: um objeto e sua imagem binária.....	46
Fig.2.16	Utilização do método de Newton para localização (figura adaptada de [Lowe.1987]).....	53
Fig.3.1	Princípio da busca em árvore utilizado por Faugeras e Hebert.....	66
Fig.3.2	Classes de atributos extraídos pelo sistema ORA (adaptação das figuras 4 e 5 de [Huttenlocher, Ullman 1988]).....	71
Fig.3.3	Atributo "par-de-vértices" usado no sistema de Thompson e Mundy.....	73
Fig.3.4	Pontos interessantes usados por Lamdan e outros nos modelos dos objetos.....	75
Fig.3.5	Exemplo da representação de um objeto no sistema de Kuno e outros (adaptada de [Kuno et al. 1991]).....	80
Fig.4.1	Esquema geral do sistema proposto.....	91
Fig.4.2	Fase de pré-processamento.....	92
Fig.4.3	Representação volume-face-aresta-vértice.....	94
Fig.4.4	Projeção Ortogonal.....	95
Fig.4.5	Projeção Perspectiva.....	96
Fig.4.6.	Sistema de herança da Hierarquia de Estratégias, onde dois antecessores apontam um mesmo sucessor.....	97

Fig.4.7	Exemplo da construção de uma pequena Hierarquia de Estratégias, composta somente pelas aparências representadas pelos nós B1, do objeto torre, e C1, do objeto cubo.....	103
Fig.5.1	Configuração típica para atuação do sistema RECTRI de reconhecimento.....	115
Fig.5.2	Conjunto de peças-teste do sistema RECTRI.....	117
Fig.5.3	Conjunto de aparências definido para o prisma, com dez planos de visualização.....	119
Fig.5.4	Conjunto de aparências definido para o cubo, com dez planos de visualização.....	119
Fig.5.5	Conjunto de aparências definido para o peça-L, com dez planos de visualização.....	120
Fig.5.6	Conjunto de aparências definido para o peça-T, com dez planos de visualização.....	120
Fig.5.7	Conjunto de aparências definido para o peça-U, com dez planos de visualização.....	121
Fig.5.8	Conjunto de aparências definido para o pirâmide, com dez planos de visualização.....	121
Fig.5.9	Conjunto de aparências definido para o torre, com dez planos de visualização.....	122
Fig.5.10	Condições de adjacência e inclusão entre atributos, aplicadas no critério de validade de nós.....	123
Fig.5.11	Critério de validade de um nó.....	124
Fig.5.12	Exemplo de teste de correspondência entre dois nós.....	126

Fig.5.13	Árvore de equivalência resultante para o exemplo de correspondência dado na figura 5.12.....	126
Fig.5.14	Hierarquia de Estratégias gerada para as aparências das peças Prisma e Cubo.....	130
Fig.5.15	Hierarquia de Estratégias gerada para as aparências das peças Prisma, Cubo e Peça-U.....	131
Fig.5.16	Nós da Hierarquia de Estratégias gerada para as aparências das peças Prisma, Cubo e Peça-U, mostrada na figura anterior.....	132
Fig.5.17	Imagens originais em níveis de cinza para as peças L e T.....	133
Fig.5.18	Resultado da aplicação do programa de [Geiger, Girosi 1991] nas imagens da figura 5.17.....	134
Fig.5.19	Resultado do algoritmo de morfologia matemática para as peças L e prisma.....	135
Fig.5.20	Aproximação poligonal.....	137
Fig.5.21	Aproximação poligonal das imagens da figura 5.18.....	137
Fig.5.22	Aproximação poligonal das imagens da figura 5.19.....	137
Fig.5.23	Condição de colinearidade.....	138
Fig.5.24	Condição de conectividade (junções).....	139
Fig.5.25	Condição de paralelismo.....	139
Fig.5.26	Relações primárias detectadas entre segmentos de uma imagem.....	140

Fig.5.27	Agrupamentos atualmente tratados por RECTRI.....	141
Fig.5.28	Agrupamentos extraídos do exemplo da figura 5.26.....	142
Fig.5.29	Reconhecimento onde uma folha da hierarquia mostrada foi alcançada, para uma imagem do cubo.....	144
Fig.5.30	Reconhecimento onde nenhuma folha da hierarquia foi alcançada, para a hierarquia aqui ilustrada.....	145
Fig.5.31	Reconhecimento efetuado para a peça L da imagem mostrada, indicando o nó correspondente (Reconhecido) e as respectivas aparências (Peça-12 e Peça-13) a ele associadas.....	150
Fig.5.32	Agrupamentos primários detectados para a imagem da peça L e a torre isoladas.....	151
Fig.5.33	Agrupamentos definidos para a imagem apresentada.....	151
Fig.5.34	Peças L e torre reconhecidas, indicando nós correspondentes da hierarquia e aparências associadas.....	152
Fig.5.35	Agrupamentos primários para a imagem das peças L e cubo, isoladas.....	153
Fig.5.36	Agrupamentos intermediários para a imagem das peças L e cubo.....	153

Fig.5.37	Peças L e cubo reconhecidas, indicando nós correspondentes da hierarquia e aparências associadas.....	154
Fig.5.38	Agrupamentos primários para a imagem das peças L e torre, em oclusão.....	155
Fig.5.39	Agrupamentos intermediários para a imagem das peças L e torre, em oclusão.....	155
Fig.5.40	Imagem onde a peça L e a torre se encontram em oclusão parcial e os respectivos nós correspondidos da hierarquia.....	156
Fig.5.41	Aparências associadas aos nós correspondidos da figura 5.40.....	157
Fig.5.42	Agrupamentos primários para a imagem das peças L e cubo, em oclusão.....	158
Fig.5.43	Agrupamentos intermediários para a imagem das peças L e cubo, em oclusão.....	158
Fig.5.44	Imagem onde a peça L e o cubo se encontram em oclusão parcial e os respectivos nós correspondidos da hierarquia.....	159
Fig.5.45	Agrupamentos intermediários detectados para a peça U, com alinhamento acidental de segmentos.....	160
Fig.5.46	Resultado do reconhecimento.....	160
Fig.5.47	Aparências associadas ao reconhecimento apresentado na figura 5.46.....	161

RESUMO

Esta tese propõe um modelo computacional para a interpretação tridimensional de uma cena, a partir de uma única imagem bidimensional. Este modelo computacional está baseado na escolha de uma representação adequada para os objetos, onde todos estejam representados com seus atributos ressaltados, evidenciando de modo integrado suas semelhanças e diferenças, visando facilitar e agilizar a busca por correspondências entre atributos da imagem e dos objetos.

Desta proposta surgiu uma nova representação, chamada *Hierarquia de Estratégias*, onde os atributos de todos os objetos são selecionados e organizados numa fase de pré-processamento. Os atributos são definidos com uma abordagem voltada para a caracterização qualitativa da morfologia das peças, diferindo da maioria dos sistemas de reconhecimento conhecidos na literatura, os quais adotam uma caracterização quantitativa das mesmas.

A escolha de tais atributos é baseada nos critérios ditados pela *Organização Perceptual*, que consiste na habilidade de permitir ao sistema visual humano detectar imediatamente determinadas relações entre elementos de uma imagem.

Este modelo foi implementado no sistema de reconhecimento RECTRI, o qual está sendo integrado na célula de montagem correntemente em desenvolvimento na Divisão de Automação e Inteligência Artificial do LSI (DAIA/LSI - EPUSP).

ABSTRACT

This thesis proposes a computational model for the 3D interpretation of a 2D view based on the choice of a suitable object representation. Object features are made explicit, providing a single, uniform mechanism for coping both with variation and similarity across objects and across views of a single object, enabling the on-line recognition algorithm to be particularly efficient.

From this proposal arrived a new representation, called *Strategy Hierarchy*, which provides the selection and organization of useful and discriminatory object features, during an off-line model preprocessing stage. These features are defined by an approach based on qualitative characterization of object morphology, differing from the greater number of previous related works, which use a quantitative approach.

The choice of these object features is based on the statements made by the phenomenon of Perceptual Organization, which is the ability that permits the human visual system to immediately detect feature relationships among image elements.

The proposed model has been implemented in a recognition system, called RECTRI, integrating the assembly cell currently under development at the Automation and Intelligence Artificial Division from LSI (DAIA - LSI / EPUSP).

1. INTRODUÇÃO

1.1. Visão Computacional

Procurando modernizar o processo produtivo, visando maiores flexibilidade, competitividade, eficiência e menor custo, as indústrias têm, nos últimos anos, investido consideravelmente em automação. Avanços nas tecnologias tanto de CAD (Projeto Auxiliado por Computador) quanto de CAM (Manufatura Auxiliada por Computador) têm provocado uma melhoria na qualidade dos produtos e uma redução no ciclo de produção, através do aumento da eficiência e produtividade do projeto e da produção, numa forma interativa. Do mesmo modo, aumentos significativos na produtividade também podem ser conseguidos com sistemas de robôs inteligentes, adaptáveis e versáteis, dotados de sensores inteligentes, dentre os quais os sistemas de visão computacional se destacam por mostrarem ser uma das características indispensáveis a qualquer entidade que necessite interagir com o mundo.

Porém, a despeito da aparente facilidade com que seres humanos são capazes de ver e entender o que está sendo visto, a visão tem se mostrado como sendo um processo extremamente complexo e ainda não completamente compreendido, mesmo no campo da fisiologia, neurologia, psicofísica e áreas afins, necessárias para o estudo da visão em seres biológicos [Marr 1982; Overington 1992].

Devido à complexidade deste processo, os sistemas de visão computacional até agora apresentados oferecem ainda muitas restrições, não havendo nenhum sistema de propósito geral com atuação tão competente que se aproxime do sistema visual humano [Davies 1990; Grimson 1990; Lowe 1985; Shirai 1987].

Para contornar o problema imposto pela extrema complexidade da visão computacional, procurou-se identificar processos isolados, tratando-os como módulos independentes [Aloimonos 1989; Ayache 1991; Pampagnin 1990]. Desse modo, pode-se solucionar cada problema separadamente e então desenvolver metodologias de integração das informações fornecidas por estes processos independentes. Nessa linha, foram identificados e isolados alguns processos, entre eles: restauração, segmentação, análise de cor, análise de movimento, identificação e classificação de texturas, estereoscopia, codificação, descrição de estruturas geométricas, reconhecimento de formas, localização de objetos, etc [Baessmann, Besslich 1989; Ballard, Brown 1982; Freeman 1990; Gonzalez, Wintz 1987; Horn 1986; Shirai 1987].

Os diferentes processos são abordados por métodos de diversas áreas, basicamente: Processamento de Imagens, Síntese de Imagens e Interpretação de Imagens (figura 1.1).

O Processamento de Imagens efetua transformações de imagens em imagens. A Síntese de Imagens parte de descrições de formas ou objetos para gerar imagens. Já a Interpretação de Imagens, também chamada de Análise e Entendimento de Imagens, tem por objetivo a construção de descrições explícitas e significativas das formas ou objetos físicos apresentados na imagem.

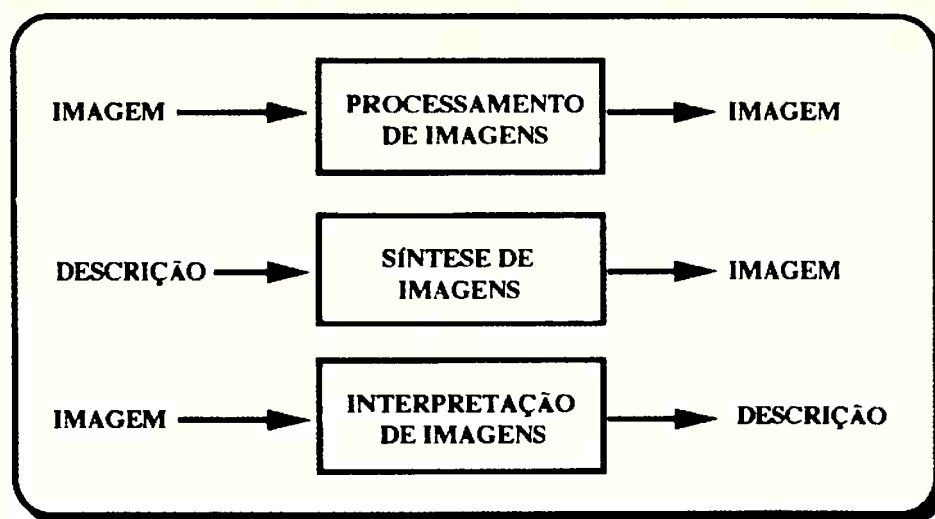


Fig.1.1 Áreas envolvidas na Visão Computacional.

Para atingir este objetivo, a Interpretação de Imagens lança mão de técnicas de Reconhecimento de Padrões, onde populações de indivíduos são classificadas por meio de análise de parâmetros invariantes, e de Inteligência Artificial, a qual procura dotar computadores com capacidade de processamento de informações comparáveis àquelas de organismos biológicos.

Vale ressaltar que o entendimento de uma imagem requer conhecimento *a priori* do domínio da tarefa, isto é, as descrições geradas pela interpretação de uma imagem devem conter toda a informação necessária para a realização de determinada tarefa [Horn 1986; Tomita, Kanade 1984; Zhao 1989]. Dessa forma, a visão é parte integrante de uma entidade maior, que interage com o ambiente.

Um dos processos fundamentais integrantes da Visão Computacional é o Reconhecimento de Objetos, pois envolve os problemas de Identificação e Localização, onde se procura definir quais objetos compõem determinada cena e qual o ponto

espacial onde exatamente se encontram [Grimson 1990; Kanade 1981; Kanatani 1990; Lowe 1985].

1.2. O Problema do Reconhecimento de Objetos

O mundo real que o ser humano vê e toca é composto primariamente por objetos. Se um objeto nunca visto antes for apresentado a um ser humano, ele é capaz de extrair informações sobre o mesmo, procurando observá-lo sob diferentes pontos, visando conhecer sua geometria, textura, funcionalidade, etc e organizando esse conhecimento. Esse processo corresponde à formação de modelos.

Uma vez familiarizado com vários objetos, o ser humano normalmente pode identificá-los a partir de pontos arbitrários de observação, tendo não somente a capacidade de identificá-los, como também de localizá-los e qualitativamente descrever sua orientação.

Todo esse processo, de maneira ampla, pode ser descrito como o processo de Reconhecimento de Objetos. Analogamente, o objetivo do Reconhecimento de Objetos na Visão Computacional, doravante chamado simplesmente de Reconhecimento, consiste em, dados a informação sensorial e o conhecimento *a priori* sobre o mundo, identificar e localizar os objetos presentes na cena visualizada (figura 1.2).



Fig.1.2 Processo de reconhecimento de objetos

As aplicações para o Reconhecimento são muitas; entre elas, cita-se: reconhecimento de caracteres, orientação de veículos autônomos, inspeção de peças industriais, reconhecimento e localização de peças industriais para tarefas de montagem e manipulação, auxílio à medicina, etc.

Muitos processos estão envolvidos na tarefa de Reconhecimento, a saber [Besl, Jain 1985]:

- processo de formação da imagem;
- processo de extração de atributos;
- processo de modelamento;
- processo de geração de hipóteses;
- processo de verificação.

O processo de formação da imagem implica em estabelecer que tipo(s) de sensor(es) é(são) usado(s) e quais princípios físicos estão envolvidos na criação da imagem. Existe uma grande variedade de métodos de obtenção de imagens digitais. As técnicas de análise das imagens podem diferir bastante, dependendo de como foram formadas. Na figura 1.3 tem-se exemplificado alguns dos possíveis métodos de formação da imagem.

O processo de extração de atributos, também chamado de processo de descrição, trata da definição e seus respectivos algoritmos de extração dos atributos relevantes da imagem. Os

modelos dos objetos reais são obtidos no processo de modelamento. Estes modelos são, então, postos em correspondência com as descrições dos atributos da imagem, gerando hipóteses de identificação e localização do objetos. Estas hipóteses serão confirmadas e apuradas ou, então, recusadas, no processo de verificação.

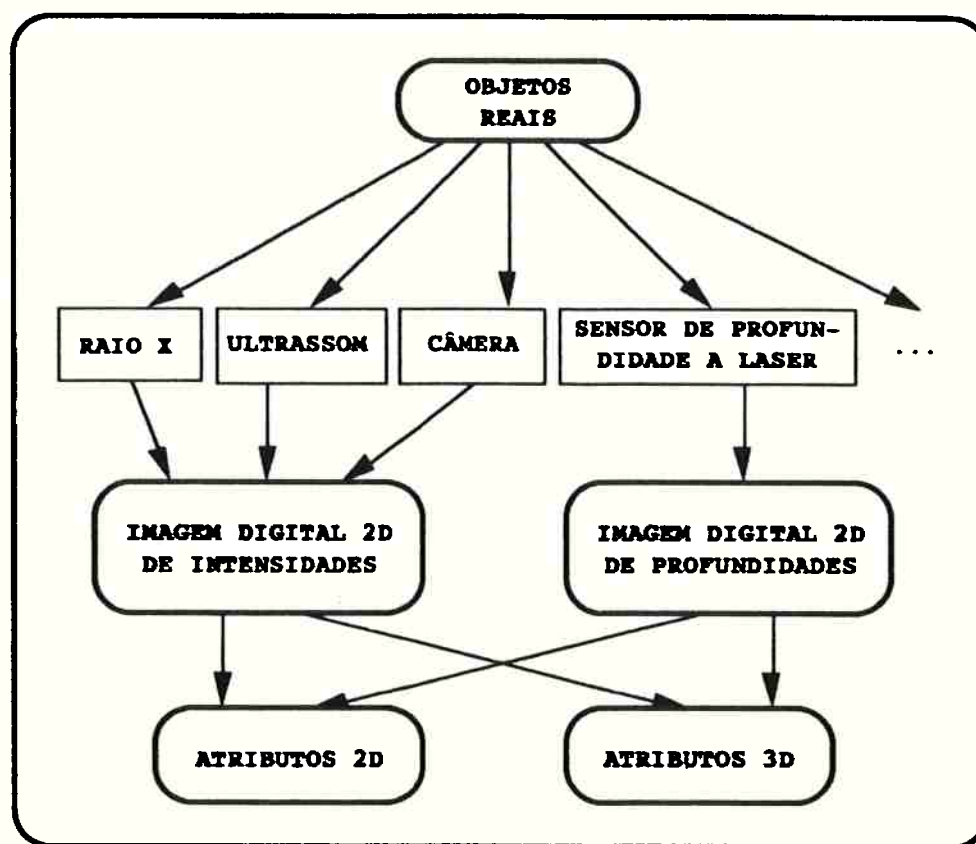


Fig.1.3 Exemplos de métodos de formação da imagem e dimensionalidade de atributos extraídos*.

As abordagens, onde a descrição de uma imagem é comparada com entidades de uma biblioteca de modelos contendo a descrição de cada objeto que pode fazer parte da cena, recebem o nome de

* Alguns autores consideram uma outra dimensionalidade de atributos extraídos (atributos 2 1/2 D), quando não se tem conhecimento por completo do espaço 3D.

Reconhecimento Baseado em Modelos [Moses, Ullman 1992]. Uma decomposição funcional típica destes sistemas é dada na figura 1.4., onde cada retângulo representa um processo e as figuras com bordas arredondadas, dados. A "imagem" representa dados adquiridos pelo processo de formação da imagem e os "conhecimentos prévios", adquiridos pelo processo de modelamento.

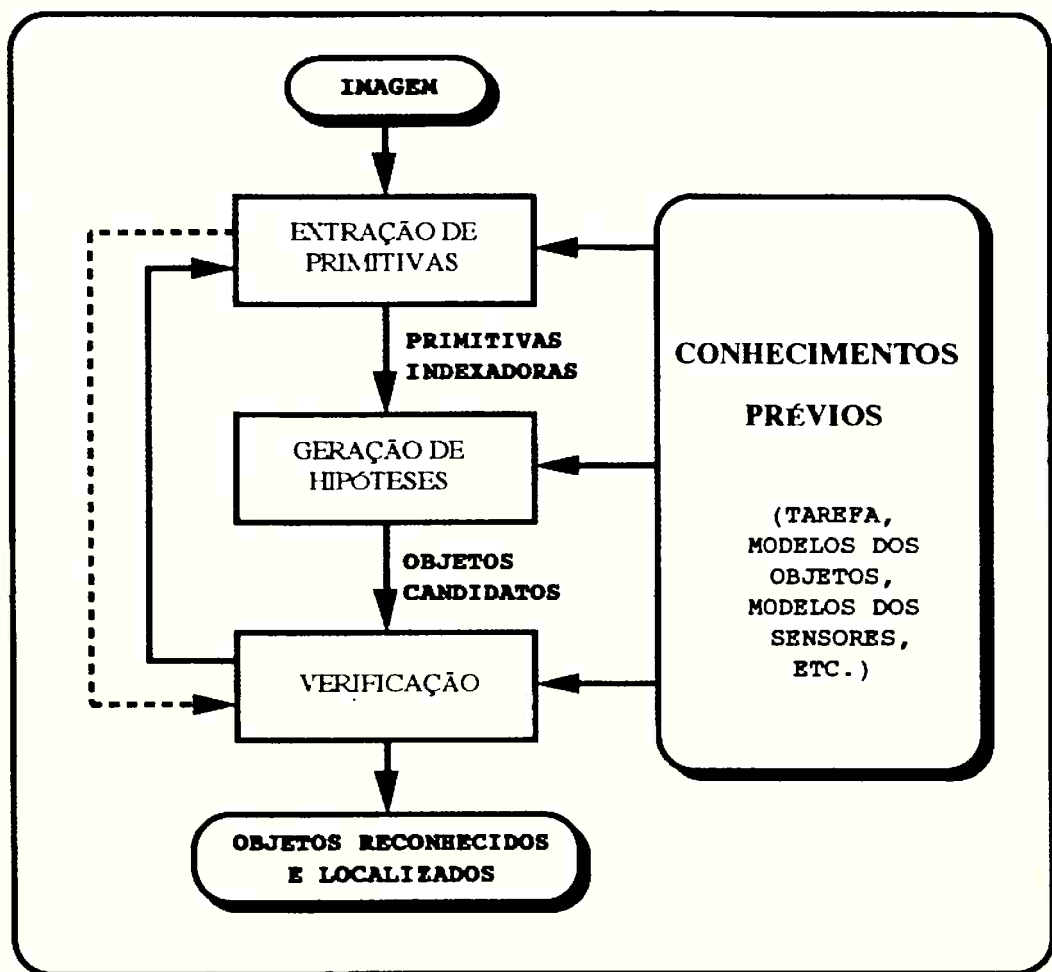


Fig.1.4 Decomposição funcional típica dos sistemas de Reconhecimento baseados em Modelos.

Um conjunto de primitivas, ou atributos, é extraído da imagem de entrada, utilizando essencialmente primitivas geométricas, embora existam propostas de uso de outros tipos de

primitivas, tais como primitivas funcionais [Bergevin, Levine 1989; 1993; Dickinson et al. 1992], de cores [Forsyth 1988; Swain 1990], texturas [Malik, Perona 1989], etc. Aqui serão abordadas apenas as primitivas geométricas.

A escolha de quais primitivas devem ser extraídas da imagem depende:

- da aplicação a que se destina cada sistema de visão computacional;
- dos objetos envolvidos no reconhecimento;
- das características dos sensores usados, entre outros fatores.

A complexidade destas primitivas geométricas pode variar desde pontos no espaço 2D até figuras volumétricas no espaço 3D.

As primitivas resultantes desse processo são chamadas primitivas indexadoras, pois elas indexam a biblioteca de objetos, evidenciando somente objetos que contenham estas primitivas em seus modelos. Dessa forma, selecionam-se objetos possuidores de determinadas primitivas da imagem, gerando hipóteses de correspondência entre primitivas da imagem e dos modelos, fornecendo, assim, objetos candidatos ao reconhecimento.

O último passo envolve a verificação das correspondências em hipótese, gerando como respostas os objetos reconhecidos e suas respectivas localizações ou, então, realimentando o processo, requerendo novas buscas na imagem. Algumas propostas preferem passar, a partir das primitivas indexadoras, diretamente à fase de verificação, sem procurar diminuir a busca por correspondências pela utilização do processo de geração de hipóteses; essas abordagens geralmente utilizam poucos objetos

em sua base de dados e preferem criar primitivas indexadoras bastante potentes, deslocando a complexidade do problema da busca para o processo de extração de atributos.

Na descrição do problema, diversos aspectos são ressaltados como sendo importantes no Reconhecimento [Grimson 1990]:

- tipo(s) de objeto(s);
- tipo de informação sensorial;
- representação da informação sobre cada objeto;
- armazenamento eficiente das representações em bibliotecas de objetos;
- métodos de aprendizado e de adição de novos objetos às bibliotecas;
- escolha de atributos (primitivas);
- métodos de extração de atributos das imagens;
- estratégia de correspondência adotada;
- métodos de indexação de objetos candidatos;
- processos para deduzir posição e orientação de objetos em imagens;
- estratégia usada para o processo de verificação;
- outros.

Muitos sistemas de reconhecimento têm sido propostos, sendo que basicamente diferem entre si na definição de cada um dos aspectos ressaltados, como: primitivas extraídas, estratégia de correspondência adotada, representação de modelos usada, estratégia de verificação ou pelo controle geral adotado.

Nos últimos anos, um paradigma muito abordado tem sido o reconhecimento de objetos tridimensionais (3D) a partir de pontos de observação desconhecidos, utilizando simples imagens

bidimensionais de intensidade luminosa (2D). Esse paradigma constitui o objeto de interesse do trabalho aqui apresentado.

Reconhecimento 3D a partir de 2D envolve a projeção de uma cena em uma imagem, resultando num problema de difícil solução, o qual consiste na ambigüidade dos dados da imagem, devido ao fato de se perder a informação espacial durante o processo de projeção. Além desse fato, problemas inerentes a situações reais da visão computacional também são muito difíceis de se tratar, como oclusão, ruído, dados espúrios, etc. Apesar disso, a visão humana não apresenta dificuldades em propriamente inferir estruturas tridimensionais às imagens.

Buscando entender e simular este fenômeno, muitos sistemas de visão computacional têm sido propostos, os quais são capazes de identificar características de imagens sob projeção e recuperar a posição e orientação de objetos a partir de uma única vista bidimensional. O trabalho aqui proposto é uma contribuição neste sentido, procurando dar uma nova e eficiente abordagem para o problema.

1.3. Objetivos, Histórico e Justificativas

O objetivo deste trabalho consiste na proposta e implementação de um sistema de visão capaz de realizar a interpretação tridimensional (3D) a partir de uma única imagem bidimensional (2D) de intensidade luminosa de uma cena contendo vários objetos, os quais podem apresentar oclusão parcial.

A escolha do tema proposto, reconhecimento de objetos, se justifica pela importância de sua solução para avançar na

direção da construção de sistemas de visão computacional eficientes, flexíveis e abrangentes.

Nosso trabalho em Reconhecimento começou no ano de 1.986, inicialmente tratando somente objetos rígidos bidimensionais (2D) isolados numa imagem binária de intensidade luminosa.

Entende-se como *rígidos* os objetos que possuam distância constante entre quaisquer dois pontos seus, sob qualquer transformação válida do objeto na cena [Grimson 1990]. Além dos rígidos, existem também outros tipos de objetos de importância na área, como: articulados (ex: tesoura), flexíveis (ex: cabo condutor) e deformáveis (ex: massa de modelar), os quais não serão abordados no trabalho.

Objetos 2D são considerados peças estampadas ou peças cujos atributos morfológicos necessários possam ser abstraídos por sua silhueta*.

Imagem binária é aquela composta por somente dois valores numéricos, cujos pontos assumem valores no conjunto $\{0,1\}$. Por exemplo, um nível representando objetos e outro nível representando fundo.

O interesse no sistema de reconhecimento 2D é proveniente de sua aplicação na Célula de Montagem correntemente em desenvolvimento na Divisão de Automação e Inteligência Artificial do Laboratório de Sistemas Integráveis da Escola Politécnica da USP (DAIA/LSI - EPUSP) [Rillo et al. 1992a; 1992b], composta por dois manipuladores, controladores lógicos programáveis [Rillo et al. 1983], câmera, digitalizador de imagens e um sistema de controle e planejamento de atividades (figura 1.5). A Célula executa tarefas simples de montagem.

* silhueta: desenho representativo do perfil de um objeto, segundo os contornos que sua sombra projeta.

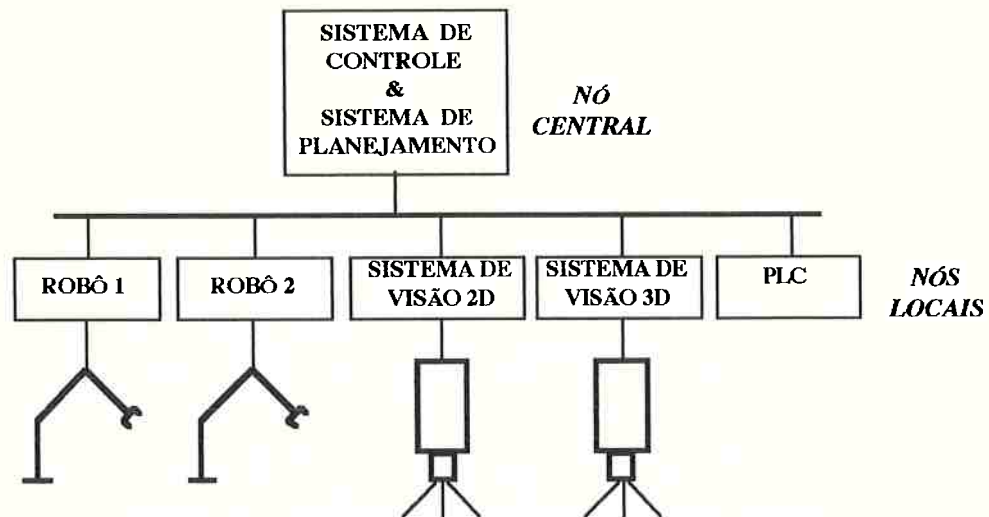


Fig.1.5 A Célula de Montagem do LSI.

A finalidade do sistema de reconhecimento 2D é a identificação (pela silhueta), além da determinação da posição e orientação (2D) das peças que chegam isoladamente à Célula, provendo o sistema de controle e planejamento de informação sensorial necessária para uma maior flexibilidade na montagem.

No sistema de visão 2D da Célula, o reconhecimento é realizado através de medidas globais, como área, perímetro, número de buracos, etc, sendo que a posição da peça é dada pelas coordenadas absolutas do centro de área das mesmas e a orientação, pela direção e sentido do eixo de alongamento da peça (ou então as peças são consideradas rotativamente invariantes). No caso, as transformações possíveis dos objetos na cena se restringem a duas translações e uma rotação (movimentações da silhueta num plano). Silhuetas diversas de um mesmo objeto, decorrentes dos diversos pontos possíveis de observação, são modeladas como sendo elementos distintos.

Foi proposta uma Fase de Treinamento para o sistema de reconhecimento, onde cada silhueta é apresentada separadamente, realizando-se a extração automática dos atributos e a formação de uma biblioteca de silhuetas. Quando todas as peças (e suas respectivas silhuetas) tiverem sido apresentadas, o sistema monta automaticamente uma árvore binária de decisão, baseada nos atributos extraídos, para a classificação de cada silhueta. Essa árvore é percorrida na Fase de Execução do sistema, determinando, de modo muito eficiente, qual silhueta corresponde aos atributos extraídos da imagem corrente.

O sistema de visão 2D foi implementado e integrado à Célula de Montagem, apresentando resultados bastante encorajadores.

Entretanto, numa cena industrial típica, as peças podem estar em contato ou parcialmente oclusas, de modo que o fato de considerar somente peças isoladas resultaria numa limitação muito forte.

Foi, então, proposta uma nova abordagem, para permitir não apenas o reconhecimento de peças isoladas na cena, mas também de peças que se tocassem ou superposicionassem, desde que partes consideráveis delas ainda estivessem visíveis [Rillo 1990a, 1990b].

Uma vez que atributos globais não têm significado quando extraídos de vista parcial de objetos, tornou-se necessário que o reconhecimento fosse baseado na definição e extração de atributos locais, como: buracos, quinas, trechos de contornos, etc.

Utilizaram-se segmentos do contorno da silhueta como atributo para se efetuar o processo de correspondência. Novamente foi utilizada uma Fase de Treinamento, onde cada

silhueta é apresentada e os atributos automaticamente selecionados, levando-se em consideração a biblioteca de silhuetas modeladas.

Comprovou-se mais uma vez a grande utilidade do uso de uma Fase de Treinamento (também chamada de Fase de Aprendizado ou de Pré-processamento) para que o sistema fosse mais eficiente na fase de Execução (ou Fase de Reconhecimento).

A abordagem integrada, permitindo peças isoladas e/ou parcialmente oclusas, foi apresentada como dissertação de mestrado em 1.989 [Rillo 1989] e permanece atuando de maneira bastante satisfatória na Célula de Montagem.

No entanto, a restrição para peças bidimensionais limitava bastante a abrangência de atuação da Célula, não possibilitando, por exemplo, tarefas de inspeção da montagem realizada, uma vez que a informação sensorial limitava-se somente a silhuetas. Deu-se, então, mais um passo no sentido de buscar um sistema de reconhecimento mais flexível e abrangente, abordando o problema do reconhecimento de objetos tridimensionais (3D) a partir de uma única imagem bidimensional de intensidade luminosa, não mais binária, porém em níveis de cinza.

O reconhecimento de objetos pode ser efetuado através de correspondências entre vários tipos de propriedades previstas e medidas, incluindo forma, cor, textura, conectividade, contexto, movimento, sombreamento, etc. No entanto, o trabalho proposto está direcionado para somente uma dessas dimensões: o problema de se conseguir correspondência espacial entre imagem e objeto.

Por correspondência espacial entende-se como sendo a perfeita concordância entre as localizações medidas dos atributos na imagem e aquelas previstas para uma projeção

particular de determinado objeto conhecido. Os atributos usados são essencialmente geométricos, como bordas ou pontos de descontinuidades na intensidade luminosa.

O estudo desse único aspecto da correspondência se justifica pelo fato de que essa informação de localização se mostra usualmente como sendo uma das mais importantes fontes de dados, tanto em número de medidas que podem ser efetuadas numa imagem típica, como também na precisão de cada medida feita na presença de ruído [Clemens, Jacobs 1991; Lowe 1985]. Além disso, resolver o problema nesta única dimensão, a correspondência espacial, freqüentemente é um pré-requisito para determinar correspondências nas outras dimensões citadas .

O interesse neste estudo também se justifica ao se fazer uma analogia com o sistema visual humano, onde se nota que o reconhecimento de objetos a partir do desenho de seus contornos é tão eficiente quanto a partir de fotografias coloridas, apesar destas conterem mais informação [Bergevin, Levine 1992a; 1992b]. Este fato sugere que a construção de descrições menos densas, como contornos, a partir de imagens de uma única vista parece ser uma alternativa apropriada, ou mesmo requerida, em relação a outras abordagens, como reconstrução de superfícies ou volumes.

O complexo desafio de Reconhecimento 3D a partir de imagens 2D foi enfrentado, gerando a formulação de uma nova abordagem [Rillo 1992a, 1992b, 1992c, 1992d, 1993]. Deu-se muita ênfase à Fase de Pré-processamento, a qual se constatou, com a experiência adquirida com os sistemas de visão 2D anteriormente desenvolvidos, ser de extrema importância.

1.4. Contribuições

Um dos principais fatores limitantes dos atuais sistemas de reconhecimento consiste na falta de definições adequadas para a descrição de objetos e imagens, assim como para a realização da correspondência entre eles.

Basicamente, são três os problemas centrais que devem ser abordados para uma solução mais adequada ao reconhecimento:

- definição dos atributos usados como elementos constituintes dos objetos e das técnicas para extraí-los eficientemente de uma imagem;
- representação adequada desses atributos e suas relações, para descrição dos objetos;
- definição da busca por correspondência entre atributos da imagem e dos objetos.

A contribuição principal desta tese se encontra na proposta e busca de uma *solução integrada* para a resolução desses problemas. O enfoque central foi dado na escolha de uma representação adequada para o modelo, onde todos os objetos estejam representados, com seus atributos ressaltados, evidenciando de modo integrado suas semelhanças e diferenças, visando facilitar e agilizar a busca por correspondências entre atributos da imagem e dos objetos.

Desta proposta surgiu uma nova representação, chamada de *Hierarquia de Estratégias*, onde os atributos de todos os objetos são escolhidos e organizados num pré-processamento, com o objetivo de possibilitar uma busca eficiente durante o processo

de reconhecimento propriamente dito. Os atributos são definidos com uma abordagem voltada para a caracterização *qualitativa* da morfologia das peças, diferindo da maioria dos sistemas de reconhecimento conhecidos na literatura [Arkin et al. 1991; Ayache 1991; Bolles, Horaud 1986; Majumdar et al. 1989, etc], os quais adotam uma caracterização quantitativa das mesmas.

A escolha de tais atributos é baseada nos critérios ditados pela *Organização Perceptual*, que consiste na habilidade de permitir ao sistema visual humano detectar imediatamente determinadas relações entre elementos de uma imagem. Critérios auxiliares também são usados para a seleção dos atributos, entre eles: frequência com a qual um atributo é encontrado no objeto, robustez de sua detecção, complexidade de sua computação e probabilidade de corretamente identificar e localizar o objeto.

Essa Hierarquia de Estratégias é construída automaticamente, numa abordagem inovadora e muito promissora.

1.5. Organização do Trabalho

Seguindo a introdução, onde foi colocado o problema a ser abordado e salientado o objetivo da tese, os próximos capítulos são organizados como segue.

O capítulo 2 define e descreve os processos envolvidos em reconhecimento de objetos tridimensionais, além de apresentar, de maneira mais formal, o que é reconhecimento e quais são os critérios a serem satisfeitos para uma boa solução.

No capítulo 3, o estado da arte em reconhecimento tridimensional é apresentado, enfocando principalmente as

abordagens que possuem somente uma imagem de intensidade luminosa como entrada, por serem estas o interesse maior da tese.

Da discussão e comparação entre esses sistemas, são ressaltados seus problemas e uma nova abordagem é proposta, voltada principalmente para o caso do reconhecimento de vários objetos, e não somente um objeto, como é o caso da maioria dos sistemas.

No capítulo 4 a nova abordagem proposta, chamada sistema RECTRI (REConhecimento TRIdimensional), é descrita em detalhes, sendo sua implementação e resultados experimentais apresentados, analisados e discutidos no capítulo 5.

Finalmente, as conclusões são apresentadas no capítulo 6, juntamente com algumas discussões sobre possíveis direções para pesquisas futuras em reconhecimento tridimensional.

2. RECONHECIMENTO BASEADO EM MODELOS

2.1. Definições

A palavra *objeto* será utilizada de maneira informal no decorrer deste trabalho, sendo que a distinção de sua semântica será decorrente do sentido da frase e do contexto. Para tanto, os seguintes conceitos serão definidos: *objeto cena*, *objeto modelo*, *objeto imagem* e *objeto interpretado*.

Objetos do mundo físico são denominados *objetos cena*. A descrição resultante da aplicação de um esquema de representação em um *objeto cena* é denominada *objeto modelo*. A semântica do *objeto modelo* resultará do esquema de representação adotado pelo processo de reconhecimento. Já a semântica do *objeto cena* será considerada como sendo a semântica usual (i.e., aquela da linguagem natural).

Objeto imagem denotará então a descrição dos dados da imagem referentes ao *objeto cena* visualizado sob determinado ponto de observação. *Objeto interpretado* é o *objeto modelo* definido para uma determinada localização no espaço tridimensional, definição esta decorrente da análise de uma imagem.

No contexto de reconhecimento, entende-se por *cena* como sendo o conjunto formado por *objetos cena*, arranjados em localizações (pré-definidas ou não) do espaço tridimensional (real). Em outras palavras, *cena* se traduz numa configuração observável de objetos.

Os termos *dado* e *imagem*, até agora usados informalmente, referem-se à informação sensorial digitalizada de uma cena, traduzida em uma matriz composta por valores numéricos, associada à informação auxiliar referente a como esta matriz de valores foi obtida. Os valores numéricos da matriz podem representar reflectância de superfícies, absorção de raios X, distância ou qualquer outro parâmetro escalar.

O objetivo do Reconhecimento Baseado em Modelos é, de um modo geral, encontrar interpretações apropriadas para os dados (imagens), em relação ao conhecimento prévio que se tem de cada objeto (modelos), da tarefa a ser executada, sensores usados, etc.

Do modo aqui abordado, visão é considerada como sendo parte integrante de uma entidade maior, que interage com o ambiente. Assim, tarefa é definida como sendo a manifestação dessa interação, objetivando determinados resultados, decorrente da interpretação efetuada da cena. No caso da Célula de Montagem do LSI, a tarefa é o reconhecimento de cada peça que será usada na montagem (direcionando os robôs para manuseá-la) e a inspeção da montagem realizada pelos manipuladores.

Para uma definição mais formal do problema de reconhecimento baseado em modelos, considera-se um modelo do mundo composto por N_{tot} objetos (rígidos). Dois objetos são definidos como sendo *não distinguíveis* quando não se consegue separá-los somente por meio de parâmetros decorrentes de suas morfologias. Considerando N_{obj} como sendo o número de objetos distinguíveis, vem:

O_{m_i} : i -ésimo objeto distinguível de N_{obj} ;

N_i : número de ocorrências de O_{m_i} ;

$$N_{\text{tot}} = \sum_{i=1}^{N_{\text{obj}}} N_i \quad (1)$$

Para o caso de $N_i = 1$, isto é, cada objeto distinguível ocorre somente uma vez, tem-se:

$$N_{\text{tot}} = N_{\text{obj}}$$

Sabe-se que:

- existe um sistema global de coordenadas, chamado referência global;
- cada objeto possui um sistema de coordenadas, com relação conhecida à referência global, chamado referência do objeto;
- os objetos são posicionados em relação à referência global por meio de parâmetros de rotação e translação:
 - vetor t : parâmetros de translação;
 - vetor r : parâmetros de rotação.

O número de parâmetros t e r depende da dimensionalidade do problema abordado. Por exemplo:

a) problema 2D (objeto chapado, num plano):

3 parâmetros (2 t e 1 r);

b) problema 3D (objeto 3D, no espaço):

6 parâmetros (3 t e 3 r).

Define-se o *modelo do mundo* W como sendo:

$$W = \left\{ \left(O_{mi}, t_i, r_i \right) \right\}_{i=0}^{N_{\text{obj}}} \quad (2)$$

$$O_{mi} = \left\{ M_j \mid j = 1 \dots m \right\} \quad (3)$$

Sendo:

O_{mi} : i-ésimo objeto do modelo do mundo;

t_i, r_i : definição da referência do objeto em relação à referência global;

M_j : j-ésimo atributo de um objeto modelo;

m : número de atributos do objeto modelo.

Para a imagem de uma cena composta por um único objeto, vem:

$$O_{ci} = \{ A_i \mid i = 1 \dots n \} \quad (4)$$

Sendo:

A_i : i-ésimo atributo da imagem;

n : número de atributos extraídos da imagem;

O_{ci} : objeto na cena.

Deseja-se determinar f , tal que:

$$f: O_{ci} \rightarrow O_{mi} \quad (5)$$

Sendo f uma função de mapeamento dos atributos de um objeto imagem nos atributos de um objeto modelo.

Isto implica que cada atributo do objeto na imagem está em correspondência com o atributo do objeto do modelo, mapeado por f , quando aplicado a uma transformação rígida T_r :

$$A_i \Leftrightarrow T_r \cdot M_{f(i)} \quad (6)$$

O reconhecimento, de modo geral, é então composto por:

- seleção de um modelo candidato O_{mi} (geração de hipótese de identificação);
- determinação (estimativa) de T_r (geração de hipótese de localização);

- estabelecimento de uma correspondência f entre O_{ci} e O_{mi} , que satisfaça (6) (verificação das hipóteses de identificação e localização).

Como resposta, deseja-se:

- a identidade do i -ésimo objeto O_{mi} do modelo do mundo W em correspondência com O_{ci} ;
- a transformação Tr aplicada ao modelo, para corresponder à imagem.

Pode-se desejar a localização em relação à referência global; para isto, basta aplicar a transformação do modelo para a referência global $T_i = (t_i, r_i)$, armazenada no modelo do mundo W , para O_{mi} .

2.2. Processos Envolvidos

Uma boa solução para o problema de reconhecimento deve satisfazer os seguintes critérios [Grimson 1990; Chin, Dyer 1986]:

- **Eficiência:** tanto em termos de tempo de execução quanto em termos de complexidade formal dos algoritmos utilizados no processo de reconhecimento;
- **Precisão:** relacionada não somente ao número médio de objetos reconhecidos corretamente numa cena, como também à precisão na determinação das localizações e orientações dos mesmos. *Falsos positivos* referem-se à identificação de objetos como pertencentes à cena, quando na verdade não o são; *falsos negativos* representam a perda de interpretações corretas. Desse modo, o interesse consiste

em que o sistema responda eliminando falsos positivos e também fazendo com que o número de falsos negativos seja mínimo;

- **Robustez:** o método de reconhecimento usado deve degradar suavemente com o aumento de ruído nas medidas feitas pelos sensores, com a diminuição de dados relevantes obtidos e com o aumento de dados irrelevantes;
- **Flexibilidade:** o sistema deve ser flexível o suficiente para acomodar variações nas dimensões físicas das diversas amostras de determinado objeto, assim como incertezas em relação à sua localização;
- **Abrangência:** na medida do possível, não deve haver restrições quanto às características dos objetos para os quais o reconhecimento se dá.

Na maioria dos sistemas de reconhecimento esses critérios não são totalmente satisfeitos; na verdade, existe um compromisso entre eles, como, por exemplo, quanto mais flexível e abrangente é o sistema, geralmente menos eficiente, preciso e robusto ele é. A cada critério é definida uma relevância diferente, e esta definição varia de acordo com a abordagem utilizada.

Um grande número de fatores limitam a competência dos atuais sistemas de reconhecimento. Um dos principais fatores limitantes é a falta de definições adequadas para representações espaciais, assim como para descrições de objetos e imagens. Conseqüentemente, três problemas centrais aparecem:

- Que atributos devem ser usados e como extraí-los de uma imagem, a fim de que descrevam eficientemente propriedades físicas e suas relações numa cena?

- Como representar adequadamente estes atributos e suas relações, para caracterizar classes semanticamente significativas de objetos, isto é, como definir eficientemente os modelos dos objetos?
- Como deve ser feita a correspondência entre atributos das imagens e os modelos dos objetos para que se possa reconhecer objetos complexos em cenas também complexas?

Uma variedade muito grande de soluções para estes problemas tem sido proposta. Para caracterizá-las melhor, os processos básicos envolvidos no reconhecimento, definidos anteriormente como formação da imagem, extração de atributos, modelamento, geração de hipóteses e verificação serão discutidos isoladamente, mostrando algumas possibilidades de abordagens a serem feitas.

Serão enfocadas somente abordagens que utilizam essencialmente propriedades geométricas de objetos rígidos, uma vez que as características geométricas são as mais usadas na grande maioria dos sistemas propostos, por serem um dos fatores mais importantes no reconhecimento de objetos.

2.3. Modelamento

Para a caracterização do processo de Modelamento, torna-se necessário definir mais precisamente alguns termos (figura 2.1).

Um *esquema de representação* E é uma relação entre o conjunto O de objetos sólidos rígidos (objetos cena) e o conjunto R de representações de objetos (objetos modelo) [Flynn 1990, Requicha 1980]:

$$E: O \rightarrow R .$$

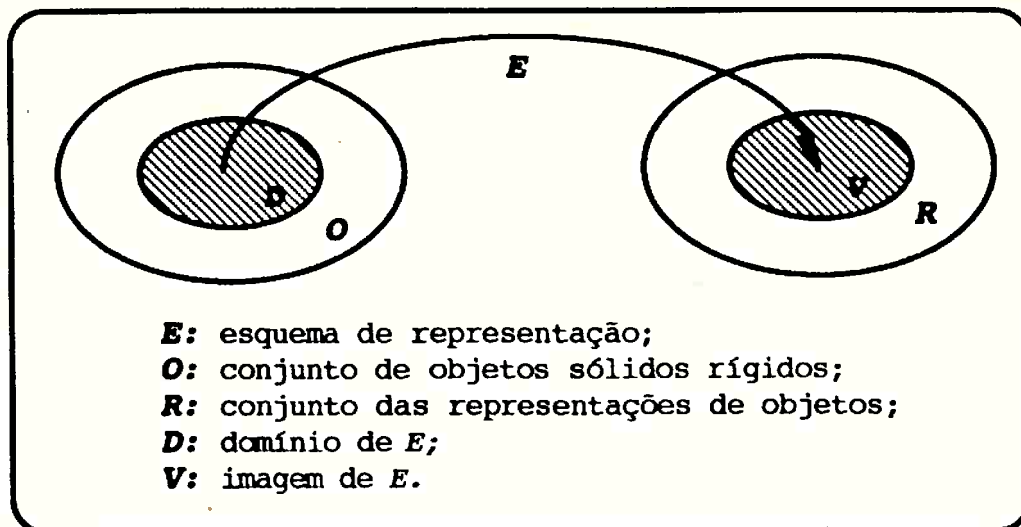


Fig.2.1 Definições envolvidas no processo de Modelamento.

Um elemento $e_i \in E$ é um par ordenado (o_i, r_i) , onde:

$o_i \in O$: um objeto do conjunto de objetos cena;

$r_i \in R$: uma representação (ou modelo) de o_i , de um conjunto de representações válidas.

Uma representação é dita válida quando for sintática e semanticamente correta.

$D \subset O$ é o domínio de E e $V \subset R$ é a imagem de E .

Os elementos $o \in (O-D)$ não são representáveis por E .

Os elementos $r \in (R-V)$ não são representações válidas.

Dentro desse contexto, são relacionadas algumas propriedades importantes que devem ser consideradas na definição do esquema de representação a ser utilizado [Flynn, Jain 1991a; 1991b; Requicha 1980]:

- **Domínio (D):** caracteriza o poder de descrição da representação;

- Validade (V): indica se as representações resultantes da aplicação do esquema de representação escolhido correspondem a objetos válidos, isto é, não possibilita a existência de estruturas simbólicas correspondentes a objetos irreais;
- Não Ambigüidade: cada representação corresponde a um único objeto:

$$(o_1, r) \in E \wedge (o_2, r) \in E \Rightarrow o_1 = o_2$$

$$\forall o_1, o_2 \in D \quad \forall r \in V$$

- Unicidade: cada objeto corresponde a uma única representação, no esquema de representação usado:

$$(o, r_1) \in E \wedge (o, r_2) \in E \Rightarrow r_1 = r_2$$

$$\forall o \in D \quad \forall r_1, r_2 \in V$$

- Concisão: indica quão concisa é a representação; o espaço de memória ocupada pelas representações pode fornecer uma medida de concisão;
- Facilidade de criação: indica quão simples e fácil de ser feita é a descrição, criada pela aplicação de um esquema de representação a um determinado objeto ou conjunto de objetos.

A escolha de um esquema de representação depende de vários fatores; entre eles, cita-se:

- o domínio do esquema de representação escolhido engloba todos elementos que se pretende modelar?
- o esquema apresenta a maioria das propriedades importantes?

- como serão utilizados os modelos gerados (a que se destinam, qual tarefa será executada)?

É importante ressaltar que, no reconhecimento baseado em modelos, os maiores interesses nos modelos residem em dois pontos principais:

- facilidade de ressaltar atributos (2D ou 3D) que serão usados em tarefas de correspondência;
- facilidade de se efetuar transformações rígidas e projeções dos objetos na imagem (síntese de imagens), para tarefas de verificação.

Visando satisfazer eficientemente estas necessidades do reconhecimento, muitas vezes são propostas representações múltiplas ou híbridas para a formação dos modelos dos objetos.

De acordo com as características da tarefa a ser executada, objetos a serem modelados e métodos de aquisição da imagem a serem usados, os modelos podem variar desde listas de atributos globais do objeto (ou silhueta, dependendo da aplicação), como área, perímetro, etc, até estruturas mais complexas, como, por exemplo grafos, onde os nós representam a descrição analítica de pequenas partes do contorno do sólido e os arcos, as relações topológicas entre os elementos representados nos nós.

Em Requicha [Requicha 1980], Flynn e Jain [Flynn, Jain 1991a; 1991b], tem-se boas revisões sobre modelamento, voltadas principalmente para aplicações em visão computacional.

A título de exemplificação da variedade de métodos utilizados, serão apresentados alguns esquemas de representação normalmente usados em CAD e visão computacional, para objetos rígidos, como: representação por estruturas aramadas, CSG, representação por borda (B-rep), EGI, representação por

varredura, geons e representação multivista. Dessa forma, Procura-se mostrar, neste ítem, alguns tipos de representação de objetos rígidos, sem contudo ter pretensão de efetuar uma revisão completa de todos os tipos possíveis existentes, uma vez que são inúmeras propostas e sempre surgem novas abordagens.

2.3.1. Representação por Estruturas Aramadas

Uma representação por estruturas aramadas (figura 2.2) é aquela onde o objeto é descrito por pontos e arestas constituintes, na forma de um grafo, onde:

- nós: são pontos 3D na superfície do objeto (tipicamente, vértices);
- arcos: representam as arestas do objeto.

Esta representação apresenta ambigüidade, mesmo para poliedros de faces planas, porém é concisa. Poucos sistemas de visão computacional a utilizam isoladamente; normalmente, é parte integrante de uma representação híbrida.

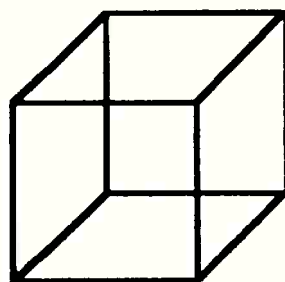


Fig.2.2 Representação por estruturas aramadas.

2.3.2. Geometria de Sólidos Construtivos

No esquema de Geometria de Sólidos Construtivos (CSG: "Construtive Solid Geometry"), um conjunto finito de formas primitivas podem ser transformadas (translação, homotetia, etc) e combinadas (união, diferença, etc) para formar o sólido desejado, conforme mostrado na figura 2.3.

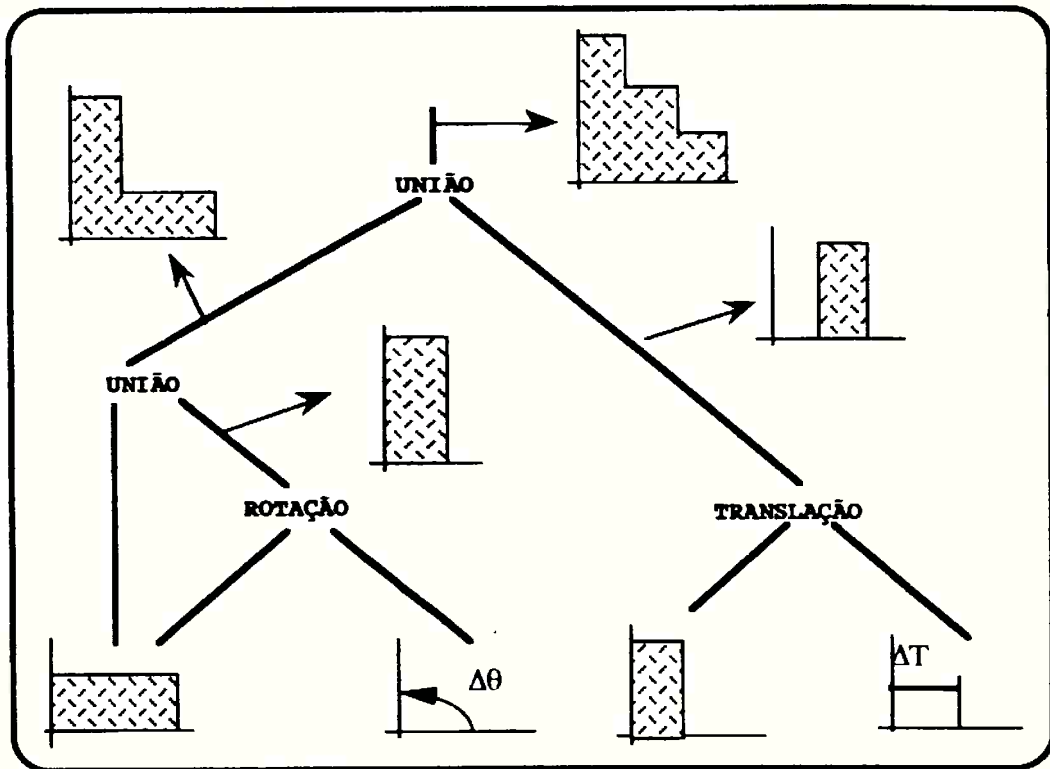


Fig.2.3 Representação CSG.

A descrição é dada numa árvore, onde:

- folhas: formas primitivas;
- nós internos: operações (transformações ou combinações);
- arcos: indicam precedência entre operações.

Para objetos de estrutura geométrica regular, CSG pode fornecer descrições concisas. No entanto, a reconstrução de uma árvore CSG a partir de uma cena é muito difícil, uma vez que este tipo de representação não é único, isto é, pode haver diferentes decomposições de um único objeto, usando o mesmo conjunto de primitivas. Isto faz com que não se consiga fazer uso de sua informação topológica [Bhanu, Ho 1987].

Existem também algumas variantes deste esquema, por exemplo:

- caso seja permitida somente a operação de "colagem", tem-se a chamada representação por Decomposição Celular (figura 2.4);

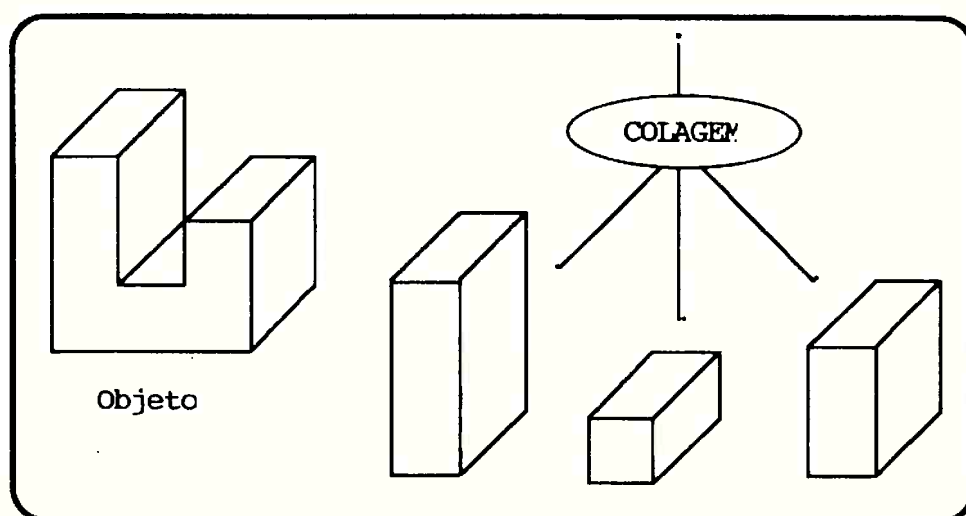


Fig.2.4 Representação por Decomposição Celular.

- na decomposição celular, se as células forem unicamente cúbicas, de tamanho fixo (voxel), tem-se o esquema de representação por Enumeração de Ocupação Espacial. Neste caso, a descrição se dá pelas coordenadas de um ponto (exemplo: centróide) de cada célula.
- caso as células sejam tetraédricas, tem-se a Triangularização Sólida.

2.3.3. Representação por Borda (B-rep)

Um sólido pode ser representado a partir da segmentação de sua superfície de contorno em superfícies mais simples, chamadas faces. Cada face pode ser representada, por exemplo, pelos seus vértices e arestas.

A informação associada aos componentes obtidos tem duas partes: uma é a geometria, que inclui a dimensão e posição do componente no espaço; outra é a topologia, que descreve a conexão entre os diversos elementos.

A estrutura de dados nessa representação é um grafo de adjacências, no qual deve ser armazenada toda informação combinatória envolvendo os elementos geométricos do contorno (figura 2.5).

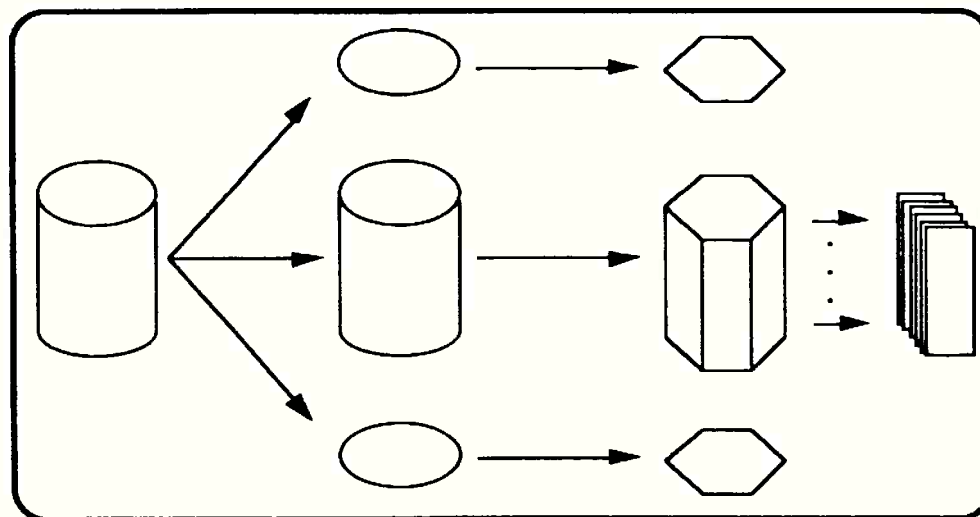


Fig.2.5 Representação por Borda, com aproximação poligonal das faces.

A representação por borda, normalmente, é menos concisa que a CSG correspondente; no entanto, sua grande vantagem consiste

na informação topológica que ela apresenta, daí sua larga utilização em CAD e computação gráfica.

2.3.4. Representação por Varredura

Muitos sólidos podem ser descritos pelo volume gerado do movimento de um conjunto bidimensional ou volume através de uma curva num espaço tridimensional.

Na varredura translacional, define-se uma secção transversal A e o deslocamento de translação, determinando o volume. A varredura rotacional é definida similarmente. A figura 2.6 apresenta exemplos de representação por varredura translacional e rotacional, onde o objeto = (A, varredura).

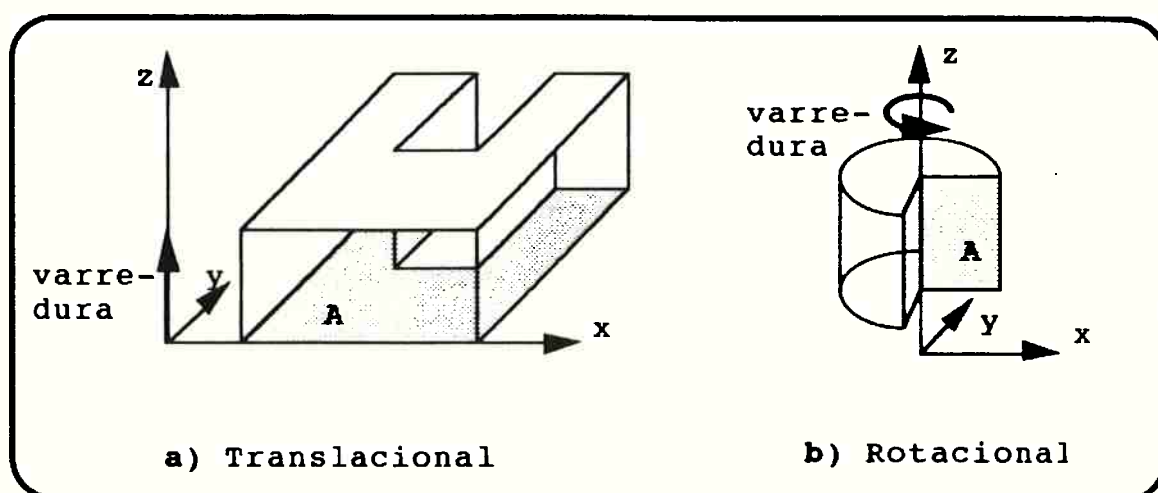


Fig.2.6 Representação por Varredura.

Na varredura generalizada, também chamada de Cilindros (ou Cones) Generalizados, o conjunto bidimensional (secção) ou volume é deslocado ao longo de uma curva espacial arbitrária e

esse conjunto pode variar parametricamente através da curva (figura 2.7).

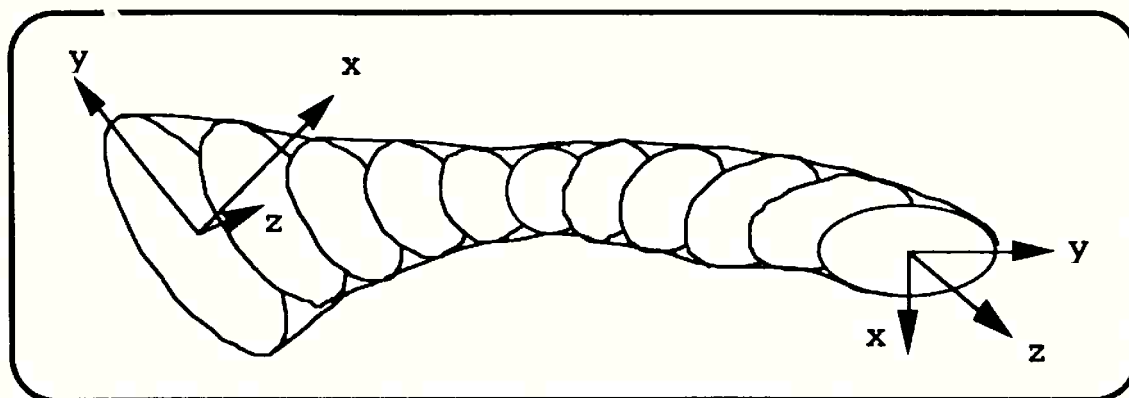


Fig.2.7 Cilindros Generalizados.

2.3.5. Imagem Gaussiana Estendida (EGI)

A Imagem Gaussiana Estendida (EGI: "Extended Gaussian Image") é um mapeamento das normais das superfícies do sólido numa esfera unitária, chamada esfera gaussiana, com a área das superfícies como peso associado a esse mapeamento [Horn 1986] (figura 2.8).

EGI é usada na análise, mas não na síntese da forma das superfícies, por representar uma métrica relacionada à forma dos objetos. Devido a esse fato, esta representação é mais utilizada em aplicações de visão computacional e não de computação gráfica.

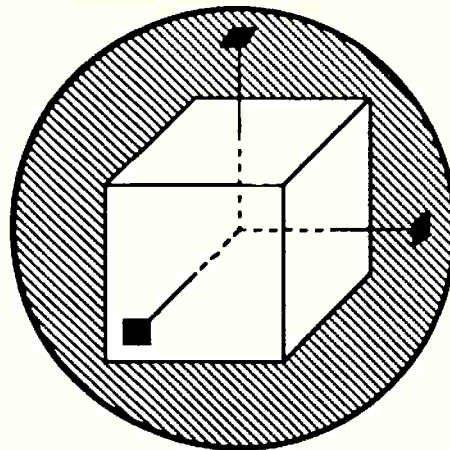


Fig.2.8 Exemplo de EGI (simplificação para melhor visualização: desenhos das projeções das faces na esfera gaussiana, em vez dos vetores correspondentes às respectivas áreas).

Esse tipo de representação possui as seguintes propriedades [Bhanu, Ho 1987, Horn 1986]:

- o centro de massa (centróide) da EGI fica na origem da esfera gaussiana;
- a massa total da EGI é igual à área total da superfície do objeto;
- ela é única para qualquer objeto convexo;
- não depende da posição do objeto, o que permite a determinação da orientação do mesmo, antes de conhecer sua posição;
- rotações no objeto não afetam a distribuição de peso relativa de sua EGI.

2.3.6. Geons

Na representação por Geons ("geometrical ions"), proposta por Biederman [Biederman 1985], as características qualitativas

dos objetos são evidenciadas, em detrimento das características quantitativas, abordadas pela maioria das representações. Em tarefas de reconhecimento, esse tipo de representação qualitativa pode ser muito interessante (e necessária), o que pode não ocorrer em tarefas como simulação, metrologia, fabricação, etc.

A descrição do objeto é feita a partir de catálogos de geons, onde cada elemento é definido por um conjunto único de quatro atributos qualitativos:

- Forma da secção transversal;
- Simetria da secção: pode apresentar simetria rotacional, refletiva ou ser assimétrica;
- Eixo da primitiva: pode ser reto ou curvo;
- Variação de tamanho: pode permanecer constante, aumentar ou diminuir e depois aumentar de tamanho, ao longo do eixo da primitiva.

Alguns exemplos de possíveis primitivas se encontram na figura 2.9.

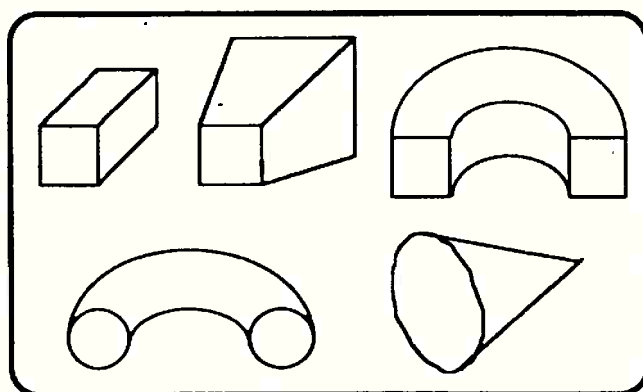


Fig.2.9 Exemplo de primitivas, chamadas geons.

Modelos baseados em primitivas volumétricas possibilitam, de modo simples, a representação de objetos articulados. A nível estrutural, estes modelos são insensíveis a mudanças dimensionais das peças.

Este tipo de representação também possibilita seu uso em reconhecimento de objetos baseado na funcionalidade [Dickinson et al. 1992]. Como desvantagem, apresenta o fato de que é bastante custosa a extração de tais primitivas volumétricas de uma imagem, considerando o uso da referida representação em reconhecimento visual.

2.3.7. Representação Multivista

Um objeto rígido também pode ser representado pela descrição de parte de, ou todas, suas possíveis projeções 2D. Desse modo, atributos característicos de cada projeção podem ser extraídos e evidenciados no modelo.

Esse é um tipo de representação bastante interessante, do ponto de vista do reconhecimento, uma vez que é um modelo centrado na observação, e não no objeto.

A desvantagem do método reside na estrutura de dados gerada, muitas vezes grande e complexa. Além disso, ainda não se encontram algoritmos bem definidos e eficientes voltadas para a representação de objetos não poliédricos. Um exemplo de sua aplicação é encontrado na figura 2.10.

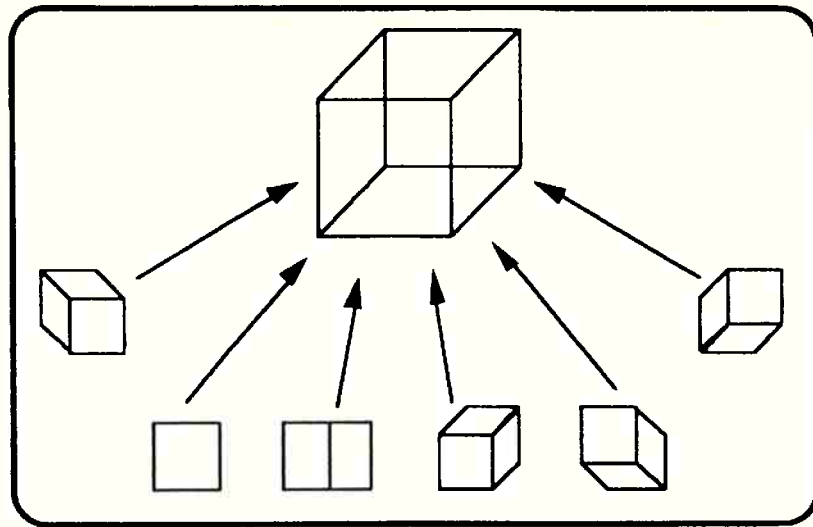


Fig.2.10 Exemplo de Representação Multivista.

2.4. Formação da Imagem

Para a determinação de quais dados captar (e como captá-los), torna-se necessário principalmente o conhecimento de:

- ambiente onde o processo se realizará (condições de iluminação, se pode ser controlado, etc);
- tarefa a ser realizada (inspeção, reconhecimento, metrologia, navegação, etc);
- tipo de objetos que serão tratados (objetos diferentes entre si ou de uma mesma família, forma geométrica simples ou não, etc).

Enfim, estes parâmetros são decisivos para a definição do(s) tipo(s) do(s) sensor(e)s a ser(em) usado(s).

2.4.1. Tipos de Imagens para Reconhecimento

Uma imagem representa uma informação sensorial digitalizada e é traduzida em uma matriz, cujos valores representam algum parâmetro escalar. Para a maioria das aplicações voltadas para reconhecimento baseado na geometria de objetos rígidos em cenas estáticas, assunto de interesse dessa tese, os parâmetros utilizados traduzem informações de reflectividade e/ou distância de superfícies. Desse modo, três categorias principais de imagens *estáticas* e *monocromáticas* são evidenciadas:

- imagens binárias;
- imagens em níveis de cinza;
- imagens com informação de distâncias.

A imagem binária é aquela que possui somente dois valores representando níveis de intensidade luminosa. São de fácil aquisição, armazenamento e processamento. No entanto, são de uso restrito, uma vez que fornecem somente informações sobre a silhueta dos objetos. Os atributos extraíveis são basicamente topológicos e geométricos, porém bidimensionais, tais como: área da silhueta, número de buracos nas peças, segmentos do contorno da silhueta, etc.

As imagens em níveis de cinza, por sua vez, oferecem informações adicionais sobre os objetos que compõem a cena, refletidas na variação da intensidade luminosa, indicando a reflectância das superfícies. Com isso, mesmo com uma só imagem em níveis de cinza, consegue-se inferir algumas características 3D da cena.

Já as imagens com informação de distâncias oferecem como informação adicional a distância (absoluta ou relativa) de cada ponto da superfície dos objetos visualizados ao sensor. Isso permite fácil extração de características geométricas tridimensionais dos objetos, como, por exemplo, formas das superfícies, etc.

2.4.2. Métodos de Aquisição de Imagens

As imagens binárias podem ser adquiridas, por exemplo, a partir de definições de limiares em imagens de níveis de cinza; dessa forma, o que estiver acima do limiar fica com valor 1 e o que estiver abaixo, com 0 (ou vice-versa).

Normalmente, para evidenciar o ponto de limiar, procura-se estabelecer maiores contrastes entre objetos e fundo com, por exemplo, a colocação dos objetos sobre um tampo translúcido, com iluminação difusa sob ele e com a câmera colocada acima da cena (figura 2.11), ou mesmo com a manipulação da informação usando técnicas de processamento de imagens (histogramas, etc).

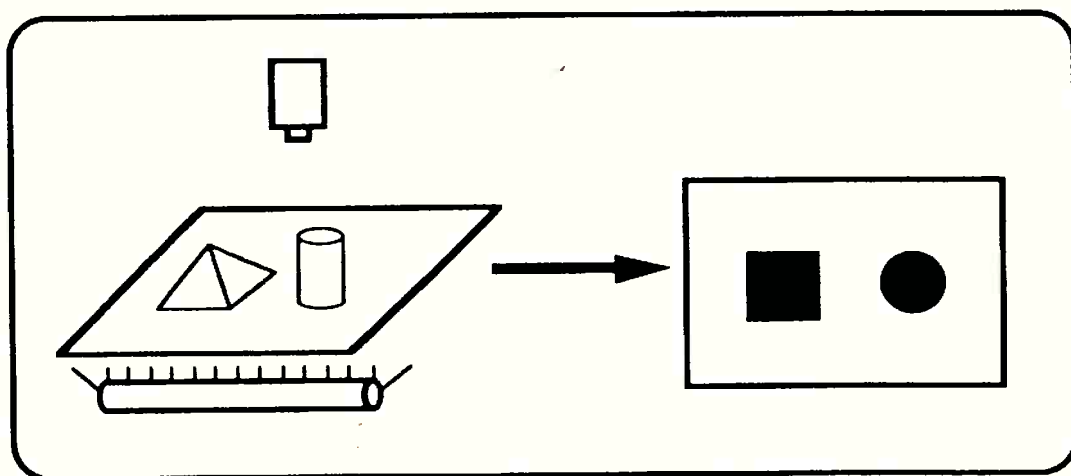


Fig.2.11 Exemplo de aquisição de Imagem Binária.

As imagens em níveis de cinza (monocromáticas), aqui abordadas, são aquelas que apresentam em sua variação de intensidade luminosa, a reflectância das superfícies que compõem a cena. Tem-se ilustrada na figura 2.12 a aquisição de uma imagem de intensidade luminosa por uma câmera, para uma determinada cena.

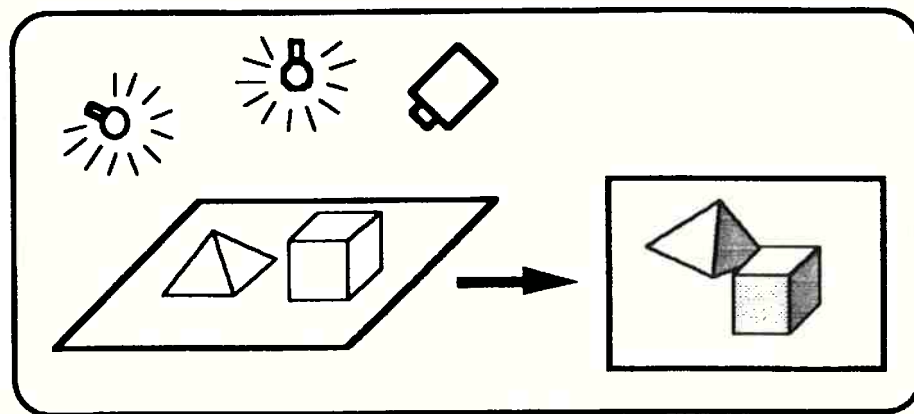


Fig.2.12 Exemplo de aquisição de Imagem de Intensidade Luminosa, em Níveis de Cinza.

As imagens de distâncias podem ser adquiridas de modo passivo (por exemplo, câmeras) ou ativo (por exemplo, sensores a laser); monocular (uma só imagem) ou multiocular (mais de uma imagem). Podem, também, ter agregada a informação de intensidade luminosa. São exemplos:

- Estéreo binocular: semelhante ao sistema visual humano, onde as coordenadas dos planos de imagens de duas câmeras estão relacionadas por uma transformação conhecida; por triangularização determina-se a distância de cada ponto da cena aos pontos correspondentes nas duas imagens, podendo, com isso, extrair a distância de cada ponto da

cena à câmera. Estéreo multiocular atua do mesmo modo, porém fornecendo mais informação para que se solucione o problema de correspondência, existente no binocular (saber qual ponto de uma imagem corresponde a qual ponto da outra) (figura 2.13);

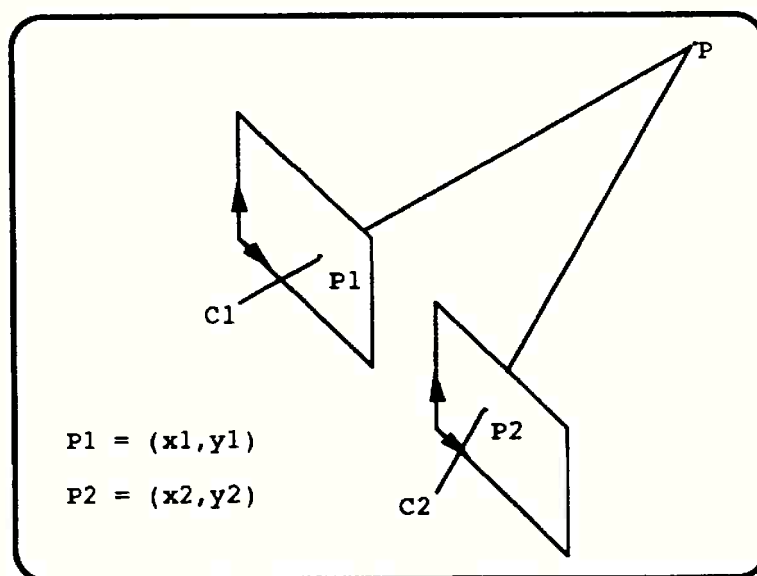


Fig.2.13 Estéreo binocular.

- Estéreo fotométrico: nesta técnica, múltiplas fontes de iluminação são utilizadas para uma mesma cena. Para cada iluminação, capta-se uma imagem monocular de intensidade luminosa. A composição destas imagens resulta numa distância relativa dos pontos da cena;
- Sensores a laser por tempo de retorno, sonares: podem extrair a distância diretamente; são sensores ativos (emissores e captadores) que emitem uma onda (luz, som) e, pelo tempo de retorno destas ondas, podem estimar a distância em que se encontra o ponto da cena onde refletiu a onda. Alguns sensores a laser possuem também

uma câmera junto ao captor, fornecendo uma imagem de intensidade luminosa em correspondência com a imagem de distâncias.

- Luz estruturada: consiste em projetar na cena uma grade conhecida (ou linha de luz, textura sintética, etc); pela deformação efetuada na grade (devido à forma dos objetos na cena), determina-se a distância (figura 2.14);

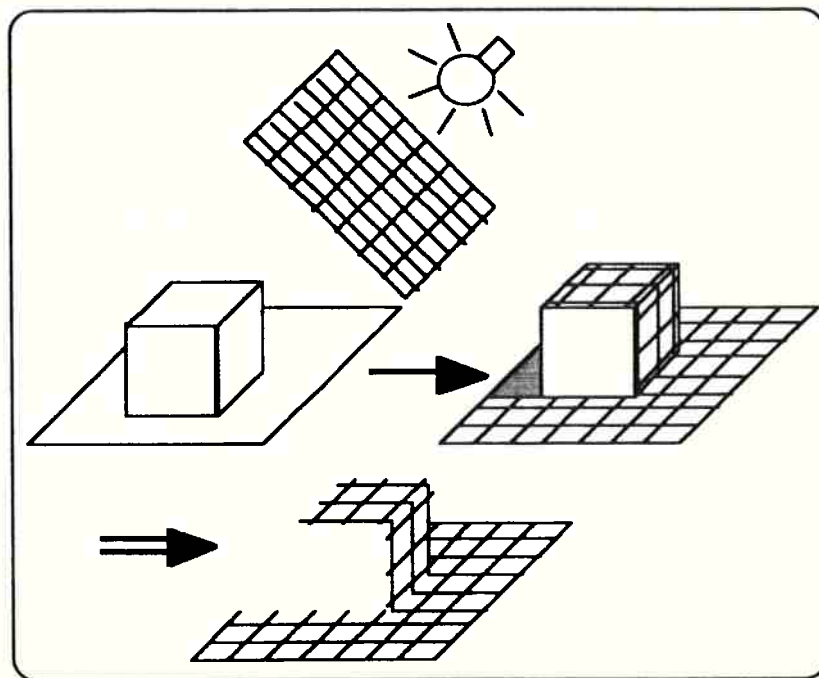


Fig.2.14 Exemplo de aquisição por luz estruturada.

Enfim, existem muitos tipos de sensores, e a análise dos dados por eles fornecidos depende essencialmente da forma como são adquiridos, isto é, torna-se necessário que se modele de modo eficiente o(s) sensor(es) utilizado(s), não esquecendo que diversos tipos podem ser usados simultaneamente, tendo integradas as informações por eles fornecidas. No entanto, vale ressaltar que integração de informação visual é uma área muito

recente de pesquisa, oferecendo, ainda, poucos resultados significativos.

2.5. Definição e Extração de Atributos

Até o momento, foram citados alguns métodos de aquisição de imagens, tipos de imagens e informação que se pode extrair. Para o estabelecimento de quais atributos usar, é necessário conhecer o tipo de dados fornecidos. Assim, referindo-se somente a uma única imagem monocromática de intensidade luminosa de uma cena estática, imagem esta que traduz a reflectância das superfícies dos objetos, serão descritas algumas categorias de atributos extraíveis da imagem.

2.5.1. Atributos Extraíveis de uma Imagem de Intensidade Luminosa em Níveis de Cinza

No processo visual, estruturas físicas tridimensionais de uma cena são projetadas em estruturas bidimensionais de uma imagem. No reconhecimento, o processo precisa ser invertido, isto é, de algum modo, as estruturas físicas 3D precisam ser inferidas de estruturas 2D das imagens.

No entanto, esta inversão do processo de formação da imagem é um problema mal posto. Para solucioná-lo, torna-se necessária a imposição de algumas restrições. Estas restrições são baseadas na observação de distintos eventos físicos e os distintos eventos ocasionados na imagem; por exemplo, um objeto opaco de forma suave produz uma imagem com intensidade luminosa variante

especialmente, mesmo que seja feito de um material com propriedades ópticas uniformes; essa variação espacial da intensidade luminosa resulta num: sombreamento na imagem, o qual fornece informações importantes sobre a forma do objeto.

Diferentes imposições de restrições são possíveis, procurando, na maioria dos casos, estimar:

- posição da fonte de luz;
- posição de uma superfície;
- orientação de uma superfície;
- propriedades de reflectância de uma superfície;
- outros.

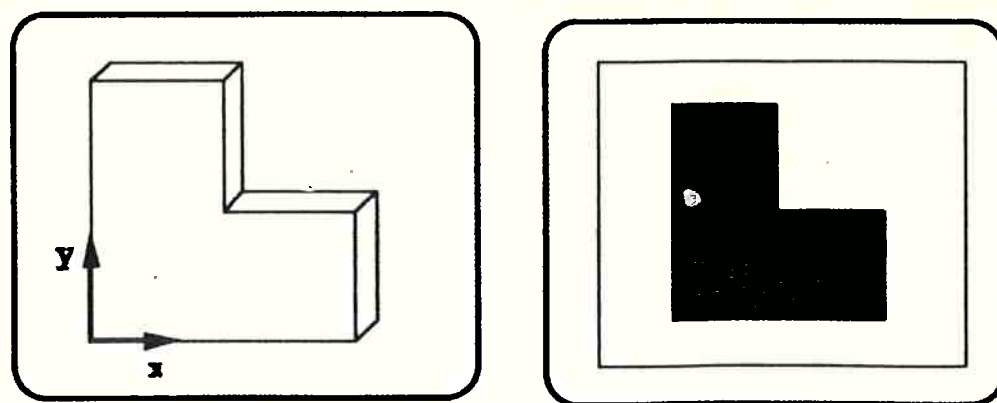
Exemplos de restrições são:

- suavidade de superfícies;
- continuidade de contornos;
- rigidez dos objetos;
- outros.

Considerando as restrições impostas, podem-se definir atributos extraíveis da imagem, os quais determinariam equivalências com os possíveis eventos físicos da cena.

Nesse contexto, variações bruscas na intensidade luminosa de uma imagem podem corresponder a descontinuidades na superfície do objeto, variações de cor, de iluminação, etc. Com base nisso, um fator de extrema importância numa imagem é a definição de suas descontinuidades de intensidade luminosa, possibilitando a detecção de diversos atributos, os quais são equivalentes a várias propriedades referentes à forma dos objetos.

Para exemplificar, observe a figura 2.15, onde se tem um objeto e uma imagem binária do mesmo.



a) Objeto.

b) Imagem binária.

Fig.2.15 Exemplo: um objeto e sua imagem binária.

Pelos conhecimentos anteriores, sabe-se tratar de um único objeto poliédrico, cuja imagem é captada por uma câmera calibrada em relação ao plano xy (figura 2.15-a) e que esta imagem resulta em dois níveis de intensidade luminosa, onde um nível representa o objeto e o outro, o fundo.

Se todos os pontos da imagem representando o objeto forem somados e a correção conhecida pela calibração da câmera for aplicada, tem-se a área da face imageada do objeto. Por outro lado, se os pontos de descontinuidade luminosa na imagem forem determinados, pode-se definir o contorno da silhueta, seu perímetro, etc.

São muitos os atributos extraíveis de uma imagem, onde procura-se definir aqueles que sejam invariantes com os pontos de observação, já que geralmente os objetos se encontram em posições desconhecidas na cena, em relação ao observador (câmera).

Dentre os atributos básicos normalmente usados, referentes à forma (geometria) dos objetos, citam-se:

- atributos globais: área, perímetro, número de Euler, centro de área, direção dos eixos principais, etc;
- atributos locais: segmentos do contorno (retilíneos, curvilíneos, elipsoidais, etc), junções de segmentos (tipo da junção, número de segmentos que a compõe), etc.

A partir desses atributos básicos, pode-se inferir características mais complexas, como:

- superfícies,
- primitivas volumétricas,
- EGI (variação da área de uma superfície com sua orientação),
- outros.

Vale mais uma vez ressaltar que a definição dos atributos extraíveis de uma imagem é altamente dependente dos seguintes fatores:

- restrições impostas para a inversão do processo de formação da imagem;
- conhecimento *a priori* dos objetos que compõem a cena;
- conhecimento *a priori* da tarefa a ser executada, mais explicitamente, quais informações pretende-se extrair da imagem;
- conhecimento do processo de formação da imagem (sensor utilizado, iluminação, etc).

2.6. Geração de Hipóteses

Reconhecimento pode ser interpretado como sendo uma busca por correspondências, isto é, como associar componentes ou parâmetros dos dados sensoriais aos correspondentes componentes ou parâmetros dos modelos dos objetos.

A correspondência define:

- quais objetos correspondem aos dados e
- qual é a localização (posição e orientação no espaço) dos objetos.

Para efetuar esta correspondência, é necessário que parâmetros similares de dados sensoriais e dos modelos sejam comparados, para que hipóteses de correspondência entre imagem e modelo sejam geradas.

Com relação ao tipo de atributos usados, pode-se diferenciar duas abordagens distintas:

- correspondência baseada em atributos globais para inferir identidade e localização;
- correspondência baseada em atributos locais (parâmetros espacialmente localizados).

Considerando o controle do processo de busca, pode-se também diferenciar três tipos de abordagens para a correspondência:

- busca guiada pelos modelos,
- busca guiada pelos dados,
- busca integrada.

2.6.1. Classificação da Correspondência de Acordo com o Tipo de Atributos Usados

a) Correspondência usando atributos globais

Atributos globais são aqueles que dependem do objeto por inteiro para serem definidos. A correspondência baseada nesses atributos envolve encontrar uma transformação de um modelo para uma imagem, sem que inicialmente se determine a correspondência entre partes individuais dos modelos e dos dados.

O domínio comum utilizado para efetuar a correspondência é um espaço de vetores de parâmetros. A idéia consiste em computar conjuntos de parâmetros globais dos modelos e dos dados sensoriais, conjuntos estes representados por vetores de parâmetros.

O vetor da imagem é então comparado com os vetores dos modelos, determinando o modelo que melhor corresponda à imagem.

Finalmente, propriedades de alguns parâmetros são usadas para a determinação da localização do objeto na imagem.

b) Correspondência usando atributos locais

Atributos locais são aqueles relacionados a pequenas partes individuais dos objetos. Assim, em abordagens baseadas nesses atributos, inicialmente é estabelecida uma correspondência entre partes individuais do modelo e da imagem, procurando, com isso, poder também efetuar correspondências apropriadas, mesmo na presença de oclusão de objetos e na presença de dados espúrios.

Em oposição à abordagem por correspondência global, são usadas relações simbólicas entre atributos locais, e não propriedades métricas globais, com o intuito de se inferir a forma do objeto. Devido a essas características, é a abordagem usada na maioria das aplicações.

2.6.2. Classificação da Busca pelo Tipo de Controle Efetuado

a) Busca guiada pelos modelos

Neste caso, examina-se cada modelo seqüencialmente, aceitando como possível solução somente aquele que apresentar uma correspondência consistente entre seus atributos e os dos dados. Este tipo de abordagem pode não ser computacionalmente eficiente quando a biblioteca de modelos for grande.

b) Busca guiada pelos dados

A busca é dita guiada pelos dados quando o processo evolui a partir da informação da imagem para níveis de descrição mais elevados, isto é, as primitivas extraídas da imagem é que são usadas para direcionar a busca no conjunto de modelos.

c) Busca integrada

Na busca integrada, as duas abordagens anteriores são misturadas, primeiramente analisando os dados até um certo nível, suficiente para selecionar alguns modelos da biblioteca

(fazendo uma pré-seleção rápida, podendo ser por intermédio da definição de classes), para então, guiada pelos atributos dos modelos pré-selecionados, continuar a análise dos dados, gerando correspondência consistente com a forma e uma transformação rígida possível. Portanto, na solução integrada, a busca é inicialmente guiada pelos dados, para, a partir de um subconjunto selecionado de modelos, tornar-se guiada pelos modelos.

Nesta solução de busca integrada, a cada novo atributo efetua-se um teste para confirmar a satisfação de uma transformação rígida e, adicionalmente, o teste das correspondências parciais efetuadas até então, para confirmar se a transformação rígida de todas as correspondências é consistente. No final, tem-se como resultados o modelo de um ou mais objetos e suas respectivas transformações rígidas efetuadas para a imagem analisada.

Normalmente, uma busca completa possui um alto custo computacional, mesmo para problemas bem simples. Por isso, lança-se mão da utilização de heurísticas para terminar a busca assim que um resultado bom o suficiente seja alcançado. Outros recursos utilizados são: ordenação da busca, enfoque em atributos privilegiados (ou salientes), etc.

2.7. Verificação

O processo de verificação é efetuado quando alguma heurística for utilizada para a interrupção da busca por correspondência. Essa interrupção é feita quando uma ou mais

hipóteses forem levantadas em relação aos modelos e às transformações correspondentes à localização dos objetos na cena. Nesse caso, utiliza-se a verificação para negar, ou confirmar e aprimorar, as hipóteses levantadas.

Esse processo normalmente requer a síntese de uma imagem para cada hipótese, a qual é comparada com a imagem inicial (analisada). A imagem gerada pode ser menos densa e completa; geralmente, somente os contornos visíveis dos objetos são utilizados.

Deste modo, pode-se comprovar se o modelo em hipótese está correto e efetuar um aprimoramento da localização do objeto.

Entende-se por localização de um objeto como sendo a definição da transformação do referencial do objeto para um sistema de coordenadas conhecido qualquer, utilizado como referencial global. Normalmente, em reconhecimento baseado em modelos, tem-se um referencial conhecido, em relação ao global, para cada modelo; com isso, a localização que se deseja é a transformação do referencial do objeto na imagem em relação ao referencial de seu respectivo modelo.

Para o processo de aprimoramento da localização do objeto, existem diversas propostas, como:

- linearizar as equações de projeção e aplicar iterativamente o método de Newton [Lowe 1987] (figura 2.16);
- utilizar um modelamento matemático do tipo explícito, permitindo a determinação simultânea dos seis parâmetros de orientação, tratado por filtragem Kalman [Tomaselli 1993];

- dividir o cálculo dos parâmetros de localização em dois estágios, primeiramente as rotações e, depois, as translações [Liu et al. 1990, Faugeras, Hebert 1986];
- outros.

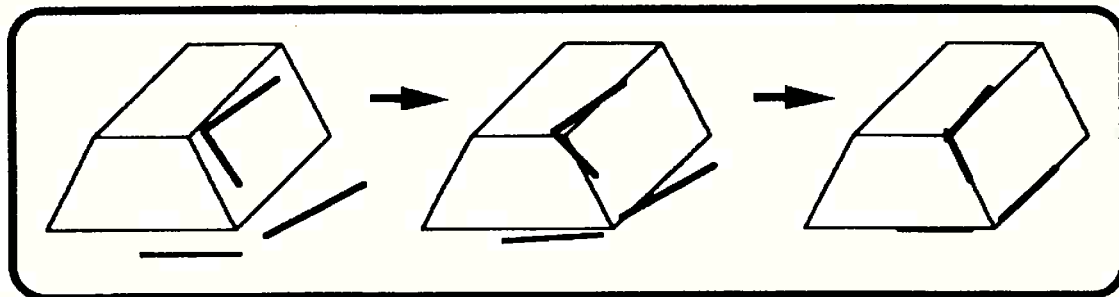


Fig.2.16 Utilização do método de Newton para localização (figura adaptada de [Lowe.1987]).

3. O ESTADO DA ARTE EM RECONHECIMENTO E UMA NOVA PROPOSTA

3.1. O Estado da Arte em Reconhecimento 3D Baseado em Modelos

Reconhecimento de objetos 3D é uma área muito ampla, onde qualquer revisão da literatura que se faça certamente não será exaustiva.

As abordagens que se enquadram nesse tema tratam desde objetos isolados e pré-segmentados, até cenas complexas envolvendo vários objetos simultaneamente; podem ter como entrada somente uma imagem de intensidade luminosa ou então uma série de informações multi-sensoriais.

Nenhum dos sistemas de reconhecimento 3D apresentados até hoje na literatura soluciona completa e irrestritamente o problema do reconhecimento, por este ser um problema extremamente complexo (veja discussão nos capítulos anteriores).

A maioria dos sistemas propostos é voltada para aplicações específicas, onde restrições são impostas, por exemplo, quanto às superfícies permitidas aos objetos, à topologia da cena, posicionamento da câmera, etc.

Os sistemas de reconhecimento existentes diferem entre si basicamente nos seguintes tópicos:

- complexidade dos atributos usados para indexação da base de dados (locais, intermediários [Aggarwal, Chien 1989, Helmke et al. 1990; Horaud, Brady 1989; McLean et al.

1992], agrupamentos [Dolan, Weiss 1989; Du et al. 1991; Horaud et al. 1990; Saroldi 1988], descrições de superfícies [Fan et al. 1988], primitivas volumétricas, etc);

- esquema de representação utilizado para modelar os objetos;
- método de indexação dos modelos (busca linear na base de dados, „hashing“ [Wolfson 1990] ou esquemas de votos [Vayda, Kak 1991], etc);
- método de correspondência (por alinhamento [Ullman, Basri 1989], baseado em verificação, etc).

3.1.1. Abordagens Pioneiras

a) Desenhos de linhas

O trabalho pioneiro na área de reconhecimento 3D é atribuído a Roberts [Roberts 1965], desenvolvido em meados da década de sessenta. Seu sistema identifica objetos poliédricos, além de determinar sua posição e orientação no espaço, a partir de uma única imagem de intensidade luminosa. Foi feita a restrição de que os objetos da cena ou seriam poliedros básicos, escolhidos como sendo cubo, cunha e prima hexagonal, ou então composição destes poliedros básicos.

Após a digitalização da imagem, um operador é utilizado para detectar variações bruscas na sua intensidade luminosa, resultando em segmentos. Segmentos colineares são agrupados em segmentos maiores, para produzir um desenho de linhas perfeito da cena (“line drawing”).

A partir daí, os modelos das formas poliédricas básicas são transformados, procurando-se ajustá-los ao desenho de linhas extraído da imagem

Objetos ou partes já reconhecidas dos objetos são retiradas da imagem e o processo prossegue, até que todas as linhas extraídas da imagem tenham sido rotuladas, isto é, até que a correspondência entre todas as linhas da imagem e as arestas dos modelos tenha sido estabelecida.

Após o reconhecimento de um objeto, o sistema é capaz de desenhá-lo a partir de qualquer outro ponto, a fim de demonstrar que interpretou corretamente a forma do objeto.

O sistema de Roberts é bastante completo e inspirou o desenvolvimento de outros trabalhos baseados em rotulações de linhas e junções para detecção de regiões poligonais [Kanade 1981; Morimoto 1991; Sugihara 1986; Waltz 1975; etc]. Para uma boa revisão sobre esse tipo de abordagem, consultar Morimoto [Morimoto 1991].

Guzman [Guzman 1968] desenvolveu uma abordagem sistemática na interpretação de desenhos de linhas poliédricas, apresentando uma teoria sobre como estabelecer ligações entre duas regiões (faces) de um mesmo objeto, baseada na análise da formação de junções, rotulando os tipos possíveis de junções (evidência local). Essas evidências locais são agrupadas para produzir interpretações válidas. Essa abordagem não utiliza modelos, sendo totalmente dirigida pelos dados. Funcionou bem para muitos exemplos, mas, para suplantiar falhas ocorridas em cenas simples, gerou regras heurísticas.

Huffman e Clowes continuaram a abordagem de Guzman, procurando suplantar as falhas com uma análise completa dos tipos de junções, formando um catálogo de rotulações possíveis para as junções. Numa imagem, uma junção é rotulada. Se esta junção estiver conectada (através de uma aresta) a outra junção, as rotulações de ambas têm que ser coerentes. Desse modo, também dirigida somente pelos dados e o catálogo de rotulações, a reconstrução da forma do objeto vai sendo desenvolvida até gerar resultados para o reconhecimento.

Waltz estendeu o catálogo de Huffman e Clowes para introduzir as restrições formadas por sombreamento na imagem.

Trabalhos posteriores expandiram o trabalho de Waltz em várias direções [Kanade 1981; Sugihara 1986], porém existem algumas configurações na imagem onde, localmente, as rotulações de junções mostram consistências, porém, a nível global, mostram-se inconsistentes. Um modo de ultrapassar esse problema é explorando algumas restrições globais. Isso é possível quando se usa o conhecimento dos modelos para eliminar interpretações globalmente inconsistentes.

No entanto, todos esses sistemas pioneiros foram muito importantes por terem abordado problemas fundamentais relacionados ao reconhecimento 3D a partir de imagens que representam a intensidade luminosa captada por câmeras, porém a maioria exige desenhos de linhas perfeitos. Mesmo assim, criaram um paradigma ainda dominante, no qual o ciclo do reconhecimento seria estabelecido por: extração de atributos, geração de hipóteses para correspondência, verificação da hipóteses e previsão das conseqüências.

b) Sistema ACRONYM

A partir da década de oitenta, surgiram vários sistemas de reconhecimento 3D.

Brooks et al. [Brooks 1983] desenvolveram o sistema ACRONYM, caracterizado pela flexibilidade e modularidade de seu projeto, o qual procurou apresentar qualidades independentes do domínio de aplicação. O sistema utiliza modelos volumétricos para os objetos (cilindros generalizados; veja capítulo anterior). Entretanto, modelos gerados por cilindros generalizados se tornam computacionalmente intratáveis quando formados por funções arbitrárias de deformações e formas genéricas de seção transversal. Devido a isso, ACRONYM restringiu as seções para figuras planas contornadas por retas ou arcos circulares e o eixo de deslocamento, para ser somente retilíneo ou circular, facilitando a predição de atributos.

ACRONYM modela um objeto complexo por intermédio de um conjunto de cilindros generalizados (restringidos), descrito numa estrutura chamada *grafo de objeto*. Classes genéricas de objetos e objetos específicos são relacionados por meio de especificações algébricas, entre classes e subclasses, numa estrutura nomeada *grafo de restrições*.

Com base nestes grafos, um planejador de predições monta uma terceira estrutura de dados, chamada *grafo de predições*, onde são organizadas as projeções dos cones generalizados permitidos.

Estes três tipos de grafos compõem a estrutura de dados conhecida pelo sistema (modelos), em relação aos objetos, classes de objetos, composições por geons básicos e suas

possíveis visualizações. É também fornecido ao sistema o modelo da câmera utilizada.

A partir da imagem digitalizada, são extraídos segmentos de contorno da imagem, os quais são agrupados em elipses ou faixas (do inglês, "ribbons"). A imagem é então descrita em termos dessas estruturas 2D básicas, num *grafo de descrição*.

Uma busca por correspondência é então realizada, acessando o grafo de descrição (atributos da imagem) e o grafo de predições (atributos dos objetos), gerando o *grafo de interpretação*. A busca evolui por propagação de restrições (cada correspondência gera restrições, que são propagadas).

Apesar do projeto do sistema ACRONYM ter sido bastante ambicioso, sua implementação não seguiu a generalidade pretendida e seus resultados, tipicamente reconhecimento de jatos em imagens aéreas de aeroportos, não são surpreendentes, sendo mesmo comparáveis aos adquiridos por sistemas bem mais simples [Besl, Jain 1985]. Um dos problemas de ACRONYM é ser um sistema de malha aberta, sem realimentação, restringindo sua robustez ao seu componente mais limitado; por isso, apresenta dificuldades quando os dados de entrada não são bastante bons.

Com a relativa decepção gerada pelo pretensioso ACRONYM, os sistemas de reconhecimento que o seguiram perderam em generalidade, voltando-se a aplicações específicas, onde o problema fica mais restrito, podendo criar soluções direcionadas e de melhores resultados, para a dada aplicação.

3.1.2. Abordagens que Utilizam Imagens de Profundidades

O trabalho de Roberts [Roberts 1965] gerou resultados tão surpreendentes para a época, que se criou uma euforia em relação ao reconhecimento visual utilizando uma única imagem, captada por uma câmera, representando a intensidade luminosa refletida por uma cena.

No entanto, após aproximadamente uma década, os pesquisadores demonstraram um certo desapontamento porque, na verdade, pouco tinha sido evoluído, em termos de resultados práticos, em relação ao trabalho original. Começaram, então, diversos questionamentos e compreenderam que, na realidade, o problema abordado era muito mais complexo do que o inicialmente suposto.

Nesta mesma época os sensores de profundidade começaram a ficar mais acessíveis (em termos de custos), apresentando, ainda, aumento da qualidade da informação extraída, chamando, com isso, a atenção de diversos pesquisadores. Além disso, é fato que a imagem de profundidades apresenta informação menos ambígua que a de intensidade, uma vez que a geometria da cena se encontra ali diretamente codificada.

Neste contexto, surgiram diversas propostas, principalmente para aplicações industriais, utilizando a informação de profundidade das cenas, em vez de simplesmente a informação de intensidade, facilitando o problema de reconhecimento 3D. Dentro dessa abordagem, são citados os representativos trabalhos de Grimson e Lozano-Perez, Ikeuchi e Kanade, Faugeras e Hebert, Bolles e Haraud, além de Oshima e Shirai.

a) Sistema RAF

O sistema RAF (Recognition and Attitude Finder) [Grimson, Lozano-Perez 1985] efetua o reconhecimento em dois passos principais:

- desenvolvimento de uma *árvore de interpretação* e busca por correspondências corretas nela, utilizando restrições geométricas locais para direcionar a busca;
- teste de consistência global.

RAF utiliza somente objetos poliédricos, os quais são modelados por intermédio das equações dos planos, que representam as superfícies planas do objeto e suas dimensões (não usa informação de conectividade).

Como entrada, possui dados esparsos, indicando posições 3D de superfícies planas dos objetos e suas normais associadas, captados tipicamente com sensores táteis, mas também pode ser usado com imagens de sonares e luz estruturada. Inicialmente, o sistema RAF só permitia cenas de um único objeto, sem oclusão; mais tarde permitiu oclusão, mas a complexidade da busca cresceu exponencialmente. São extraídos pontos de interesse da imagem captada.

Uma *árvore de interpretação (IT)* é criada, representando a busca por correspondência com cada modelo (uma IT é construída para cada modelo testado), onde cada nó representa uma correspondência possível do ponto extraído da imagem com uma superfície do objeto. Utiliza-se busca em profundidade na IT. Porém, a busca só é desenvolvida para os nós ditos *corretos*, que são aqueles que satisfazem restrições geométricas locais; por

exemplo, um par de pontos que deve pertencer a uma mesma face podem distanciar entre si até o máximo permitido pelas dimensões da referida face, dadas no modelo.

Para tratar a generalização introduzida mais tarde, de permitir vários objetos, com oclusão, a cada expansão de um nó da IT foi introduzido um nó especial, para possível expansão. Quando um ponto da imagem é correspondido a esse nó especial, indica que o ponto é inconsistente com o modelo. Essa generalização fez com que a busca crescesse significativamente.

Para aumentar a eficiência, alguns recursos foram usados:

- uso de segmentos de bordas para a correspondência, em vez de simplesmente pontos;
- utilização de heurísticas para guiar a busca e para terminá-la prematuramente;
- direcionamento da busca para os nós que possuam transformações com maior evidência de correspondência correta; essas evidências são determinadas a partir de aglomerações num espaço paramétrico de possíveis transformações, cujos eixos são parâmetros quantizados de transformação.

No entanto, estes recursos ressaltaram a característica dicotômica da abordagem: eficiência versus qualidade.

b) Sistema proposto por Ikeuchi e Kanade

No sistema proposto por Ikeuchi e Kanade [Ikeuchi, Kanade 1988a; 1988b], foi utilizado o processo estéreo fotométrico dual como processo de imageamento, onde duas câmeras são usadas para a aquisição de múltiplas imagens da mesma cena, com câmeras na

mesma posição, porém com diferentes condições de iluminação. Neste caso, não há problema de correspondência entre as imagens, determinando a distância absoluta, em cada ponto (informação 3D).

A tarefa para a qual o sistema foi destinado consiste em retirar um único objeto, num amontoado de amostras empilhadas do mesmo objeto. Com isso, apenas o modelo de um único objeto é usado e não se tem necessidade de tratar oclusões, uma vez que o objeto de interesse se encontra no topo do amontoado. Foi utilizado objeto poliédrico.

Numa fase de pré-processamento, o objeto é modelado em termos de suas propriedades geométricas e fotométricas, assim como o sensor, em termos de detectabilidade e confiança nos dados. São então determinadas suas *aparências* (também chamadas de *vistas* ou *projeções*), as quais são aglomeradas, formando o *conjunto mínimo dos aspectos* do referido objeto. Aspectos são definidos como sendo uma classe topologicamente equivalente de aparências de um dado objeto. Esses aspectos são organizados numa estrutura, onde são definidos dois tipos de mudanças na forma:

- linear, quando a mudança ocorre dentro de um mesmo aspecto;
- não linear, quando entre aspectos.

Para discriminação dos aspectos, dentro da mudança não linear da forma, são usados momentos de inércia e geometrias locais. EGI e outros são usados para determinar mudança linear, quando se deseja definir qual aparência, daquele aspecto, gerou os referidos atributos da imagem.

Os problemas apresentados por este sistema são:

- não trata oclusões;
- não é simples generalizar para objetos não poliédricos;
- o grafo de aspectos de um objeto pode ser muito grande, se o objeto for relativamente complexo;
- trata de um único objeto por vez, não considerando o problema com diferentes objetos em cena.

c) Sistema proposto por Faugeras e Hebert

Um dos bons resultados apresentados na literatura de reconhecimento 3D foi adquirido pelo sistema de Faugeras e Hebert [Faugeras, Hebert 1986], onde são realizadas correspondências geométricas entre superfícies primitivas (no caso, quádricas).

Os dados usados também refletem a profundidade de cada ponto na cena. Os objetos são modelados usando representação por borda. As primitivas usadas são lineares (pontos, linhas e superfícies), onde:

- pontos: vértices ou centros de simetria;
- linhas: contornos internos ou eixos de simetria;
- superfície: planos ou superfícies quádricas;

No entanto, a implementação realizada utilizou a aproximação poliédrica dos objetos; portanto as linhas são segmentos de reta e as superfícies, planos.

No reconhecimento também utiliza, como Grimson/Lozano-Perez, uma árvore de interpretação IT, com extensão para o nó especial

para tratar oclusões. A IT é explorada usando a restrição de rigidez.

Inicialmente, a busca por correspondência evolui até que se faça uma estimativa da transformação rígida T . Essa estimativa exige um número mínimo necessário de primitivas em correspondência, por exemplo:

- para pontos: 3 primitivas para a translação (t) e 3 para a rotação (r);
- para retas: 2 para t e 2 para r (não paralelas);
- para planos: 3 para t e 2 para r (planos independentes).

A partir daí, com uma transformação T estimada, entra-se na fase de predição-verificação.

Para cada passo na árvore, correspondendo a um reconhecimento parcial $(M_1, S_i), (M_2, S_i), \dots (M_k, S_i)$, a melhor transformação T_k do modelo é computada para as primitivas. M_j representa uma primitiva (atributo) do modelo e S_i , da cena. S_i pode ser o referido nó especial, indicando que o M_j correspondente não está presente na imagem (está oculto ou não foi captado). Define-se, então, no modelo, a próxima primitiva M_{k+1} a buscar, aplicando T_k a ela. Só considera para correspondência as primitivas da cena que estiverem próximas a $T_k.(M_{k+1})$. Além disso, a consistência global com relação à restrição de rigidez deve ser mantida (figura 3.1). A localização, efetuada na fase de verificação, foi desmembrada em duas: primeiro, cálculo da rotação, depois, da translação; foi utilizado o cálculo de um erro composto para medida da qualidade da resposta, com os erros parciais de cada módulo do sistema contribuindo nesta medida.

d) Sistema 3DPO

O sistema de Polles e Horaud [Bolles, Horaud 1986], chamado 3DPO ("Three-Dimensional Part Orientation system"), é voltado para tarefas onde um objeto deve ser retirado do topo de um amontoado formado por objetos idênticos empilhados.

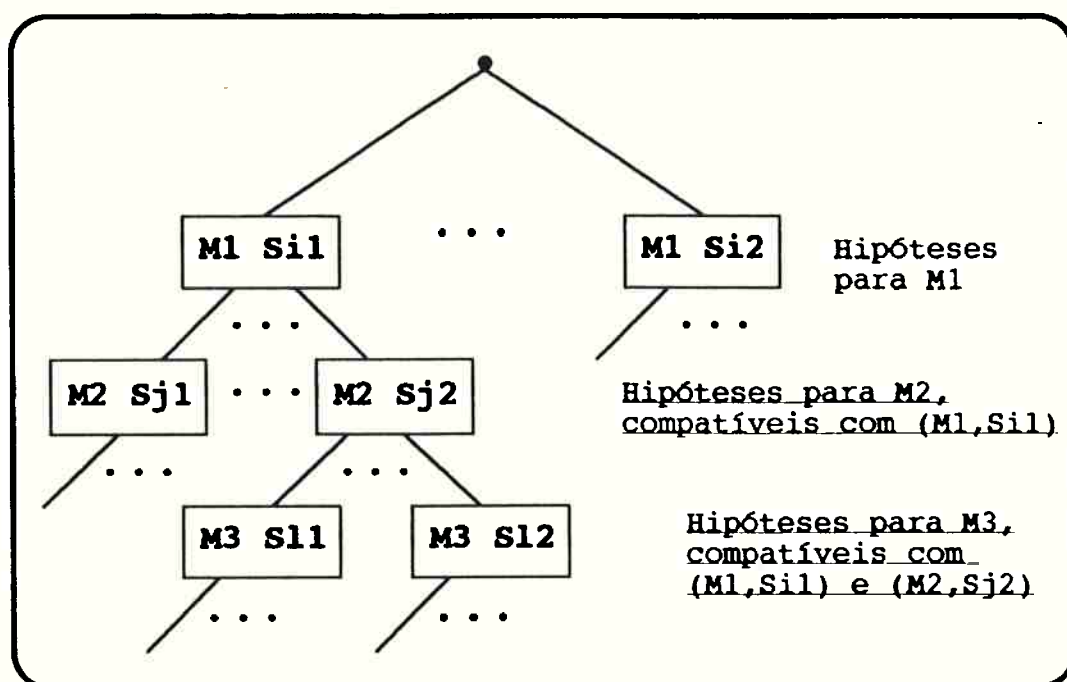


Fig.3.1 Princípio da busca em árvore utilizado por Faugeras e Hebert.

Preferencialmente, esses objetos devem ser moderadamente complexos, apresentando vários atributos característicos. Isso faz com que a seleção de atributos, para serem usados no processo de correspondência, seja facilitado, podendo escolher dois ou três deles altamente salientes e ricos em informação.

O modelo completo do objeto é composto por:

- uma árvore de classificação de atributos;
- modelo resultante de um sistema CAD, ligeiramente modificado (volume-superfície-aresta-vértice);
- lista dos trechos planos de superfície;
- modelo por estruturas aramadas.

O reconhecimento se dá do seguinte modo: inicialmente, a imagem de profundidade da cena é captada. Seguindo a seleção feita no pré-processamento, são extraídos os atributos da imagem, como arcos no final de um cilindro, reta de intersecção de dois planos, etc.

Com o auxílio da árvore de classificação construída para o referido objeto, na fase de pré-processamento, procura outros atributos (pré-selecionados e ordenados na árvore), a partir do atributo saliente determinado, fazendo com que a busca por correspondências evolua até que uma hipótese de identificação e localização seja efetuada (na verdade, confirmação de identificação, uma vez que trata de somente um objeto).

O processo de verificação faz, com o auxílio do modelo CAD estendido, a predição dos dados de profundidade que deveria ter na imagem, caso a hipótese seja correta. Compara os dados preditos com os medidos, refinando ou recusando a hipótese.

O interessante nessa abordagem é justamente como a aglomeração de atributos é dirigida e efetuada, porém fica completamente dependente da extração dos atributos salientes da imagem. Como o interesse é tratar o objeto do topo do amontoado, onde normalmente não ocorre oclusão, este problema (de buscar atributos salientes), não se torna tão relevante.

e) Sistema proposto por Oshima e Shirai

O sistema de reconhecimento de objetos 3D proposto por Oshima e Shirai [Oshima, Shirai 1981] é baseado na imagem de profundidades adquirida com o auxílio da técnica de luz estruturada.

Os pontos da imagem são agrupados em trechos planares de superfície, que, por sua vez, são agrupados em regiões, classificadas como planares ou curvas (superfícies quádricas). É então gerada uma descrição da cena, em termos dos atributos das regiões.

Os atributos das regiões são:

- tipo: curva ou planar;
- equação da superfície no espaço (3D);
- atributos 2D globais: perímetro, área, raios mínimos e máximos da região em relação ao seu centróide, etc;
- centróide 3D;
- número de regiões adjacentes.

As relações de adjacência entre as regiões são:

- tipo de intersecção: côncava ou convexa;
- ângulo entre regiões, ou entre seus planos tangentes, caso sejam curvas;
- posição relativa entre centróides.

Existe uma fase de aprendizado do sistema, onde os modelos são montados automaticamente a partir da descrição gerada para cada vista de cada objeto.

Na imagem, uma região é escolhida como sendo a *semente*. Os critérios para esta escolha são:

- tipo: preferência para planares;

- área: preferência para as maiores;
- número de regiões adjacentes: preferência para aquelas com maior número de vizinhos.

É então feita uma busca linear nos modelos, selecionando aquele que tiver uma região com as mesmas propriedades da região semente. Feito isso, é executada uma busca por correspondência dirigida pelo modelo selecionado. O critério usado para evoluir a busca é *proximidade de regiões*. No final, obtém-se a confirmação (ou não) da adequação do modelo candidato. O processo é repetido até que todas regiões da imagem tenham sido processadas.

3.1.3. Abordagens que Utilizam Imagem de Intensidades

Apesar do crescente número de contribuições voltadas para reconhecimento 3D usando imagens de profundidades, pesquisas continuaram sendo feitas em reconhecimento visual realizado diretamente a partir de imagens 2D, sem qualquer reconstrução preliminar das informações de distância.

Essa abordagem contrasta com a maioria das pesquisas em visão computacional realizadas no início da década de oitenta, e também com muitas teorias sobre a visão humana, as quais assumem que nosso conhecimento tridimensional da aparência de um objeto é comparado com algum tipo correspondente de reconstrução 3D dos dados de uma imagem. No entanto, apesar de processos de medição de profundidades, como visão estereoscópica, e de interpretação de movimentos serem aspectos muito importantes da visão humana, muitos pesquisadores argumentam que é mais provável que, para

típicas situações de reconhecimento, esses processos sejam, na verdade, realizados após o reconhecimento, e não antes dele.

Baseados nessas considerações, várias abordagens foram propostas, onde o reconhecimento de objetos tridimensionais é realizado tendo como única fonte de dados uma imagem bidimensional de intensidade luminosa refletida por uma cena. Nesta linha de pesquisa, enquadram-se os trabalhos de Huttenlocher e Ullman, Horaud, Thompson e Mundy, Lowe, entre outros.

a) Sistema ORA

O sistema ORA ("Object Recognition by Alignment"), proposto por Huttenlocher e Ullman [Huttenlocher, Ullman 1988], consiste em determinar possíveis alinhamentos de um modelo com uma imagem, para então verificar cada posição e orientação em hipótese, através da transformação do modelo em coordenadas da imagem.

Para realizar a transformação de alinhamento, são usadas duas classes de atributos (figura 3.2):

- Classe I: somente um desses atributos é suficiente para fazer o alinhamento. Exemplo: uma aresta completa conectada a dois segmentos de arestas;
- Classe II: são necessários dois atributos para o alinhamento. Exemplo: arco circular.

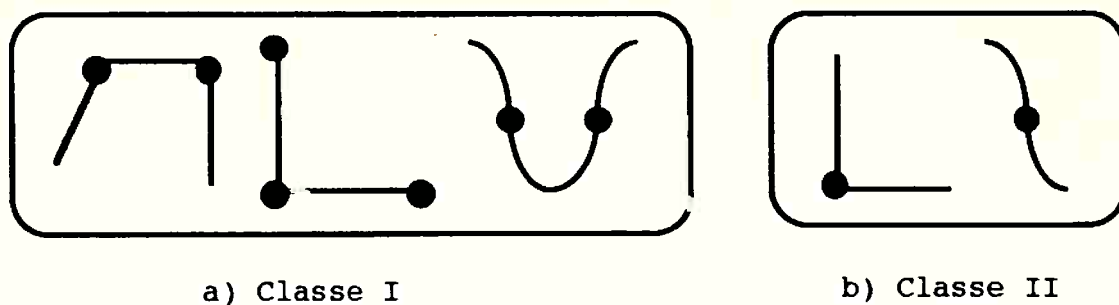


Fig.3.2 Classes de atributos extraídos pelo sistema ORA (adaptação das figuras 4 e 5 de [Huttenlocher, Ullman 1988]).

O sistema extrai atributos (classe I e II) de uma imagem representando a intensidade luminosa refletida por uma cena e captada por uma câmera. Conjuntos mínimos de atributos são comparados com os atributos de mesmo tipo, de todos modelos. Para cada modelo, e para cada atributo em correspondência, é calculada uma possível transformação de alinhamento.

Usando a transformação de alinhamento calculada, todos os outros pontos de cada modelo são transformados, projetados e comparados com os pontos da imagem. Os alinhamentos são então ordenados de acordo com o número de atributos, de mesmo tipo e orientação, em correspondência entre modelo e imagem. O alinhamento com maior número de correspondências entre objeto e imagem é computado como sendo correto.

Esta técnica apresenta dificuldades quando aplicada a um sistema de grande base de dados, uma vez que é realizada uma enumeração exaustiva sobre todos os possíveis pares de conjuntos mínimos de atributos dos modelos e da imagem.

b) Sistema proposto por Horaud

Horaud propôs um sistema de reconhecimento de objetos 3D, usando uma única imagem representando intensidade luminosa, onde a correspondência é efetuada no espaço tridimensional, e não no bidimensional dos atributos. Com isso, os atributos 2D da imagem são projetados de volta para o espaço 3D.

Os modelos usados para os objetos são compostos por:

- CAD estendido (volume-superfície-aresta-vértice), onde é construída uma rede de ponteiros ligando as primitivas conectadas topologicamente;
- várias listas, constituídas da listagem de atributos sob diversos tipos de classificação (por exemplo, mais saliente, primeiro).

Primeiramente, o sistema efetua a aquisição da imagem, constituída basicamente por objetos compostos por faces planas e arestas retas. Atributos lineares são extraídos (junções compostas por dois ou mais segmentos de reta) e projetados de volta para o espaço 3D. Para isso, foi feita a consideração de que linhas que se cruzam na imagem também devem se cruzar no espaço, o que, no entanto, nem sempre corresponde à verdade.

No espaço 3D é efetuada a busca por correspondências, utilizando o controle de busca em profundidade. O paradigma usado é de hipótese-e-teste. Foi usada a restrição de rigidez para o controle da busca.

c) Sistema proposto por Thompson e Mundy

A abordagem proposta por Thompson e Mundy [Thompson, Mundy 1987] é voltada para processamento paralelo, determinando aglomerações de transformações derivadas das especificações feitas por múltiplos atributos.

Os atributos usados são bastante simples e de fácil e rápida extração. Consistem nos chamados *pares de vértices*, compostos por (figura 3.3):

- coordenadas de um vértice e das duas arestas que o formam;
- coordenadas de um segundo vértice.

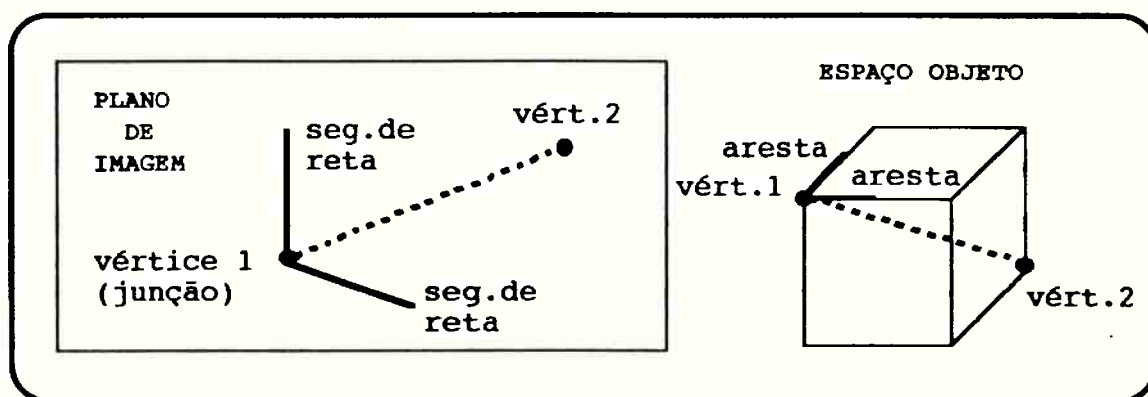


Fig.3.3 Atributo "par-de-vértices" usado no sistema de Thompson e Mundy.

O número de pares de vértices numa cena pode ser bastante grande. Na implementação do sistema, foi usado um único objeto poliédrico, com modelamento 3D com representação por borda.

Com base na restrição de rigidez dos objetos usados em cena, sabe-se que a transformação resultante de todos os pares de vértices do modelo para mapear a imagem tem que ser consistente.

Desse modo, cada correspondência possível entre um par de vértices da imagem e do modelo vota por uma transformação rígida num espaço de transformações possíveis.

O reconhecimento consiste na detecção de um ponto no espaço de possíveis transformações, cujas coordenadas mapeiem um número significativo de atributos da imagem aos dos modelos, isto é, consiste em determinar o ponto com maior peso de votação no espaço de transformações possíveis.

As grandes capacidades de computação e de armazenamento exigidas por tal abordagem são atenuadas com recursos como: decomposição do espaço de transformações possíveis, eliminação de regiões vazias (sem votos) neste espaço logo ao iniciar a busca por picos, etc. Mesmo assim, este sistema requer muita capacidade de armazenamento e é principalmente indicado para cenas com poucos objetos, sistemas com poucos modelos e objetos poliédricos.

d) Sistema proposto por Lamdan e outros

O sistema proposto por Lamdan, Schwartz e Wolfson [Lamdan et al. 1988] também é de interesse, pois, apesar da implementação ter-se realizado somente para objetos 2D (imagem binária), foram propostas algumas extensões para o caso de objetos 3D. No entanto, a nível geral, o algoritmo do sistema seria o mesmo, só mudando a dimensão do atributo usado (vale ressaltar que nem sempre essa extensão é direta e simples de ser realizada).

O sistema é dividido em duas fases, pré-processamento e execução, além de também ser voltado para processamento

paralelo, como o de Thompson e Mundy [Thompson, Mundy 1987], nas duas fases.

Os objetos são modelados pelo seu contorno, no caso 2D (ou pela representação por borda, no 3D); além disso, possui uma lista de pontos interessantes associados. Os pontos interessantes usados foram baseados em concavidades profundas e convexidades agudas das linhas do contorno, como exemplificado na figura 3.4.

Na fase de pré-processamento, é definida uma base de coordenadas para cada trinca de pontos interessantes não colineares. A coordenada de todos os outros pontos interessantes é calculada, em relação à base estabelecida.

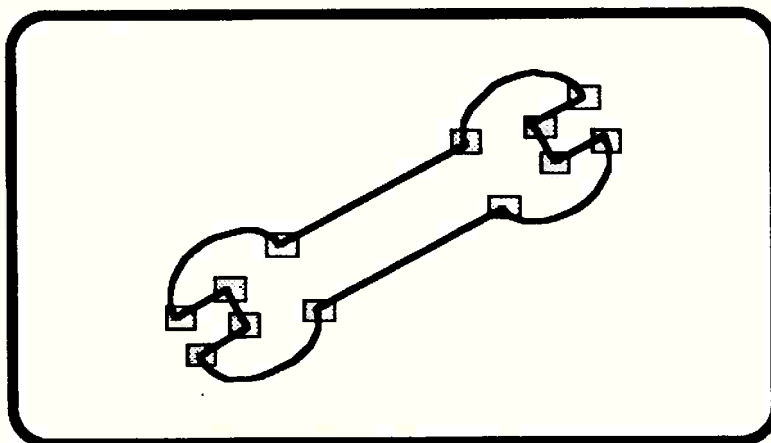


Fig.3.4 Pontos interessantes usados por Lamdan e outros nos modelos dos objetos.

Para cada modelo é então construída uma tabela, chamada "hash-table", onde tem-se associada a base correspondente a cada coordenada.

Na fase de execução, dada uma imagem, são determinados seus pontos interessantes. Escolhe-se então uma trinca de pontos não colineares como base e calcula-se as coordenadas dos outros

pontos em relação a essa base. Cada uma dessas coordenadas acessa a "hash-table" e vota por um par (modelo,base). Começando do par mais votado (acima de um limiar), os outros pontos interessantes do modelo são projetados na imagem. Se os pontos corroborarem com a imagem, faz-se então uma verificação mais apurada; caso contrário, repete-se a operação para outro par mais votado, se existir, ou retorna-se ao passo de escolher uma nova base na imagem.

A verificação mais apurada consiste em transformar o contorno do objeto para o espaço imagem, dada a correspondência inicial, feita pelos pontos interessantes. O contorno projetado é comparado com o da imagem, aceitando o resultado ou refutando-o e voltando ao passo anterior do algoritmo.

e) Sistema OPTICA

O sistema OPTICA (Object recognition using Probabilistic Three-dimensional Interpretation of Component Aspects), de Dickinson, Pentland e Rosenfeld [Dickinson et al. 1992], é dividido em duas fases: pré-processamento e reconhecimento.

Os objetos que fazem parte do banco de dados são modelados a partir de primitivas volumétricas constituintes (geons).

As primitivas volumétricas usadas no modelamento dos objetos formam um conjunto que é representado, a partir da projeção de suas superfícies, numa hierarquia de aspectos. Associada à hierarquia foi criada uma representação matricial que descreve as probabilidades condicionais associadas aos mapeamentos de agrupamentos de contornos a faces, faces a aspectos e aspectos a primitivas volumétricas.

Na fase de reconhecimento, a imagem de entrada é segmentada e os atributos são agrupados, formando os aspectos.

Uma vez recuperados os aspectos, a hierarquia é usada para inferir, através de isomorfismos de subgrafos, o conjunto de primitivas volumétricas que possivelmente se encontra na imagem, bem como a conectividade entre elas. Esse agrupamento de primitivas volumétricas é então usado para indexar a base de objetos modelos.

O uso de tal vocabulário finito de primitivas volumétricas, a partir das quais os objetos podem ser construídos, faz com que a maior dificuldade do sistema seja deslocada para a detecção e agrupamento de primitivas.

f) Sistema SCERPO

No sistema SCERPO, proposto por Lowe [Lowe 1985; 1987], faz-se o reconhecimento de objetos (basicamente poliedros) a partir de uma imagem de intensidade luminosa.

Num pré-processamento, os objetos 3D são modelados usando representação por estruturas aramadas, porém com informação de superfícies para que possa ser feita a visualização de somente arestas visíveis, na fase de verificação. Além disso, ainda é feito, nesta fase, o equacionamento da combinação da informação probabilística de múltiplas fontes, para cada modelo.

Uma imagem de intensidade luminosa é captada e métodos tradicionais de extração de contornos e aproximação poligonal são utilizados para determinar segmentos de reta na imagem.

Um processo de Organização Perceptual é usado, para formar agrupamentos e estruturas na imagem, os quais são escolhidos

dentre aqueles que normalmente são invariantes numa grande faixa de observação.

Os agrupamentos são classificados pela sua significância perceptual (mais importantes, primeiro) e probabilidade de ocorrência acidental (menos provável, primeiro).

Se existirem vários modelos, também é usado um método de ordenação probabilística dos modelos para reduzir o tamanho do espaço de busca durante o processo de correspondência entre modelos e imagem. Este método indica a probabilidade de ocorrência de cada modelo, de acordo com a informação extraída da imagem (agrupamentos diversos). Estas probabilidades contribuem num esquema de votos, onde os votos são acumulados, dados pelos agrupamentos detectados na imagem, e usados para ordenar os modelos, com base no total de votos recebidos.

Finalmente, um processo de correspondência espacial compara a projeção do modelo 3D (mais provável) com a imagem, minimizando as diferenças com a utilização do método de Newton (ver figura 2.16). Com isso, ou o modelo é recusado e nova correspondência espacial com outro modelo é realizada, ou o modelo é aceito e os parâmetros de posicionamento do objeto na imagem são aperfeiçoados.

Uma dificuldade apresentada pelo sistema SCERPO é que, em tarefas que interpretam cenas com vários objetos e que contenham um grande número de modelos, o mecanismo de votos utilizado não seria suficientemente discriminatório. Além disso, outro problema apresentado é o custo das transformações entre modelos 3D e atributos 2D no desenvolver do processo de correspondência.

g) Sistema proposto por Kuno e outros

No sistema de Kuno, Okamoto e Okada [Kuno et al. 1991] o modelo utilizado é composto por uma hierarquia que representa o objeto (somente um objeto é modelado), onde o primeiro nível representa a descrição 3D completa do objeto, o segundo nível, suas peças constituintes, o terceiro nível, as primitivas 3D que compõem as peças, o quarto nível, atributos intermediários, formados por agrupamentos dos atributos 2D, representados no quinto e último nível da hierarquia (figura 3.5).

Essa hierarquia é construída em três passos, considerando os cinco critérios abaixo para ordenação da busca pelos atributos:

- início do processo de reconhecimento procurando primeiramente por aqueles atributos que possam ser vistos pelo maior número de pontos de observação;
- probabilidade de detecção acidental do atributo;
- custo computacional de detecção de cada atributo;
- proporção visível da peça;
- proporção de distorção.

Para a compilação da estratégia de reconhecimento de cada objeto, são fornecidos:

- lista das descrições das aparências do objeto, para um determinado número de pontos de observação;
- lista dos atributos 3D (quarto nível da hierarquia);
- descrição do objeto em termos de peças e primitivas volumétricas constituintes.

O compilador então monta um grafo de estratégias, organizando a informação de acordo com os cinco critérios citados.

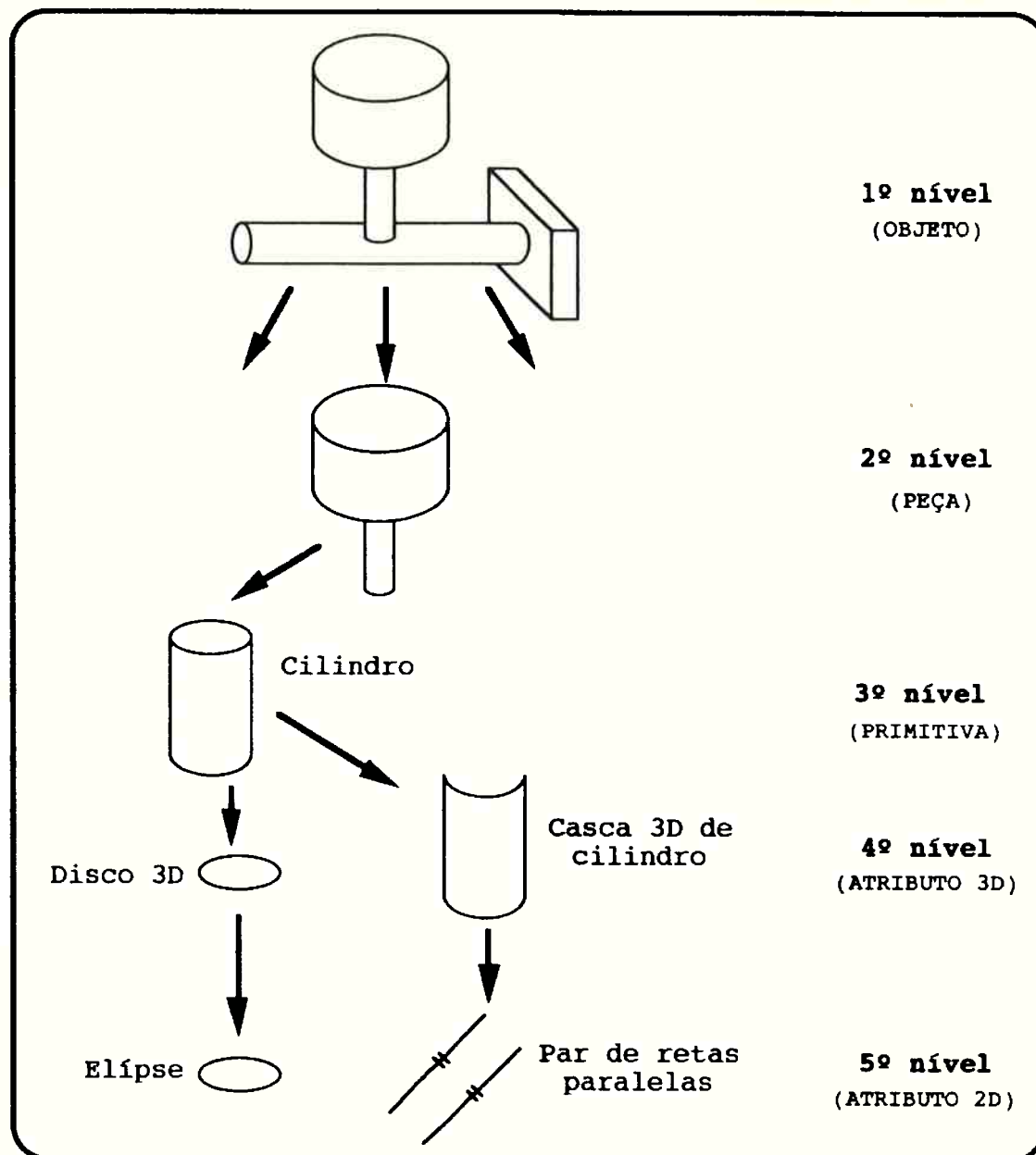


Fig.3.5 Exemplo da representação de um objeto no sistema de Kuno e outros (adaptada de [Kuno et al. 1991]).

No reconhecimento, a busca é efetuada para cada objeto, com controle em profundidade, guiada pelo respectivo grafo de estratégias. A busca termina com a definição de um conjunto de

atributos da imagem em correspondência com a descrição da aparência de uma das vistas amostradas. Com isto, tem-se uma vaga estimativa da localização do objeto na cena.

O sistema então aplica essa estimativa de transformação no modelo 3D do objeto e faz, iterativamente, uma aproximação dos atributos projetados do modelo transformado aos atributos da cena. Para isso, não é utilizado um método simultâneo de localização, como Lowe usou, mas sim, é modificado um parâmetro por vez, procurando a minimização das distâncias. Esse método foi satisfatório, pois normalmente a estimativa inicial gerada para a transformação é bastante boa.

h) Sistema proposto por Burns e Kitchen

O sistema de Burns e Kitchen [Burns, Kitchen 1988] também realiza, numa fase de pré-processamento, a compilação de uma estrutura para guiar a busca por correspondência, durante o reconhecimento. Esta estrutura é chamada Hierarquia de Predições.

O domínio escolhido é para aplicações com objetos poliédricos.

As descrições de como o objeto pode apresentar-se na imagem (aparências), são nomeadas predições. Essas predições são organizadas na hierarquia de predições, baseadas em operações de combinação (PART-OF) e especialização (IS-A).

Uma predição é representada como um grafo relacional; os elementos desse grafo são projeções de segmentos de reta. As relações associadas aos arcos do grafo restringem mutuamente as posições, orientações e tamanhos relativos dos pares de

segmentos de reta. Cada predição pode ser satisfeita por mais de um objeto da base de modelos, e mesmo por mais de uma vista do próprio objeto que a gerou.

Na hierarquia, as predições são os nós e as especializações ou combinações de nós geram sucessores. No sistema, somente o processo de combinação foi implementado.

O processo de compilação da estrutura inicia com predições bem simples, válidas para grande número de aparências, de vários objetos. Uma busca iterativa é feita, procurando combinações e, eventualmente, isolando predições características de alguma vista específica de um objeto.

Os critérios para a construção da hierarquia, são:

- para maior eficiência na busca por correspondência, a hierarquia deve se aproximar ao máximo de uma estrutura binária; para isso, cada nó sucessor deve procurar satisfazer a metade dos predecessores;
- a estrutura não deve ser discriminatória (para poder tratar oclusões e falhas na aquisição da imagem), sendo que, para cada subconjunto de possibilidades, deve estar associado um sucessor que requisite evidência positiva adicional na imagem;
- o número de sucessores por predição não deve ser muito grande;
- cada predição (não folha da hierarquia) deve ter pelo menos um sucessor.

A busca nesta hierarquia gerada, durante o reconhecimento propriamente dito, não havia sido, até então, implementada. Porém, a proposta consiste em implementação distribuída espacialmente na imagem, para duas predições iniciais

simultâneas, procurando intersecções de sucessores na hierarquia. Um processo de verificação e refinamento da localização deve ser efetuado após adquirir hipóteses, geradas com a busca na hierarquia.

i) Sistema proposto por Swain

Swain [Swain 1988] descreve um sistema capaz de efetuar o reconhecimento para um número significativo de objetos poliédricos no modelo, usando como testes a topologia, comparações de comprimentos e ângulos, e consistência da transformação rígida aplicável.

Dadas as probabilidades *a priori* de determinado poliedro aparecer na cena, dos pontos de observação a partir dos quais o poliedro é visto e dos erros que podem ocorrer na determinação de seus atributos na imagem, uma árvore de decisão é construída, usando o critério de minimizar o tempo de reconhecimento esperado.

A construção da árvore se dá baseada na minimização da entropia esperada da distribuição *a posteriori* das possíveis correspondências. No cálculo da entropia está embutido o conceito de saliência dos atributos, fazendo com que a árvore chegue o mais rápido possível em discriminações entre os modelos. Este fato pode ser problemático quando tratar oclusões, pois os atributos salientes podem ficar ocultos. Além disso, na maioria das aplicações, é muito difícil ter conhecimento de todas as probabilidades *a priori* requeridas para a construção da árvore de decisão. Por isso, foi proposta a utilização de exploração múltipla concorrente da árvore de decisão; isto é,

sempre que dúvidas são geradas durante a exploração da árvore, processos investigam simultaneamente os ramos descendentes, evitando, com isso, uma dependência muito grande dos valores das probabilidades.

3.1.4. Resumo e Comentários

Nos trabalhos apresentados ressaltou-se que, apesar dos algoritmos diferirem na escolha dos atributos, nos modelos, nos métodos de indexação dos modelos e busca usados, o paradigma dominante é aquele constituído pelos seguintes passos:

- extração de atributos;
- geração de hipóteses de correspondência entre atributos da imagem e modelos;
- verificação das hipóteses e
- previsão das conseqüências.

A escolha dos atributos dependem dos seguintes fatores:

- tarefa a ser executada;
- ambiente onde a tarefa será realizada;
- sensores usados e
- tipos de objetos tratados.

Os trabalhos descritos apresentaram métodos de aquisição da imagem como:

- câmeras [Roberts 1965; Guzman 1968; Brooks 1983; Huttenlocher, Ullman 1988; Horaud 1987; Lowe 1985; etc];
- sensores tácteis [Grimson, Lozano-Perez 1985];
- estéreo fotométrico dual [Ikeuchi, Kanade 1988a; 1988b];
- sensores a laser [Faugeras, Hebert 1986]

- "scanner", onde um plano de luz é projetado na cena e, por triangularização, as distâncias são calculadas [Bolles, Horaud 1986];
- luz estruturada [Oshima, Shirai 1981].

Os sistemas de reconhecimento apresentados se voltaram a tarefas como:

- retirar um objeto geometricamente simples ou poliédrico do topo de um amontoado formado por objetos idênticos empilhados [Bolles, Horaud 1986; Ikeuchi, Kanade 1988a; 1988b];
- reconhecimento de aviões em imagens aéreas de aeroportos [Brooks 1983];
- reconhecimento de peças poliédricas simples [Lowe 1985; 1987; Roberts 1965; Waltz 1975];
- reconhecimento de peças automobilísticas complexas [Faugeras, Hebert 1986];
- reconhecimento de objetos rígidos ou articulados compostos por primitivas volumétricas simples [Bergevin, Levine 1989; 1992a; 1992b; Dickinson et al. 1992].

Os atributos usados variaram desde segmentos de retas e junções dos mesmos [Faugeras, Hebert 1986; Guzman 1968; Huttenlocher, Ullman 1988; Kanade 1981; Roberts 1965; Waltz 1975; etc] até primitivas volumétricas [Bergevin, Levine 1989; 1992a; 1992b; Dickinson et al. 1992], passando por equações de planos [Faugeras, Hebert 1986; Grimson, Lozano-Perez 1985; Oshima, Shirai 1981], EGI [Ikeuchi, Kanade 1988a; 1988b], curvas de intersecção de superfícies [Bolles, Horaud 1986], agrupamento de curvas [Clemens, Jacobs 1991; Jacobs 1988; Lowe 1985; 1987], entre outros.

Os trabalhos também descreveram diversas abordagens para a realização da indexação da base de dados e busca por correspondências, onde diferentes recursos foram utilizados para restringir a busca. Assim, características especiais dos objetos foram usadas para diminuir o espaço de busca, incluindo:

- esquema de votos num espaço de transformações;
- terminação heurística da busca;
- focalização em atributos interessantes ou salientes.

No esquema de votos em um espaço de transformações, também chamado de Transformada de Hough Generalizada, todos os pares de atributos de modelo e imagem são usados para computar conjunto de transformações consistentes com esses pares. Cada transformação recebe um voto num espaço quantizado de transformações possíveis. Picos de votos neste espaço indicam subespaços promissores para que se desenvolva a busca por correspondência (solução) [Grimson 1990; Thompson, Mundy 1987]. Esquemas de "Hashing" Geométrico ("hash-table") também se enquadram neste contexto.

Por outro lado, pode-se também atribuir "notas" (custos) às interpretações, a fim de reduzir a busca. Sempre que uma interpretação hipotética é encontrada no processo de busca, calcula-se uma medida (nota) e, se esta medida ultrapassar um limite aceitável, toma-a como solução e a busca é terminada [Grimson, Lozano-Perez 1985; Lowe 1985; etc].

Finalmente, pode-se também focalizar a busca em atributos interessantes (ou salientes), considerando conjuntos mínimos de pares de atributos em correspondência, entre modelo e imagem, para gerar hipóteses de identidade e localização de objetos na cena. Essa técnica é usada em [Bolles, Horaud 1986] e,

implicitamente, em [Chen, Kak 1989; Huttenlocher, Ullman 1988] etc.

Todas essas abordagens são inerentemente dirigidas pelos modelos, uma vez que usam particularidades dos modelos para reduzir o processo de busca. No entanto, quando a base de dados possui um grande número de modelos, geralmente não se tem possibilidade de selecionar particularidades dos objetos para reduzir a busca.

Usa-se, então, recursos mais genéricos, sendo mais baseados nos dados extraídos da imagem para indexar os modelos. Esse é o princípio do agrupamento de atributos, baseado na literatura de Percepção, fornecida principalmente pelos psicólogos gestaltistas da década de vinte e trinta. Esse princípio inspirou os trabalhos de [Jacobs 1988; Lowe 1985; McAndrew, Wallace 1989; etc], assim como o presente trabalho.

3.2. Proposta de um Sistema de Reconhecimento 3D a Partir de uma Única Imagem 2D

A partir do estudo e análise dos fatores referentes ao reconhecimento visual e das abordagens feitas por outros pesquisadores, sentiu-se a necessidade de buscar uma nova solução, visando solucionar alguns dos problemas levantados.

3.2.1. Motivação

Os pontos críticos a serem tratados para uma solução adequada ao reconhecimento são os seguintes:

- definição, extração e descrição de atributos das imagens e dos objetos;
- correspondência entre atributos das imagens e dos objetos.

Uma abordagem a esses problemas foi pensada, visando aplicações de reconhecimento 3D a partir de uma única imagem 2D, usando o contorno dos objetos para este fim, por entender que esta única dimensão de resolução do reconhecimento é muito importante, podendo ser aplicada como um pré-requisito para determinar correspondências em outras dimensões, como estereoscopia, determinação de movimento, etc.

O interesse nesta abordagem também é reforçado quando se busca analogias com o sistema visual humano, onde se nota alta eficiência no reconhecimento de objetos a partir do desenho de seus contornos.

Porém, com o interesse voltado a esta única dimensão para uma abordagem ao reconhecimento visual, sentiu-se a necessidade da criação de estratégias de reconhecimento mais eficientes do que as apresentadas até então. Estas estratégias deveriam utilizar *todo ou grande parte* do conhecimento disponível (características dos objetos e dos sensores usados, tarefa a ser executada, níveis desejados de soluções, etc), de modo *integrado*, visando maior eficiência ao reconhecimento.

Para isso, deveria também ser definido algum tipo de representação dessas estratégias que fosse de fácil acesso, para a condução da busca por correspondências.

Além disso, esta representação deveria permitir níveis de abstrações diferentes para que se pudesse resolver o problema de reconhecimento de acordo com o detalhamento desejado.

Seria do mesmo modo interessante que a organização das estratégias de reconhecimento *interagisse* com o conhecimento adquirido durante a repetição da tarefa, permitindo um certo grau de *aprendizagem* [Draper, Riseman 1990].

As estratégias de reconhecimento deveriam, no entanto, ser indexadas por informações extraídas das imagens, isto é, essas informações extraídas de modo relativamente simples, deveriam restringir o acesso a determinadas estratégias e selecionar outras, possibilitando uma redução do conjunto de estratégias a serem acessadas para a solução do problema.

Estes interesses foram aglutinados numa solução, resultando na proposta aqui apresentada: sistema RECTRI.

3.2.2. Proposta do Sistema RECTRI

RECTRI surgiu da necessidade de fornecer uma abordagem que satisfaça a gama de interesses acima apresentada.

RECTRI segue o paradigma dominante de: extração de atributos; geração de hipóteses de identificação e localização; verificação das hipóteses e previsão das conseqüências.

No entanto, utiliza-se extensivamente da fase de pré-processamento para buscar estratégias de reconhecimento. Estas estratégias estão representadas numa solução integrada, chamada Hierarquia de Estratégias, onde todos os objetos se encontram representados, com seus atributos ressaltados, evidenciando suas semelhanças e diferenças para conduzir eficientemente a busca por correspondências entre atributos da imagem e dos objetos. Essa Hierarquia de Estratégias constitui uma abordagem inovadora e promissora, sendo totalmente construída de modo automático.

Caracterizações qualitativas da morfologia das peças são refletidas nos atributos representados na Hierarquia e naqueles extraídos das imagens. Esses atributos são agrupamentos de primitivas, escolhidos por apresentarem boa invariância com diferentes pontos de observação, além de outras características, como frequência com a qual é encontrado no objeto, robustez de sua detecção, complexidade de sua computação e probabilidade de corretamente identificar e localizar cada objeto. Esses atributos atuam como indexadores da base de dados, selecionando estratégias eficientes para o reconhecimento.

4. SISTEMA RECTRI

4.1. Introdução

O sistema RECTRI é composto por duas fases: pré-processamento e reconhecimento. Essa divisão é uma importante característica, uma vez que possibilita significativa redução no tempo de execução do algoritmo de reconhecimento por efetuar, previamente, ordenação dos conhecimentos. A figura 4.1 apresenta um esquema geral do sistema proposto, onde cada uma das etapas será descrita no decorrer do capítulo.

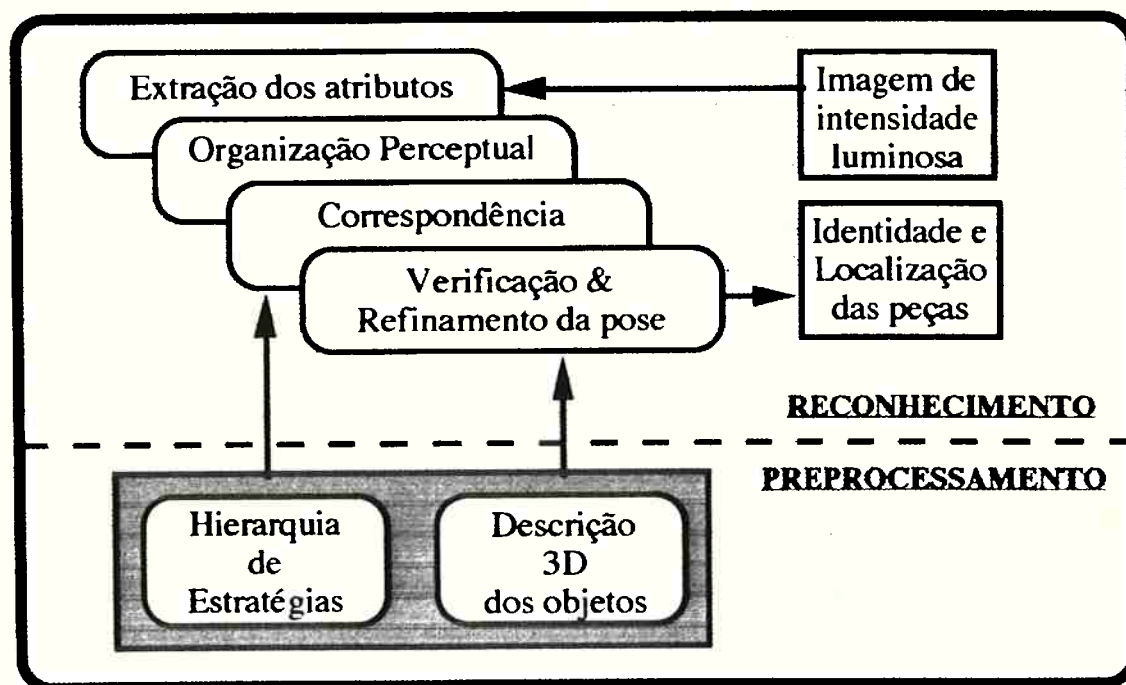


Fig.4.1 Esquema geral do sistema proposto.

4.2. Fase de Pré-processamento

RECTRI, na sua base de dados, utiliza um *modelo híbrido*, composto por três entidades básicas (figura 4.2):

- *descrição tridimensional* de cada objeto (modelo do tipo volume-superfície-aresta-vértice);
- *representação multivista modificada* de cada descrição;
- *Hierarquia de Estratégias*, onde todos os objetos se encontram representados e estratégias de busca por correspondência entre atributos são definidas.

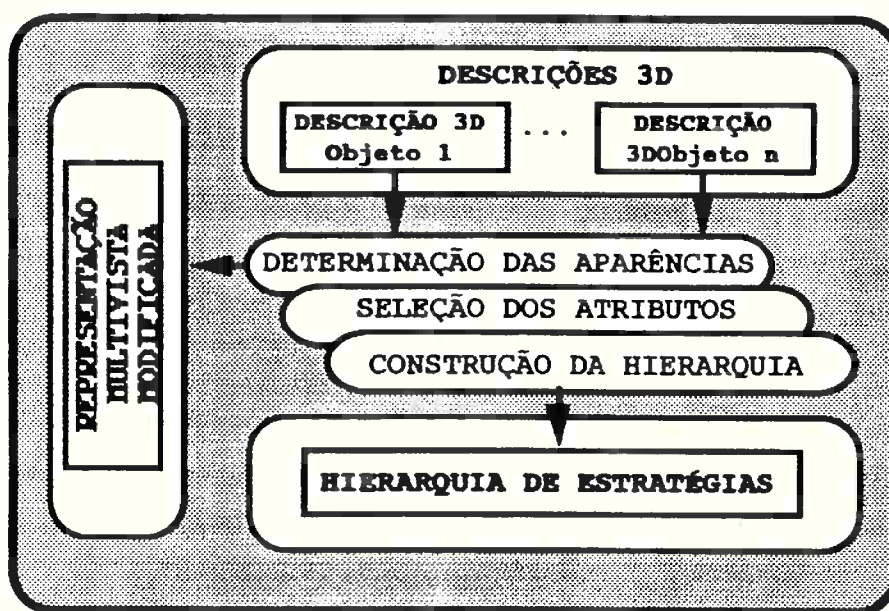


Fig.4.2 Fase de pré-processamento.

A fase de pré-processamento tem a função de construir esse modelo híbrido *automaticamente*, a partir da descrição de cada objeto, editada pelo usuário. Esta descrição dos objetos poderá

ser fornecida por um sistema CAD comercial [Hansen, Henderson 1989; Flynn 1990], porém esta facilidade ainda não se encontra disponível na atual versão do sistema RECTRI.

4.2.1. Esquema de Representação para a Informação Tridimensional

Um grande problema em sistemas de reconhecimento é a escolha de uma boa representação para os modelos dos objetos. Para os propósitos de reconhecimento, é fundamental representar as relações entre atributos e estruturas que podem ser detectadas na imagem, além dos componentes correspondentes do modelo. A representação dessas relações pode ser mais importante do que mesmo a descrição da aparência visual do objeto [Lowe 1987; Dickinson et al. 1992].

Portanto, um requisito básico de um modelo é representar associações entre atributos que podem ser detectados na imagem a partir de um procedimento *guiado simplesmente pelos dados*, e os *correspondentes* atributos de modelos particulares, para serem utilizados durante os *estágios iniciais* do reconhecimento. Este requisito é de extrema importância, pois uma *indexação eficiente* da base de dados dos modelos pode agilizar em muito o reconhecimento.

Os modelos podem ser especificados em termos das propriedades geométricas dos objetos e/ou de outras propriedades disponíveis. A escolha do modelo é altamente dependente da tarefa a que se destina o sistema, não existindo, até então, um modelo que apresente generalidade, abrangência e flexibilidade suficientes a ponto de se adaptarem à diversidade de tarefas

tanto quanto, por exemplo, o sistema visual humano. Conforme visto no capítulo anterior, vários sistemas de visão propostos divergem na definição do modelo [Huttenlocher, Ullman 1988; Ikeuchi, Kanade 1988; Kuno et al. 1991; Lamdan et al. 1988; Lowe 1985; Thompson, Mundy 1987; etc].

A *representação tridimensional* dos objetos da base de dados utilizada neste trabalho consiste na representação *volume-face-aresta-vértice*, resultando em descrições usadas durante a *fase de verificação* do sistema e durante a *fase de construção do conjunto de vistas parciais* dos objetos. Esta descrição consiste em definir cada sólido por suas faces constituintes; cada face, por suas arestas constituintes e, cada aresta, por seus vértices e respectivas coordenadas (figura 4.3).

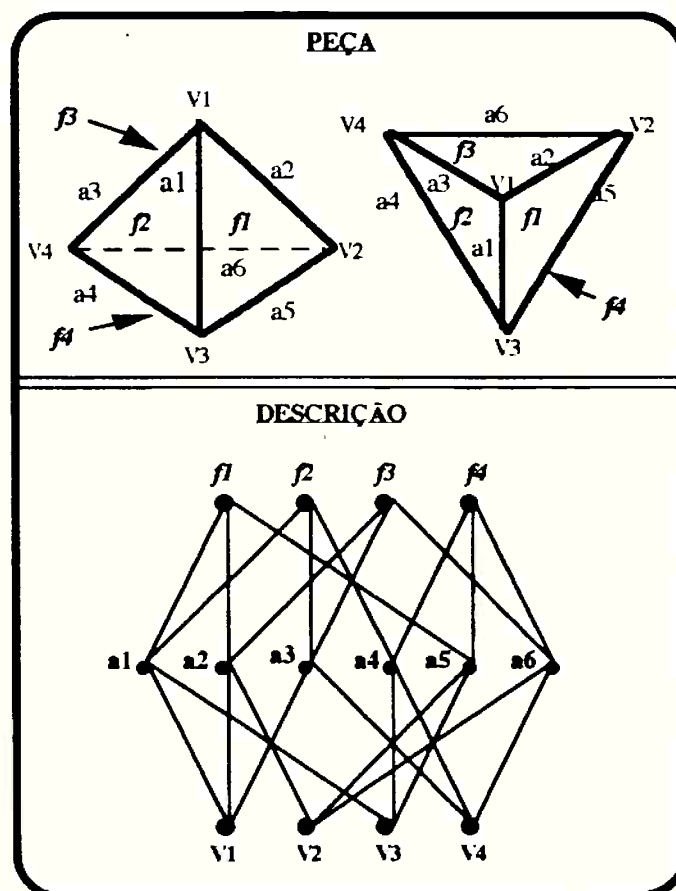


Fig. 4.3 Representação volume-face-aresta-vértice.

4.2.2. Modelos do Processo de Imageamento

São usados basicamente dois modelos do processo de imageamento no sistema aqui apresentado. Na fase de pré-processamento do modelo é usada a projeção ortogonal (ou paralela). Porém, durante o processo de verificação e refinamento do posicionamento, deve-se utilizar a projeção perspectiva como modelo do processo de imageamento, por este ser um modelo mais preciso.

a) Projeção ortogonal

Sob projeção ortogonal, cada ponto P na cena é conectado a sua imagem P' no plano I da imagem por uma reta perpendicular a I , como mostrado na figura 4.4.

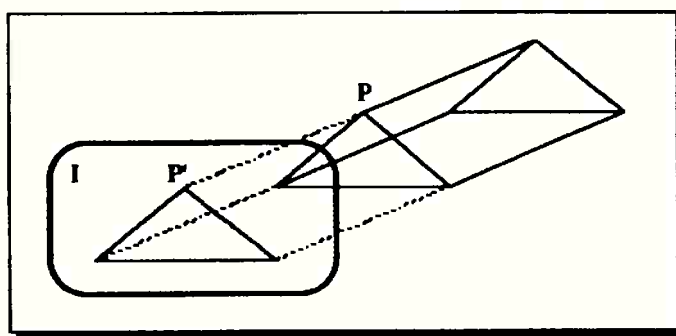


Fig.4.4 Projeção Ortogonal.

Na fase de reconhecimento, o sistema controla o processo de correspondência de acordo com a Hierarquia de Estratégias. Os tamanhos dos atributos detectados na imagem podem ser usados para modificar os valores da Hierarquia de Estratégias, visando tornar de mesmo tamanho os atributos extraídos da imagem e seus

correspondentes homotéticos armazenados na hierarquia, resultando, dessa forma, no modelo de perspectiva normal (ou „weak“ perspectiva) para o processo de imageamento.

Esta aproximação é válida quando se assume que a distância entre câmera e cena seja grande o suficiente comparada ao tamanho dos objetos, tanto na construção da Hierarquia de Estratégias quanto no processo de correspondência.

No entanto, durante o processo de verificação e refinamento do posicionamento, deve-se utilizar um modelo mais preciso do processo de imageamento: a projeção perspectiva, para que resultados mais precisos sejam encontrados [Wolfe et al. 1991].

b) Projeção perspectiva

Sob projeção perspectiva, o centro da projeção se encontra no ponto focal F . O plano I da imagem é perpendicular ao eixo de observação, o qual contém F a uma distância f (distância focal). Cada ponto P na cena é conectado a sua imagem P' em I por uma reta passando por F , como mostrado na figura 4.5.

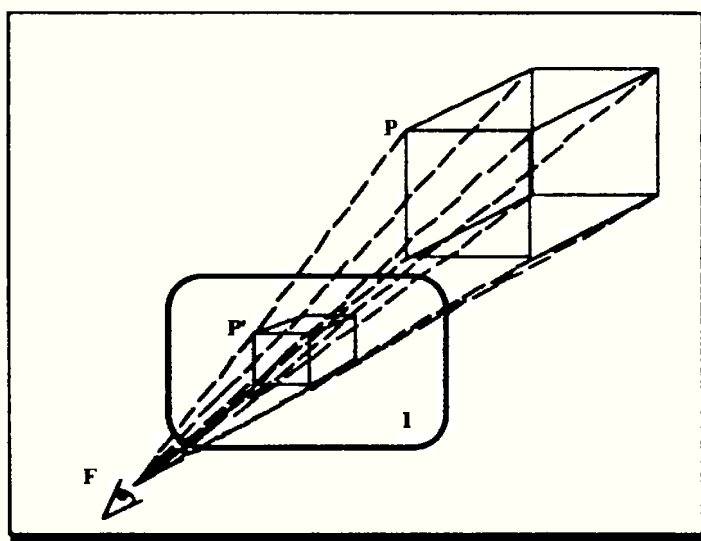


Fig.4.5 Projeção Perspectiva.

4.2.3. Uso de Vistas Parciais

A partir das descrições tridimensionais geradas, uma estrutura discriminatória é automaticamente construída, formando um *grafo acíclico direcionado*, chamado *Hierarquia de Estratégias*, porém apresentando uma singularidade no sistema de herança, ilustrada na figura 4.6, onde um sucessor poder ser apontado por mais de um antecessor ("tangled hierarquies").

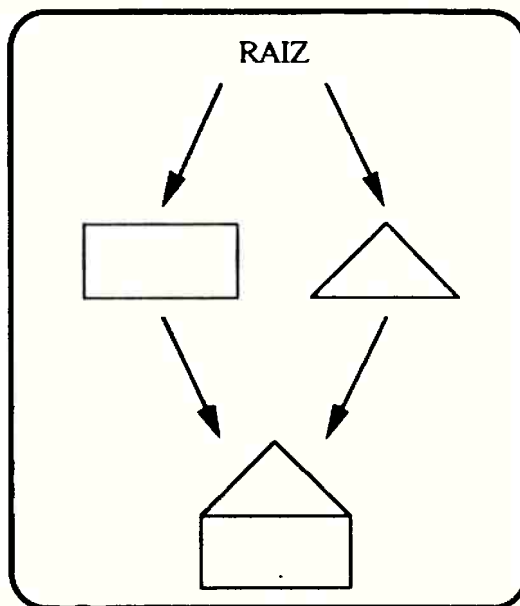


Fig.4.6. Sistema de herança da Hierarquia de Estratégias, onde dois antecessores apontam um mesmo sucessor.

A Hierarquia de Estratégias tem a finalidade de reduzir para bidimensional o processo tridimensional de correspondência, através da representação de associações entre atributos detectados na imagem e a base de dados dos modelos, usando a geometria dos objetos para acelerar o processo de correspondência. A redução de 3D para 2D é adquirida através da

utilização das vistas parciais dos objetos para a determinação dos atributos usados no processo de correspondência.

O uso de vistas parciais no reconhecimento tem a vantagem de transformar as informações tridimensionais em projeções no plano bidimensional, facilitando o processo de correspondência, uma vez que são bidimensionais as informações extraídas das imagens. A desvantagem está na grande capacidade de armazenamento de dados requerida.

No entanto, estudos e experimentos no campo da psicofísica realizados por Edelman [Edelman 1993] determinaram ser consistente o uso de vistas parciais com uma teoria de reconhecimento que envolve duas fases distintas: normalização e comparação.

Na fase de normalização, a imagem e o modelo são postos numa atitude comum (mesmo posicionamento). Esta operação poderia ser feita por um processo análogo à rotação mental, a qual teria tempo proporcional à diferença de atitude entre imagem e modelo.

A seguir, uma comparação seria feita entre ambos. O tempo para realizar a comparação dependeria, por exemplo, da complexidade do objeto, mas não de sua atitude, de tal forma que a fase de comparação contribuiria com uma quantidade constante no tempo total de reconhecimento. Por outro lado, a taxa de erro do reconhecimento seria fortemente determinada pela fase de comparação. Com a prática, um maior número de vistas parciais de estímulo poderiam ser armazenadas pelo sistema visual, resultando num menor tempo na fase de normalização. Com isso, uma possível explicação para o efeito da familiaridade, que faz com que o tempo de reconhecimento seja diminuído com a prática, seria em termos de se tornarem desnecessárias as rotações

mentais das representações dos objetos, quando vistas específicas de objetos são armazenadas.

Com base nesta teoria, propõe-se a utilização de vistas parciais dos objetos para redução do tempo de reconhecimento. No sistema aqui proposto armazenam-se todas as vistas parciais necessárias para o reconhecimento, como se já houvesse uma grande familiaridade com todos os objetos que podem compor a cena, procurando o menor tempo de reconhecimento. Essa informação é utilizada na construção da Hierarquia de Estratégias.

O cálculo das vistas parciais é realizado tendo a projeção ortogonal como modelo para o processo de imageamento.

4.2.4. Construção da Hierarquia de Estratégias

A determinação da Hierarquia de Estratégias é realizada em três passos:

- definição do conjunto de aparências,
- detecção dos atributos e
- construção da hierarquia.

a) Definição do conjunto de aparências

O procedimento adotado para a definição das vistas parciais necessárias para o reconhecimento consiste em colocar cada objeto (modelo tridimensional) com seu centro de massa coincidindo com o centro de uma *esfera gaussiana de visualização*, a qual envolve o objeto.

Cada ponto na superfície da esfera gaussiana corresponde a um *ponto de observação* do objeto. A projeção do objeto no plano tangente a um ponto de observação da superfície da esfera de visualização corresponde a uma *vista* do objeto [Goodrich 1992]. Cada vista corresponde a uma relação ponto de observação-objeto.

Aglomeraciones de vistas isomórficas vizinhas resultam em *aparências* (ou *aspectos*), onde se estabelece uma relação região de observação-objeto.

As aparências representam as projeções qualitativamente representativas, em relação à forma dos objetos visualizados e podem, dependendo da geometria do objeto, ser calculadas analiticamente ou por amostragem.

Alguns trabalhos significativos na abordagem analítica para a determinação das aparências são o proposto por Gigus e Malik [Gigus, Malik 1990], voltado para a partição exata da esfera de visualização, para objetos poliédricos, e o trabalho de Eggert e Bowyer [Eggert, Bowyer 1993], voltado para a partição exata da esfera de visualização para objetos de superfícies curvas (sólidos de revolução, isto é, aqueles que se adaptam à representação por cilindros generalizados).

Na definição das aparências por amostragem, um método muito difundido é o chamado método por simulação [Ikeuchi, Kanade 1988; Kuno et al. 1991; Dickinson et al. 1992; Zhang et al. 1992a; 1992b; etc], utilizado neste trabalho e descrito no próximo capítulo.

Todas as aparências, de cada objeto, são utilizadas em conjunto para a construção da Hierarquia de Estratégias.

b) Detecção dos atributos das aparências

Procedimentos de inferência de vários tipos são aplicados às aparências de todos os objetos, com a finalidade de extrair atributos úteis para o processo de reconhecimento.

Assim, a informação geométrica das aparências de cada objeto é analisada, determinando uma seleção dos atributos que devem possibilitar o acesso à Hierarquia de Estratégias. Esta seleção tem a finalidade de determinar a melhor estratégia para o reconhecimento, considerando as seguintes propriedades:

- restrições impostas pelos sensores (processo de formação da imagem),
- tarefa a ser realizada e
- tipo de objetos que compõem a base de dados do sistema.

Os atributos devem ser propriedades na imagem que podem ser formadas na ausência do domínio do conhecimento, porém precisam ser específicas o suficiente para servirem como termos de indexação da base de dados. O principal critério para determinar a utilidade de um atributo é baseado nas seguintes qualidades:

- frequência com a qual é encontrado no objeto,
- robustez de sua detecção,
- complexidade de sua computação,
- característica discriminatória em relação aos outros atributos e
- probabilidade de corretamente identificar o objeto e sua localização.

Procedimentos idênticos aos aplicados à imagem, para a detecção de atributos na fase de reconhecimento, são aqui utilizados. A descrição destes procedimentos e os atributos

resultantes de sua aplicação às imagens e aparências dos objetos é feita adiante. É importante salientar que os mesmos atributos 2D extraídos tanto da imagem, quanto das aparências são utilizados no processo de correspondência e, portanto, armazenados na Hierarquia de Estratégias.

c) Construção da Hierarquia de Estratégias

O terceiro e último passo consiste em *organizar* os conjuntos de aparências numa estrutura hierárquica. Desse modo, os atributos comuns às aparências são evidenciados e representados somente uma vez na hierarquia, fornecendo um mecanismo simples e uniforme de *cooperação entre variações e similaridades* entre os objetos e entre as aparências de um objeto.

A hierarquia é construída através de um *procedimento iterativo, incremental e totalmente automático*. O processo de construção parte das folhas da hierarquia até chegar à raiz, num sentido *inverso* ao utilizado durante o reconhecimento.

Cada folha está associada a uma aparência única de um objeto ou a aparências cujas diferenças em suas projeções não são representáveis e/ou perceptíveis ao sistema.

Cada folha pode representar uma *adjacência* ou *inclusão* de atributos extraídos e selecionados das aparências. *Iterativamente* retira-se um atributo representado no nó analisado, formando outros nós (portanto, com um menor número de atributos), até alcançar os *nós mínimos*, formados por um só atributo.

Nós que contenham os mesmos atributos são *aglomerados* em um só nó.

Considera-se como *válidos* somente os nós que possuam *conectividade* entre todos os seus atributos. Este fator é importante no controle da busca por atributos conectados (adjacentes ou inclusos) na imagem, durante a fase de reconhecimento, e está baseado na suposição de que atributos conectados são mais prováveis de pertencerem a um mesmo objeto.

Um trecho da hierarquia se encontra exemplificado na figura 4.7.

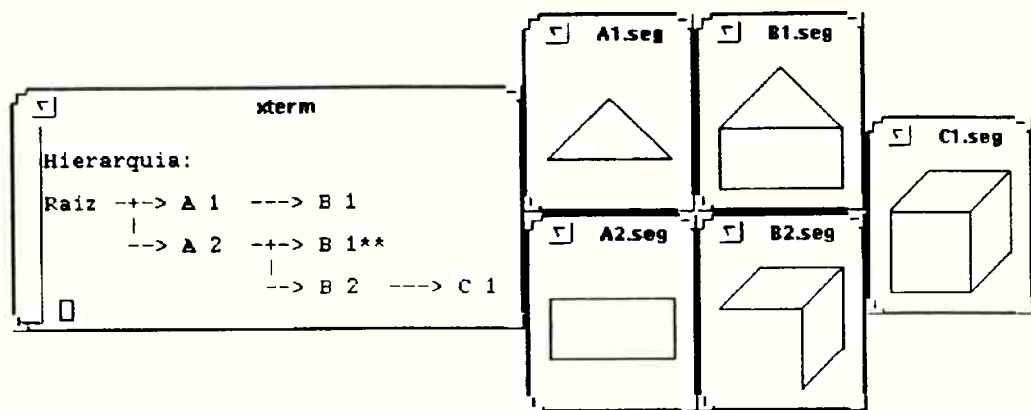


Fig.4.7 Exemplo da construção de uma pequena Hierarquia de Estratégias, composta somente pelas aparências representadas pelos nós B1, do objeto torre, e C1, do objeto cubo.

No quadro à esquerda tem-se a hierarquia resultante, que partiu das aparências B1 e C1 e chegou à raiz, durante a construção. Cada nó recebeu uma letra e um número identificando-o. Nós do primeiro nível, compostos por somente um atributo, recebem letra A; do segundo nível, letra B e assim por diante. O número identifica cada nó de um referido nível. Para efeito de clareza da representação na figura, nós que na verdade são

únicos foram desmembrados em outros no desenho, sendo este desmembramento identificado por **. Assim, o nó B1 na verdade é único e é sucessor de A1 e A2. Na representação, ele foi desmembrado em dois (um para cada antecessor), porém o símbolo ** indica sua unicidade.

Diversos critérios podem ser utilizados para organizar a hierarquia, a saber:

- mecanismo de busca utilizado;
- frequência de ocorrência de cada atributo (maior importância dada a atributos que aparecem no maior número de aparências, para garantir acesso à hierarquia);
- característica discriminatória dos atributos (atributos mais salientes, ou discriminantes, são mais importantes);
- grau de detectabilidade acidental de cada atributo (menor importância dada a atributos mais possíveis de serem detectados acidentalmente);
- robustez da detecção de cada atributo (maior importância para atributos detectados mais robustamente);
- custo operacional de detecção de cada atributo (maior importância para atributos com menor custo de detecção);
- proporção visível da peça: atributos que correspondam a menores proporções das peças são melhores (trata melhor oclusão);
- outros.

Esses critérios podem ser considerados *filtros* aplicados à informação disponível. A ordem de aplicação desses filtros influencia a organização final da hierarquia, sendo a seleção e ordenação da aplicação desses filtros dependentes da aplicação a

que se destina o reconhecimento (tipos de objetos, sensores, tarefa, etc).

A organização da hierarquia, juntamente com uma *terminação heurística* da busca por correspondência entre atributos da imagem e dos objetos da base de dados, visa uma diminuição do espaço de busca, resultando num menor tempo de reconhecimento.

É importante ressaltar as vantagens oferecidas pelo uso de uma estrutura que procure *similaridades e discriminações* entre os diversos conhecimentos (no caso, morfologia dos objetos). Esses conhecimentos, em relação ao reconhecimento, não precisam necessariamente ser somente geométricos, podendo incluir propriedades como cor, textura, funcionalidade, etc.

Essas propriedades podem ser usadas para classificar os objetos segundo similaridades representadas nos diversos níveis da hierarquia. Assim, por exemplo, objetos de cores diferentes poderiam ser agregados em classes separadas; dentre os objetos de mesmas cores, poder-se-ia classificá-los de acordo com suas texturas, para a seguir, por exemplo, organizá-los segundo sua geometria, inicialmente de modo global, porém seguindo um aumento gradativo dos detalhes observados, até chegar a correspondências únicas entre objeto e imagem. O grau de detalhes que deve ser alcançado é dependente da base de dados, do resultado esperado e das informações disponíveis; quanto maior for a característica discriminatória entre os componentes da base de dados, mais rápido e seguramente se atinge a classificação e maior poderá ser o nível de abstração (e menor número de propriedades utilizado) para se adquirir a discriminação entre todas as classes.

A ordem de distribuição das classes na estrutura e de aplicação das propriedades podem ser definidas de acordo com:

- disponibilidade de informação necessária,
- complexidade de sua determinação,
- número de classes
- quantidade de objetos dentro dessas classes,
- tarefa a ser executada,
- tipos de objetos tratados,
- método de controle de busca utilizado no acesso à estrutura,
- outros.

Deve-se ressaltar que um processo de aprendizado pode interagir nessa ordenação, visando maior eficiência, conforme adquira experiência.

No entanto, não se deve esquecer que a forma (geometria) dos objetos tem papel importantíssimo dentro do reconhecimento, pois esta se encontra intrinsecamente relacionada com a identidade da classe do objeto. Segundo Biederman [Biederman 1985], as características como cor e textura são tipicamente secundárias no reconhecimento; pode-se saber, por exemplo, que uma cadeira tem uma determinada cor e textura, simultaneamente com sua descrição volumétrica; no entanto, somente a descrição volumétrica fornece um acesso eficiente à representação cadeira. Devido a essa característica (e também às informações disponíveis na aplicação abordada) somente as características geométricas das peças são abordadas no corrente trabalho.

Deve-se notar também que, apesar do processo de construção da hierarquia ser computacionalmente custoso, ele é executado no

pré-processamento, portanto, uma única vez, para cada banco de dados, permitindo uma maior eficiência no reconhecimento.

4.3. Fase de Reconhecimento

Tendo organizado, na fase anterior, seus conhecimentos e informações disponíveis, RECTRI se encontra então apto a efetuar reconhecimentos, de acordo com sua base de dados. Se for necessária a inclusão ou modificação de algum elemento da base de dados, deve-se executar novamente a fase de pré-processamento.

A fase de reconhecimento consiste em: detecção de atributos na imagem; processo de correspondência entre atributos da imagem e dos modelos, gerando hipóteses de identificação e localização e processo de verificação, onde as hipóteses são refutadas ou aprimoradas. Essas etapas serão descritas no decorrer deste item.

4.3.1. Escolha dos Atributos

Os estágios iniciais do processamento da imagem num sistema de reconhecimento devem ser realizados sem o uso de conhecimento sobre objetos e suas poses. No entanto, em algum ponto é necessário usar atributos extraídos da imagem para indexar a base de dados dos modelos, associando atributos com os modelos dos objetos que mais provavelmente devem estar presentes na cena. Esse mecanismo de indexação requer atributos que sejam

menos ambíguos em sua interpretação e que, além disso, permaneçam estáveis sob mudanças nos pontos de observação.

A detecção de tais atributos é um dos objetivos do trabalho em Organização Perceptual [Jacobs 1988; Lowe 1985; Mohan, Nevatia 1989], que é a habilidade que permite ao sistema visual humano detectar imediatamente determinadas relações entre elementos de uma imagem, como colinearidade, paralelismo, conectividade, padrões repetitivos, etc. Com base nisso, atributos isolados são agrupados em relações estruturais que se mantêm estáveis em uma grande faixa de variação do ponto de observação, evidenciando propriedades como junções, segmentos colineares, segmentos paralelos, paralelogramos, etc.

O método usado para extrair estruturas geométricas e relacionais a partir de imagens de intensidade luminosa utiliza duas propriedades básicas: a habilidade da estrutura em constituir ponteiro eficiente na base de dados e sua robustez em relação aos processos de formação e segmentação da imagem. Essas propriedades representam um compromisso: quanto mais complexa é a estrutura, mais atrativa ela é para ser um ponto inicial de correspondência, no entanto, menor é a probabilidade de que seja extraída robustamente da imagem. Por isso, a decisão de quais agrupamentos de atributos usar depende da análise dos objetos que podem estar presentes na cena e do processo de formação e segmentação da imagem.

4.3.2. Extração dos Atributos

A extração dos atributos é organizada em duas etapas. Na primeira, é realizado um processamento na imagem para detectar

segmentos retos. Métodos tradicionais de detecção de variações bruscas na intensidade luminosa são usados, gerando cadeias de pontos* conectados [Aoyama, Kawagoe 1991; Baessmann, Besslich 1989, 1991; Ballard, Brown 1982; Canny 1986; Freeman 1974; Gonzalez, Wintz 1987; Korn 1988; Martelli 1976]. Após isso, essas cadeias são aproximadas para segmentos de retas através da seleção da representação de reta mais significativa para cada ponto, ao longo da cadeia [Lindenbaum, Bruckstein 1993; Lowe 1985]. A significância de cada reta pode ser calculada pela razão de seu comprimento pelo desvio.

4.3.3. Agrupamento dos Atributos

Na segunda etapa, é executado um processamento de nível intermediário, chamado *agrupamento*, que consiste em, a partir de uma lista de segmentos de reta produzida pelo estágio anterior, fornecer como saída um grafo relacional que expresse as relações entre esses segmentos em termos das propriedades utilizadas. O agrupamento de atributos de uma imagem exige procedimentos computacionais substanciais, porém representam um custo relativamente modesto e constante para cada atributo, além de causar considerável diminuição do processo de busca [Clemens, Jacobs 1991; Grimson 1990; Jacobs 1988; Lowe 1985].

As propriedades básicas, também chamadas *propriedades primárias*, utilizadas no sistema RECTRI para o agrupamento perceptual são: *colinearidade*, *conectividade* e *paralelismo*.

A partir destas propriedades primárias são realizados agrupamentos maiores, gerando descrições de mais alto nível de

* Ponto ou elemento da imagem. Do inglês "pixel: picture element".

abstração, como segmentos de reta interligados, formando estruturas como *triângulos, trapézios, paralelogramos, etc.*

Os agrupamentos (de modo mais amplo, também chamados de atributos) a serem determinados foram definidos considerando os objetos que compõem a base de dados e o custo computacional de suas detecções. O limite estipulado para o custo computacional permitido foi determinado empiricamente.

4.3.4. Busca por Correspondência entre Imagem e Modelo

De maneira genérica, pode-se considerar o problema da correspondência entre imagem e modelo como sendo encontrar um conjunto de atributos na imagem obtida que corresponda aos atributos do modelo de um objeto. Portanto, este processo pode ser visto como uma *busca por correspondências consistentes* entre atributos da imagem e do modelo.

O processo de correspondência aqui usado utiliza os atributos intermediários (agrupamentos) extraídos da imagem para acessar a Hierarquia de Estratégias, construída na fase de pré-processamento. A detecção de tais atributos, conforme dito, é baseada nas regras da Organização Perceptual.

A Hierarquia de Estratégias controla o processo de correspondência, gerando hipóteses de objetos e seus pontos de observação, através da predição de atributos dos modelos que ainda não foram incluídos na correspondência, mas que foram considerados importantes na fase de pré-processamento.

No processo de geração de hipóteses, o aumento da certeza de uma hipótese provoca dois efeitos: aumento do *custo da geração da hipótese* e diminuição do *custo da verificação* [Huttenlocher,

Cass 1992]. Para contrabalançar esses dois efeitos, o processo de correspondência é evoluído até que se obtenha uma estimativa inicial aproximada para o objeto e seu ponto de observação e então o processo de verificação é aplicado. Note que para isso é necessário somente uma estimativa inicial dos parâmetros para que o processo de verificação possa validar a hipótese.

Outra opção, mais elaborada, seria a de comparar as funções de custo da geração da hipótese com o custo da verificação, delimitando a busca até o ponto onde a função de custo de verificação aponte um valor menor (ou igual) ao custo da geração da hipótese, mesmo que não se tenha definido uma única aparência (ou objeto) relacionada com aqueles atributos extraídos da imagem analisada. Para este procedimento, torna-se imprescindível a exata definição das correspondentes funções de custo, o que não é uma tarefa simples.

4.3.5. O Processo de Verificação

O processo de verificação consiste em resolver a *correspondência espacial* [Lowe 1985], isto é, dado um conjunto de correspondências conhecidas entre pontos 3D de um modelo e pontos 2D de uma imagem, determinar quais seriam os valores dos parâmetros do modelo e da projeção desconhecida que irão resultar na projeção dos dados pontos do modelo nos pontos correspondentes da imagem [Heller, Stenstrom 1989].

Entretanto, na maioria dos casos as correspondências irão ocorrer entre contornos projetados de um modelo e bordas parciais de uma imagem. Por isso, a idéia pode ser expandida para permitir o uso de correspondências entre curvas, em vez de

somente entre pontos. Esse processo de verificação deve utilizar projeção perspectiva para a transformação do modelo do objeto, para que se adquira maior precisão nos resultados.

O processo de verificação deve ser utilizado quando já existe uma previsão quanto à identidade de um objeto, dentre aqueles da base de dados, e à localização do mesmo e tem como finalidade permitir uma decisão final quanto a aceitar ou recusar a previsão em questão (hipótese de identidade e localização). O processo de verificação deve utilizar todas as evidências disponíveis.

Assim, dada uma correspondência entre segmentos da imagem e do modelo, estima-se a localização do objeto na cena. Com isso, efetua-se a transformação correspondente na descrição 3D do objeto em hipótese e projeta-o na imagem, eliminando as arestas e faces não visíveis naquela posição. Feito isso, é realizada uma busca em todos os segmentos da imagem ao redor da região da projeção, procurando evidências que comprovem (ou recusem) a hipótese em questão.

Este processo não será abordado neste trabalho, deixando esta tarefa para uma etapa seguinte de desenvolvimento do sistema RECTRI.

Apesar da existência de diversas abordagens para esta etapa, sugere-se atenção especial para as propostas feitas por Lowe [Lowe 1987, 1991] e por Tomaselli [Tomaselli 1993].

A primeira pode ser aplicada a objetos rígidos ou articulados, de superfície curva, e as exigências de previsão inicial são bastante flexíveis; a abordagem efetua uma linearização da função de parâmetros, permitindo a utilização do

método de Newton, aliado ao método de Levenberg-Marquardt, para garantir a convergência para a solução.

Já a abordagem feita por Tomaselli define um modelo matemático do tipo explícito, que é tratado por filtragem Kalman, permitindo a determinação simultânea dos seis parâmetros de orientação. É um trabalho bastante interessante, apesar de exigir um desvio menor das estimativas preditas do que aquele exigido pela abordagem de Lowe.

5. IMPLEMENTAÇÃO E RESULTADOS DO SISTEMA RECTRI

5.1. Implementação

O sistema RECTRI foi inteiramente desenvolvido em estação de trabalho sob ambiente UNIX, na linguagem C, usando o compilador C da Gnu, resultando em aproximadamente 20.000 linhas de código.

O programa está dividido em três blocos: Global, Reconhecimento e Pré-processamento.

No Global encontram-se os seguintes módulos:

- rotina principal;
- variáveis e estruturas globais do sistema;
- parâmetros de ajuste;
- diversas funções de apoio, usadas também de maneira global.

No bloco Reconhecimento existem três módulos:

- extração de atributos, que possui diversas funções de integração com programas externos ao RECTRI, para interfacear a aquisição de imagens segmentadas por outros sistemas;
- agrupamento, que possui funções referentes à determinação das relações primárias entre atributos e ao agrupamento das mesmas em estruturas intermediárias;
- correspondência, que possui funções referentes ao processo de correspondência entre atributos da imagem e dos modelos.

O bloco Pré-processamento é constituído pelos módulos de:

- edição das descrições 3D dos objetos;
- construção da descrição 3D interna dos objetos;
- determinação dos planos de visualização para o cálculo das vistas;
- determinação das aparências, através da aglomeração de vistas isomórficas;
- geração da Hierarquia de Estratégias de busca por correspondência.

A configuração típica para uma aplicação do sistema RECTRI é ilustrada na figura 5.1, onde as descrições dos n objetos formam a base de dados, indicando todos os objetos reconhecíveis pelo RECTRI.

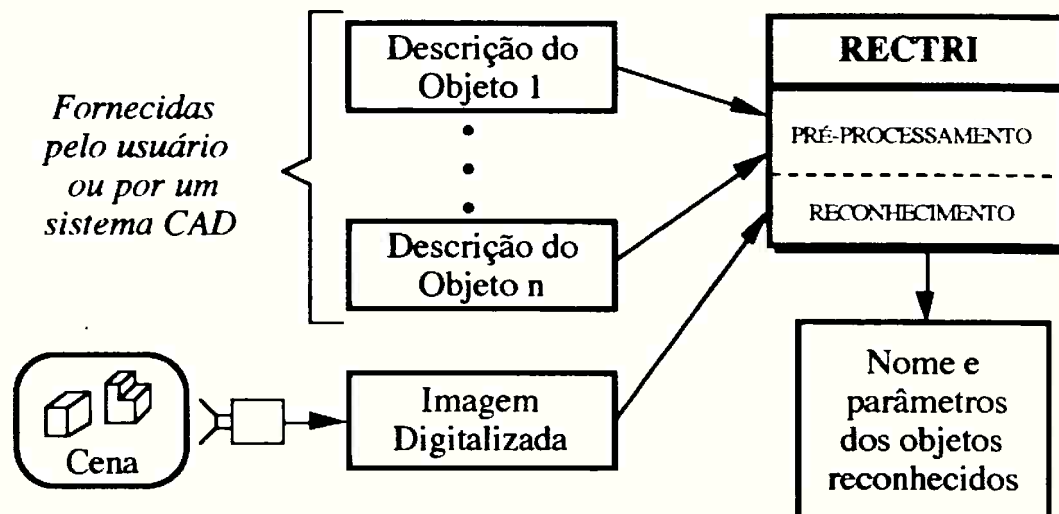


Fig.5.1 Configuração típica para atuação do sistema RECTRI de reconhecimento.

Uma cena formada por um ou mais objetos desta base de dados, dispostos aleatoriamente, é captada por uma câmara, formando uma imagem que, digitalizada, é processada, extraíndo-se atributos

da mesma. Esses atributos são comparados com os atributos extraídos dos objetos da base de dados, seguindo estratégias pré-definidas de correspondência. Hipóteses de correspondências são geradas a partir desta comparação, resultando em identidade (nome) e parâmetros iniciais para um processo de verificação de hipóteses e localização.

5.2. Domínio Escolhido

O domínio escolhido para a aplicação atual do sistema RECTRI consiste no reconhecimento de objetos poliédricos simples.

Esta escolha está baseada no fato de que sistemas que reconhecem objetos poliédricos possuem a maioria dos aspectos essenciais para análise e interpretação de uma cena, como formação da imagem, modelos, relações espaciais entre os objetos da cena (occlusão, sombreamento, etc), mas, no entanto, são bastante simples para modelar tridimensionalmente, extrair atributos, efetuar correspondência espacial, etc.

O conjunto de peças definido para a realização de testes e levantamento de dados do sistema RECTRI é constituído pelos objetos apresentados na figura 5.2 e foi escolhido com base nos objetos normalmente apresentados na literatura de reconhecimento de objetos poliédricos.

As peças foram nomeadas por (da esquerda para a direita, de cima para baixo): cubo, peça L, peça T, peça U, pirâmide, torre, prisma.

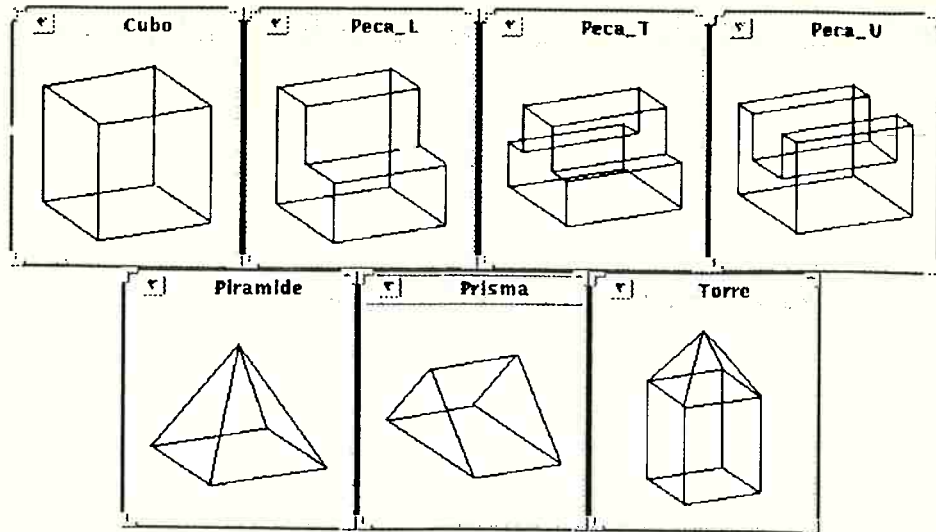


Fig.5.2 Conjunto de peças-teste do sistema RECTRI.

A descrição de cada peça é feita pelo usuário, seguindo um menu auto-explicativo, fornecendo os seguintes parâmetros:

- nome do arquivo para gravar a descrição de cada objeto, para que, numa segunda utilização, não seja necessário reeditar o mesmo objeto, bastando para isso ler o referido arquivo;
- quantidade de faces, arestas e vértices do objeto;
- quantidade e identificação das arestas que compõem cada face;
- identificação dos vértices que compõem cada aresta;
- coordenadas 3D de cada vértice;
- normal de cada face.

Esta descrição é armazenada internamente, seguindo uma estrutura pré-definida, contendo inclusive informações redundantes, para facilitar o acesso às informações durante a execução do RECTRI.

5.3. Geração do Conjunto de Aparências

A definição do conjunto das aparências é feita a partir da descrição interna (*modelo 3D*), gerada para cada objeto. O método utilizado é chamado *método por simulação*. Seu algoritmo consiste nos seguintes passos:

- colocar uma esfera gaussiana discretizada envolvendo o modelo do objeto, onde cada ponto da grade de discretização é considerado um ponto de observação;
- calcular a projeção do modelo 3D no plano de visualização tangente à esfera, passando pelo ponto de observação;
- aglomerar (fundir) as células com projeções isomórficas;
- quando não houver mais aglomeração possível, atinge-se o grafo de aspectos.

A esfera de visualização é discretizada segundo o modelo de discretização por cortes longitudinais e latitudinais, sendo que o número de cortes desejado em cada direção é fornecido pelo usuário. Quanto maior a granularidade na discretização da esfera de visualização, tanto maior será a precisão na resposta do sistema, porém maiores serão a capacidade de armazenamento requerida e o tempo de resposta do sistema.

As figuras 5.3 - 5.9 ilustram o conjunto de aparências definido para cada objeto, para uma esfera de visualização discretizada em dois cortes longitudinais e dois cortes latitudinais, resultando em dez planos de visualização, definidos como sendo tangentes à casca da esfera, nos pontos de cruzamento entre as linhas de corte.

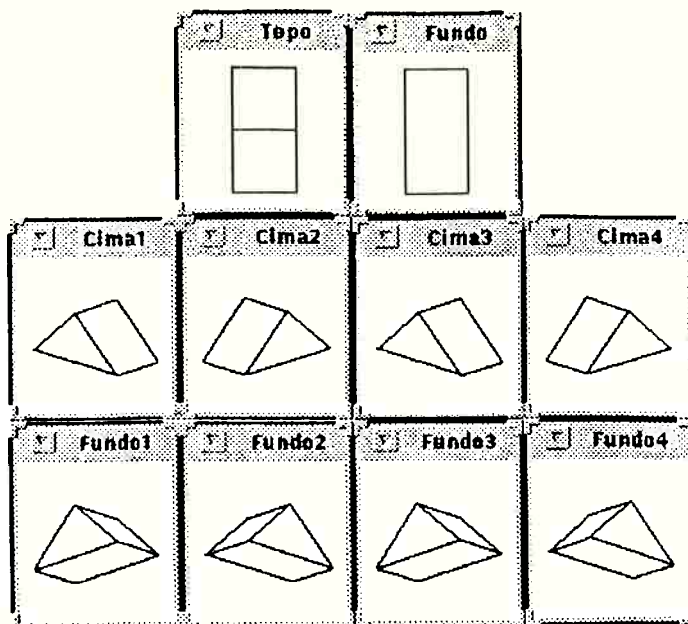


Fig.5.3 Conjunto de aparências definido para o prisma, com dez planos de visualização.

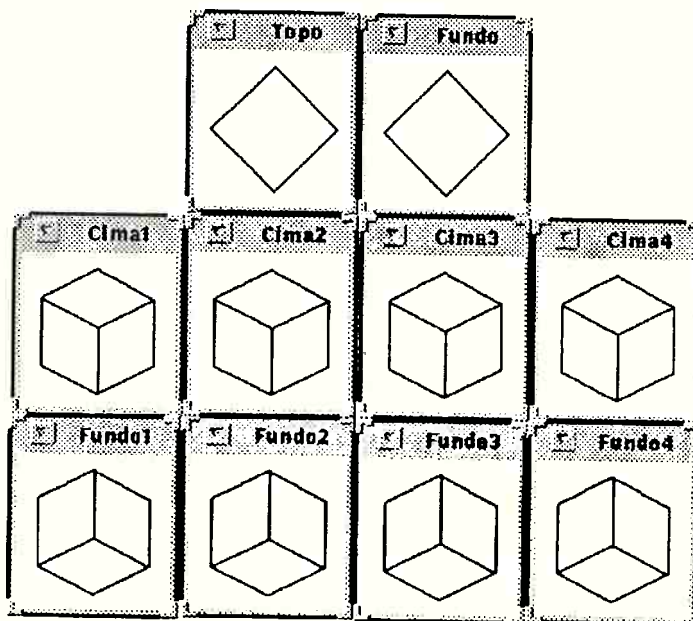


Fig.5.4 Conjunto de aparências definido para o cubo, com dez planos de visualização.

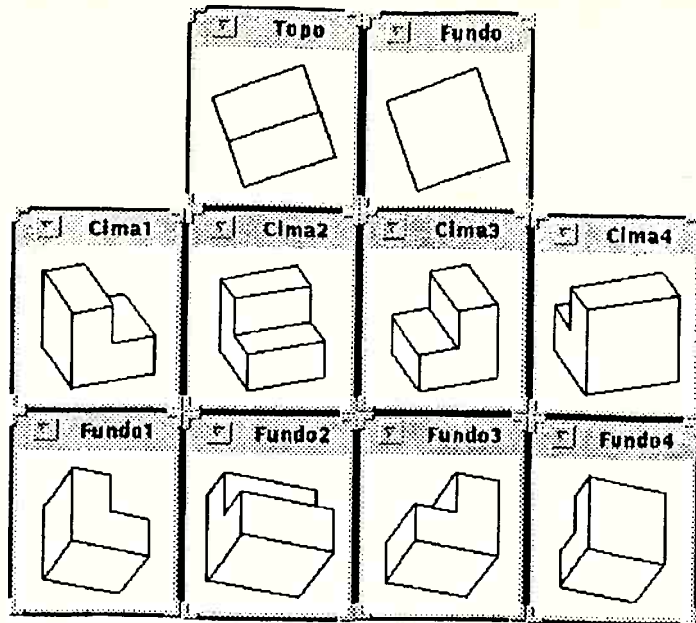


Fig.5.5 Conjunto de aparências definido para o peça-L, com dez planos de visualização.

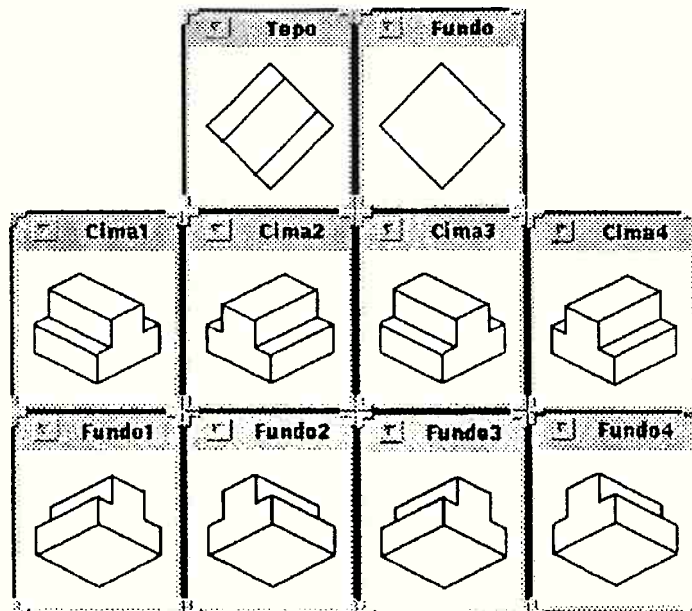


Fig.5.6 Conjunto de aparências definido para o peça-T, com dez planos de visualização.

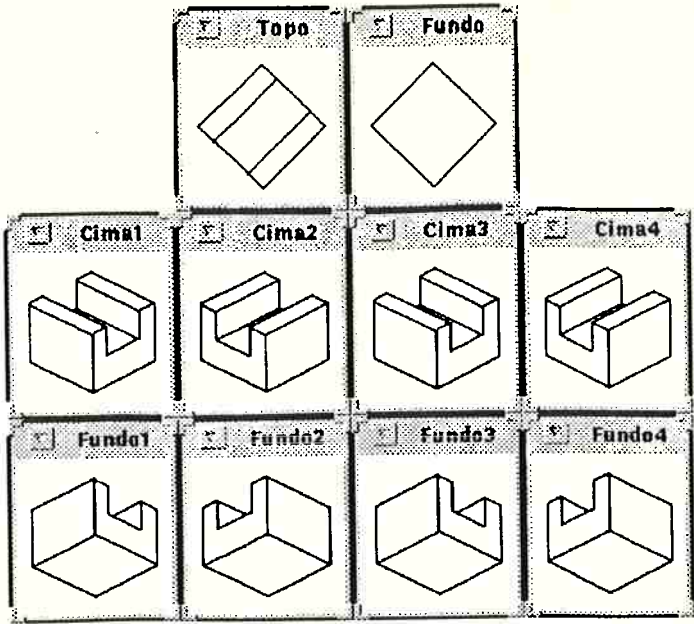


Fig.5.7 Conjunto de aparências definido para o peca-U, com dez planos de visualização.

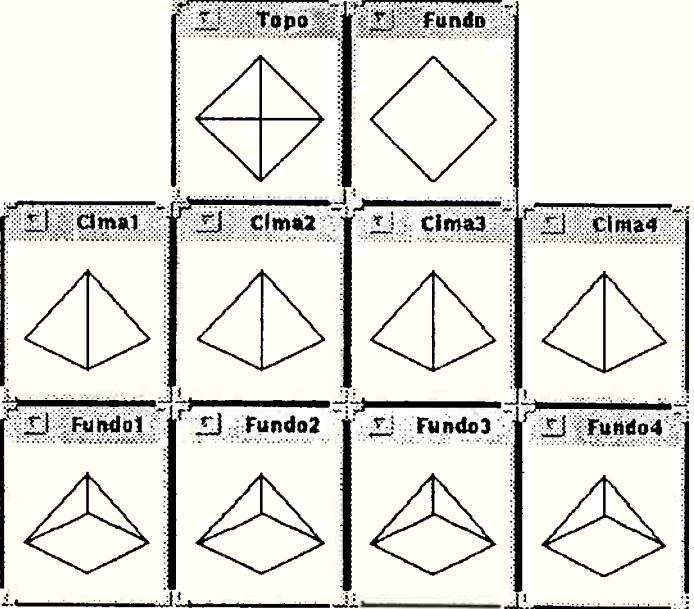


Fig.5.8 Conjunto de aparências definido para o pirâmide, com dez planos de visualização.

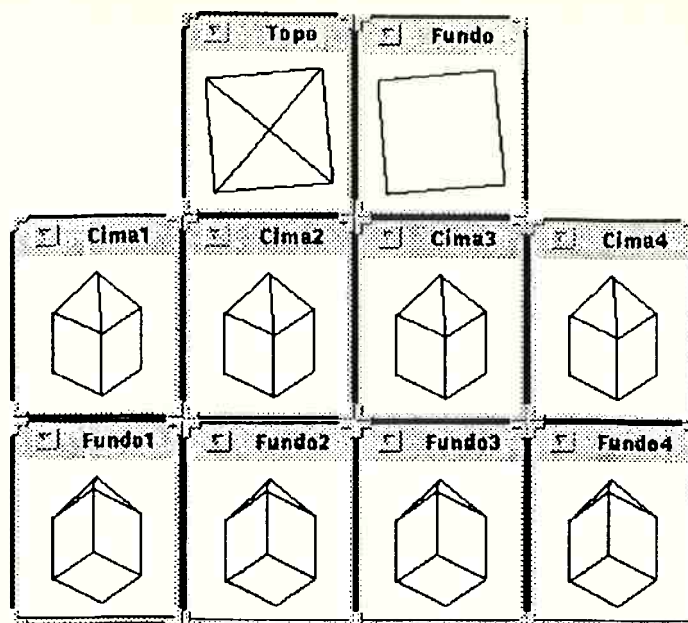


Fig.5.9 Conjunto de aparências definido para o torre, com dez planos de visualização.

5.4. Construção da Hierarquia de Estratégias

A hierarquia é construída através de um *procedimento iterativo, incremental e totalmente automático*, iniciando a construção nas folhas da hierarquia e terminando na raiz, para cada ramo, no sentido *inverso* ao utilizado durante o reconhecimento.

Cada folha está associada a uma aparência única de um objeto ou a aparências cujas diferenças em suas projeções não são representáveis e/ou perceptíveis ao sistema.

Cada nó, folha ou não, é composto por um ou mais atributos. Os atributos pertencentes a um nó são iterativamente retirados, um a um, formando outros nós com um menor número de atributos. O ramo é finalizado ao alcançar os nós mínimos, que contém somente

um atributo, formando o primeiro nível da hierarquia, apontado pela raiz.

Um critério de validade é aplicado a cada nó criado. São considerados válidos somente os nós que possuam conectividade entre todos os seus atributos. Conectividade pode ser representada por adjacência ou inclusão de atributos, conforme ilustrado na figura 5.10.

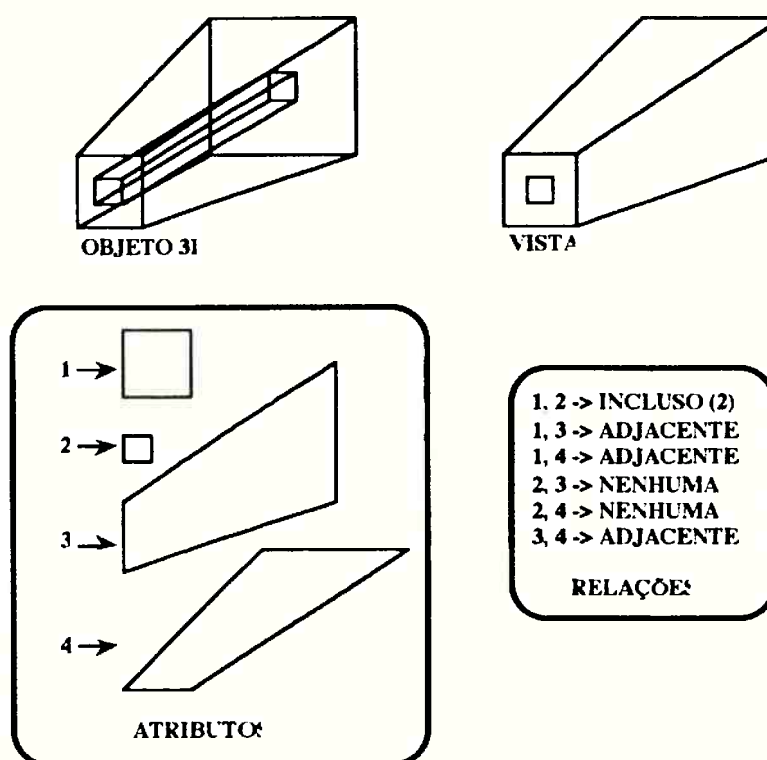


Fig.5.10 Condições de adjacência e inclusão entre atributos, aplicadas no critério de validade de nós.

Atualmente, somente a condição de adjacência entre atributos está implementada nesta versão do sistema RECTRI, sendo que dois atributos são considerados adjacentes quando pelo menos um de seus segmentos de contorno forem comuns. A figura

5.11 exemplifica nós válidos e inválido, pela condição de adjacência.

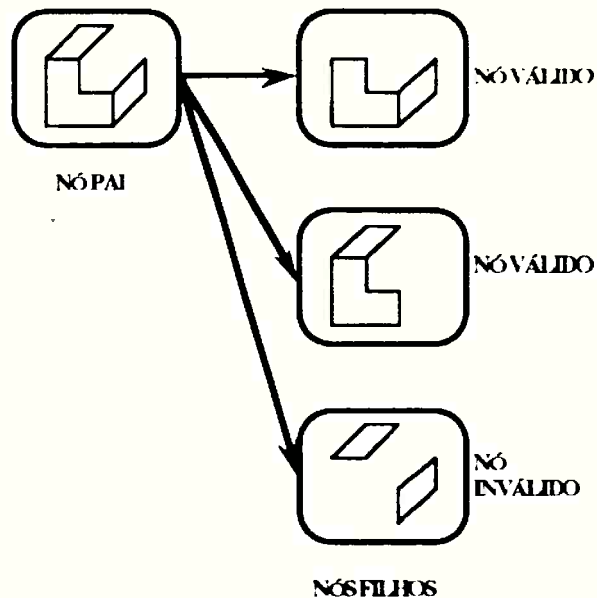


Fig.5.11 Critério de validade de um nó.

A condição de conectividade entre atributos é um fator importante no controle da busca por atributos na imagem, durante a fase de reconhecimento, e está baseado na suposição de que atributos conectados são mais prováveis de pertencerem a um mesmo objeto.

Cada nó válido criado é testado com todos os outros de mesmo número de propriedades já existentes na hierarquia. Nós correspondentes são aglomerados em um só nó. A correspondência entre dois nós é determinada por meio de uma busca recursiva numa árvore de equivalência entre atributos pertencentes aos nós testados, construída automaticamente.

Os nós postos em correspondência são inicialmente desmembrados em seus atributos constituintes. Cada nó da árvore

de equivalência representa dois atributos equivalentes, um de cada nó. A equivalência é determinada pelos seguintes critérios:

- tipo do atributo (triângulo, paralelogramo, etc);
- quantidade de atributo(s) vizinho(s) (adjacente);
- tipo(s) do(s) atributo(s) vizinho(s);
- número de segmentos adjacentes com cada atributo vizinho.

A árvore de equivalência tem a função de estabelecer critérios iniciais, fáceis de serem conferidos, para filtrar a passagem a um teste mais apurado dos pares de nós postos em correspondência. A árvore não analisa a relação de todos atributos pertencentes aos nós, mas só a de cada atributo individualmente. Se os nós não satisfizerem as condições de equivalência para construção da árvore, fica evidente que não se correspondem.

Caso exista uma árvore de equivalência para um par de nós postos em correspondência, os testes prosseguem, agora considerando incrementalmente cada par de atributos equivalentes, até que todos atributos pertencentes aos nós tenham sido analisados. Cada par de atributos equivalentes determinados pela árvore é acrescentado aos atributos já postos em correspondência, segundo três critérios:

- análise de *conectividade* entre os atributos (atributos vizinhos devem ser do mesmo tipo e a conectividade também de mesmo tipo e quantidade);
- análise da *condição de contorno* entre ambos os nós, para cada atributo agregado;
- análise do *paralelismo* mantido pelos segmentos pertencentes aos nós.

A figura 5.12 mostra um exemplo de dois nós postos em correspondência.

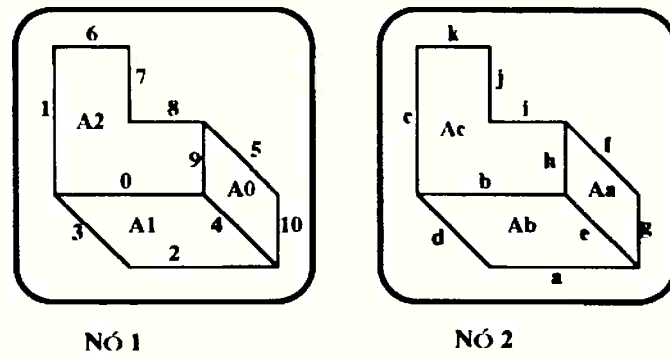


Fig.5.12 Exemplo de teste de correspondência entre dois nós.

A figura 5.13 mostra a árvore de equivalência resultante para estes nós.

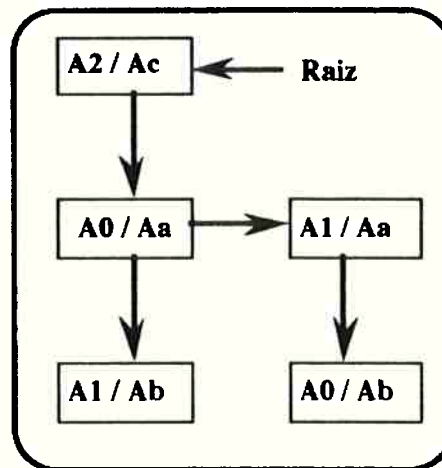


Fig.5.13 Árvore de equivalência resultante para o exemplo de correspondência dado na figura 5.12.

Uma busca é realizada na árvore de equivalência, com mecanismo de controle de busca em profundidade. Desse modo, inicialmente os atributos A2 (nó 1) e Ac (nó 2) são postos em correspondência. A condição de conectividade neste passo é

irrelevante. A condição de contorno indica que a correspondência entre os segmentos resulta:

```

0 1 6 7 8 9  ->  b c k j i h
                   c b h i j k

```

A condição de paralelismo indica, para cada nó:

```

Nó 1:    0 // 2 // 6 // 8
          1 // 7 // 9 // 10
          3 // 4 // 5

```

```

Nó 2:    a // b // i // k
          c // j // h // g
          d // e // f

```

Ao ser acrescida a correspondência entre os atributos A0 e Aa (segundo nó a ser acessado na árvore de equivalência), a condição de conectividade forçará a correspondência do segmento 9 do nó 1 com o segmento h do nó 2, tornando falsa a correspondência 016789 -> cbhijk, restringindo um pouco mais as correspondências possíveis entre os segmentos dos dois nós.

Enfim, o processo de correspondência entre dois nós prossegue, até que uma *contradição* seja encontrada, indicando que os nós não se correspondem, ou que sejam geradas correspondências válidas para os nós, segundo os critérios estabelecidos.

Caso encontre *contradição*, o novo nó criado é introduzido na hierarquia. Caso contrário, os nós são aglomerados, combinando as informações pertencentes ao nó criado (a que aparência pertence, número de filhos, etc) com aquelas do nó

correspondente já existente na hierarquia, e liberando o nó criado.

É também importante notar que a *restrição de rigidez* dos objetos está integrada na hierarquia, uma vez que direciona a busca não só para obter hipóteses de identidade, mas também de localização, esta representada pelos pontos de observação associados a cada aparência.

Alguns critérios de organização da hierarquia foram utilizados, juntamente com uma terminação heurística da busca por correspondência de atributos, visando agilizar o processo de reconhecimento, através da diminuição do espaço de busca.

O mecanismo utilizado de controle da busca por correspondência entre *imagem e modelos* é a *busca em profundidade*. De acordo com este mecanismo adotado, os critérios são aplicados, implícita ou explicitamente, para organizar a hierarquia.

Na atual versão do sistema RECTRI, somente alguns critérios de organização da hierarquia foram aplicados. Eles não são globais, para toda a hierarquia, porém somente a nível de ramos derivados de uma única aparência. A aplicação global de todos os critérios exige um segundo passo na construção, que seria o repasse total na hierarquia, reorganizando-a segundo um ou mais critérios globais, o que ainda não foi implementado no RECTRI.

Desse modo, durante a construção de um ou mais ramos derivados de uma aparência, foi aplicado inicialmente o critério de custo operacional de detecção de cada atributo. Desta forma, triângulos são consultados primeiro pela maior simplicidade de sua extração, seguidos por quadriláteros, trapézios, paralelogramos, contorno L, etc.

O critério de detectabilidade acidental foi imbutido na ordenação dos atributos selecionados e no ajuste dos parâmetros de agrupamento dos atributos. Desse modo, atributos constituídos por segmentos de maior tamanho são escolhidos primeiro, por considerar que segmentos maiores são mais prováveis de serem extraídos corretamente da imagem (menor probabilidade de representar ruído). Além disso, quanto mais rígido for o ajuste dos parâmetros (indicadores de paralelismo, colinearidade, junção, etc), menor será a possibilidade de detectá-los acidentalmente.

No entanto, considerando um nível de abstração maior na interpretação deste critério, a detectabilidade acidental de um atributo é dependente dos objetos (cena) que formarão a imagem e do processo de formação e aquisição da imagem. Conforme o sistema vai adquirindo maior experiência com tratamento de diversas imagens, pode-se determinar quais atributos são detectados acidentalmente e em que circunstâncias. Esta característica de aprender o grau de detectabilidade acidental dos atributos, neste maior nível de abstração, e de permitir que este conhecimento modifique a organização da hierarquia não se encontra implementada na versão atual do RECTRI.

Os critérios de maior freqüência de aparecimento de um atributo e o de saliência são antagônicos, pois se um atributo é muito freqüente, sua característica discriminatória é pequena. O primeiro é importante no primeiro nível da hierarquia, para garantir o acesso à mesma. O segundo normalmente apresenta um compromisso com o custo computacional de sua detecção, pois geralmente os atributos mais discriminatórios possuem um custo de detecção maior.

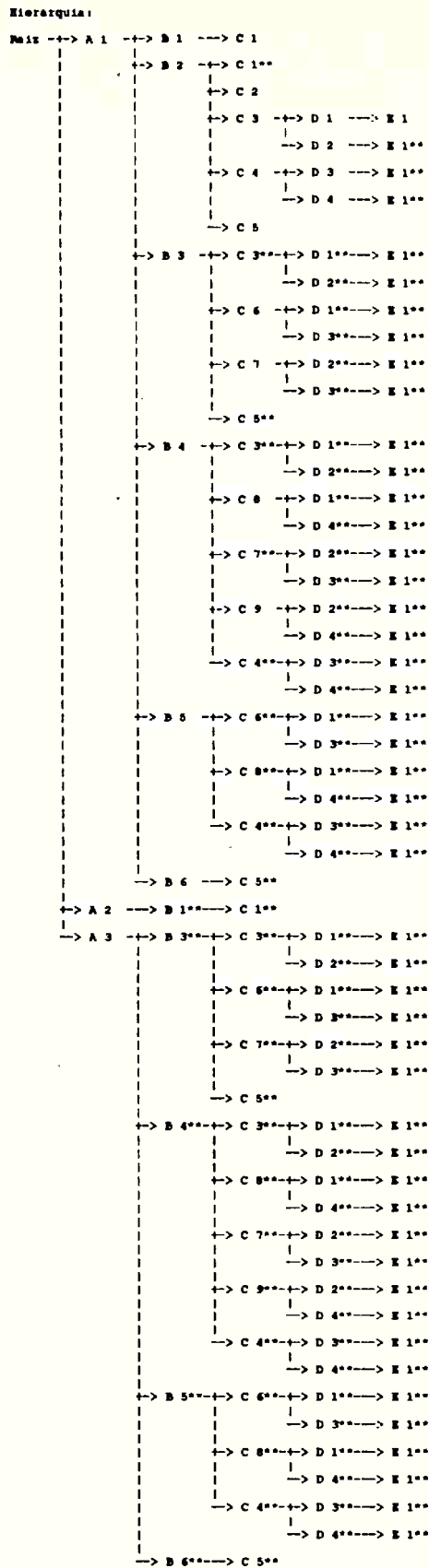


Fig.5.15 Hierarquia de Estratégias gerada para as aparências das peças Prisma, Cubo e Peça-U.

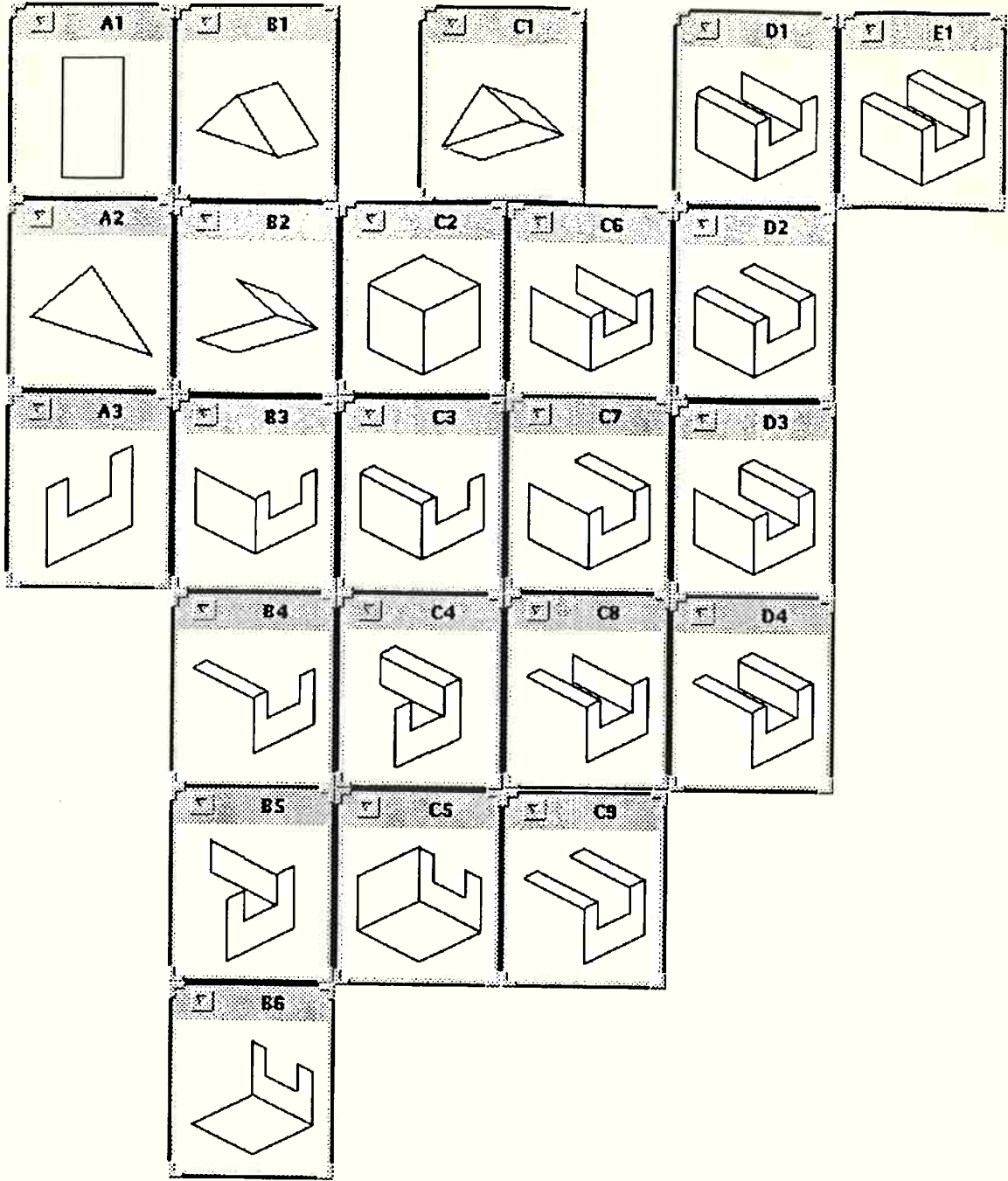


Fig.5.16 Nós da Hierarquia de Estratégias gerada para as aparências das peças Prisma, Cubo e Peça-U, mostrada na figura anterior.

5.5. Aquisição e Segmentação da Imagem

A aquisição de imagens de uma cena é realizada por uma câmera, tendo o processo de iluminação controlado para aumentar o contraste das bordas das peças e diminuir o efeito do sombreamento na imagem, a fim de facilitar a extração das bordas.

As bordas podem ser caracterizadas por mudanças bruscas da intensidade luminosa na imagem, determinando contornos de regiões que podem corresponder a superfícies ou descontinuidades de propriedades físicas de uma superfície.

Dessa forma, diversos algoritmos foram utilizados para detecção de descontinuidades na intensidade luminosa em imagens em nível de cinza. As imagens tratadas são do tipo das ilustradas na figura 5.17.

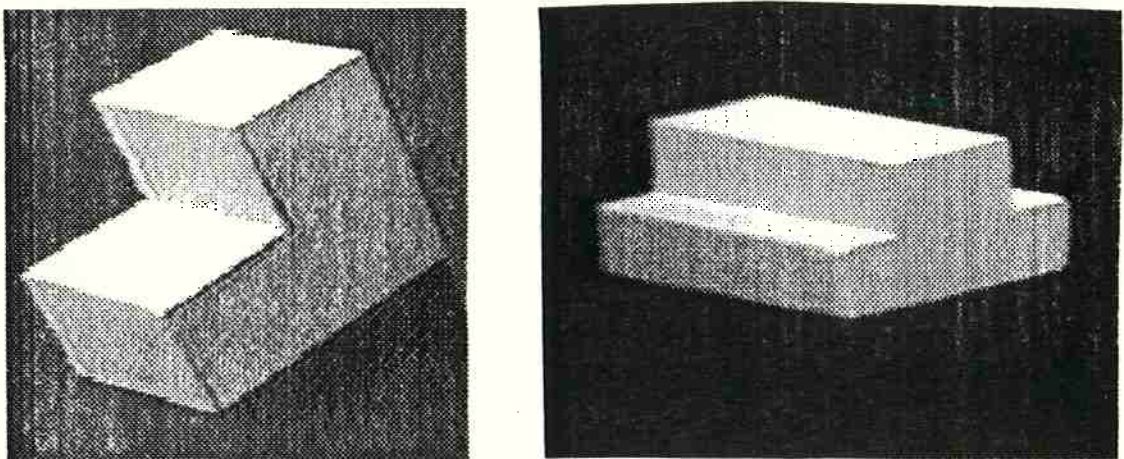


Fig.5.17 Imagens originais em níveis de cinza para as peças L e T.

Uma das abordagens utilizadas para segmentação da imagem é derivada da técnica de recozimento por campo médio ("mean field annealing") [Geiger, Yuille 1991]. O modelo utilizado é o modelo de membrana fraca ("weak membrane model") [Munford, Shah 1985], inserido em um paradigma estatístico [Geman, Geman 1984]. O algoritmo de recozimento por campo médio [Geiger, Yuille 1991] foi implementado por Geiger e Girosi [Geiger, Girosi 1991]*. Os seguintes parâmetros foram utilizados, para a referida implementação:

- número de iterações: 10;
- parâmetro de suavidade: 2.5;
- custo de produção de bordas: 240.1;
- parâmetro de escala inicial (Beta): 0.0001
- parâmetro de escala final (Beta): 1.0.

O resultado obtido está ilustrado na figura 5.18, para as imagens da figura 5.17.

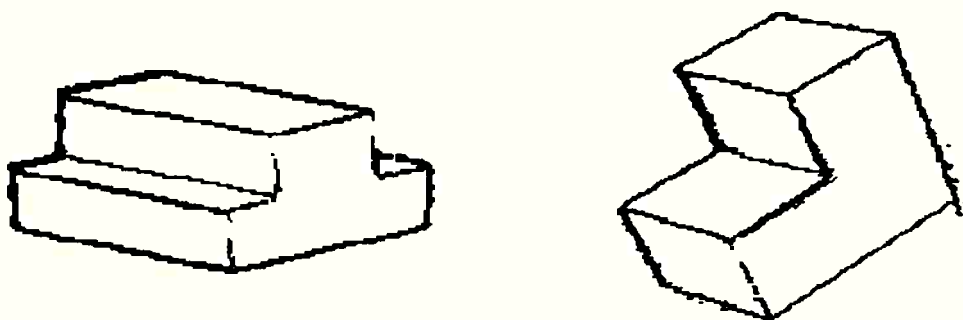


Fig.5.18 Resultado da aplicação do programa de [Geiger, Girosi 1991] nas imagens da figura 5.17.

* O algoritmo foi adaptado por João Kogler e Cláudio Katsumi, ambos da DAIA/LSI, para executar em ambiente sunview e depois novamente adaptado por João Kogler para executar em ambiente Xwindow.

Resultados bastante impressionantes também foram conseguidos através de uma abordagem feita utilizando Morfologia Matemática. Esta abordagem consiste no algoritmo de *partição de águas* [Beucheur, Vincent 1990], implementado no ambiente KHOROS por Júnior Barrera e outros [Barrera et al. 1994], com parâmetros ajustados por Júnior Barrera para as imagens tratadas. Uma grande vantagem apresentada por este algoritmo é a garantia de obtenção de contornos fechados para as bordas. Os resultados da aplicação deste algoritmo estão ilustrados na figura 5.19.

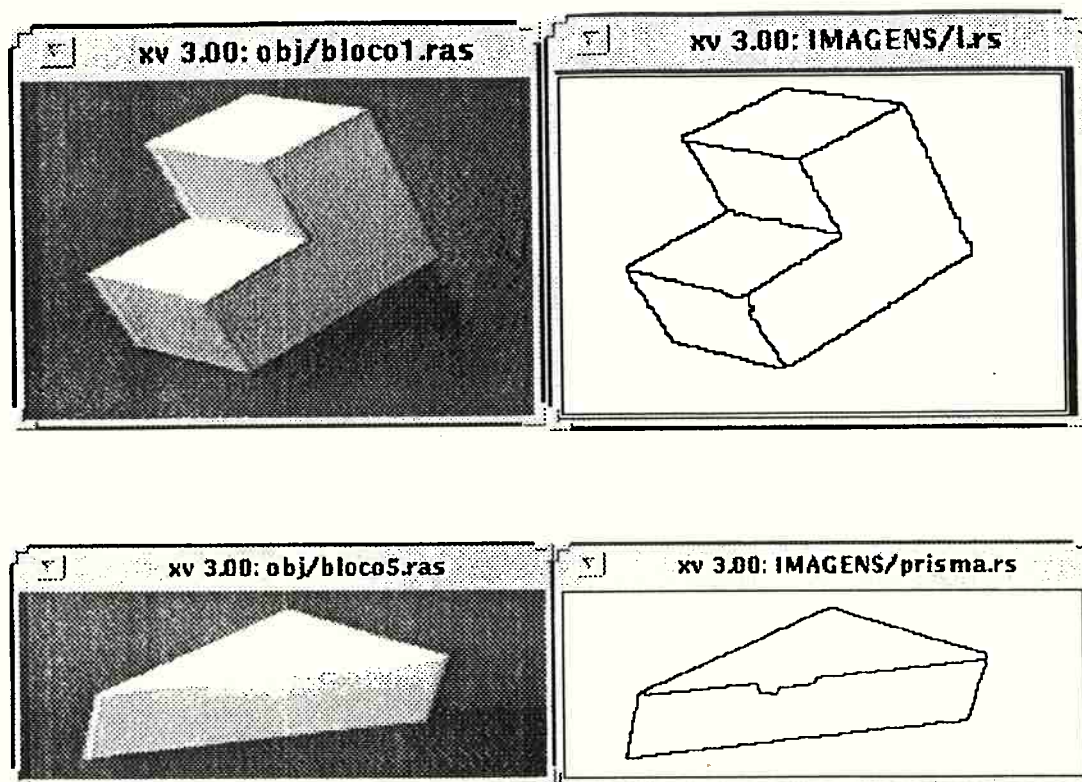


Fig.5.19 Resultado do algoritmo de morfologia matemática para as peças L e prisma* .

* Agradeço ao Prof. Dr. Júnior Barrera (IME / USP) pela gentileza de ceder estas imagens.

Após a realização deste processo de detecção de contornos, pode-se obter uma matriz de valores binários, onde os pontos de contorno possuem valor 1 e os outros pontos, valor 0.

No entanto, é necessário que se represente esses pontos de modo conectado, num nível de abstração maior do que o fornecido por uma matriz. Um processo de conexão é realizado, gerando listas de pontos conectados, formando segmentos. Este é um processo de busca local, onde a conectividade é analisada de acordo com uma vizinhança, para um modelo quadriculado de matriz.

Esses segmentos (cadeia de pontos conectados) são aproximados para segmentos de reta, seguindo o critério abaixo:

- gerar um segmento de reta entre os pontos terminais do segmento analisado;
- calcular o ponto mais distante do segmento analisado em relação ao segmento de reta gerado;
- se a distância entre eles for maior do que um valor previamente estipulado, dividir o segmento de reta em dois e repetir o procedimento para os dois segmentos gerados.

Este algoritmo está ilustrado na figura 5.20. O conceito utilizado de "distância" entre segmentos está na verdade baseado na relação de comprimento do segmento de reta pelo desvio entre a reta e o ponto mais distante, resultando numa medida mais precisa.

Todo esse processo de detecção de segmentos de reta nas imagens foi realizado utilizando algoritmos já desenvolvidos [Ballard, Brown 1982; Gonzalez, Wintz 1987; Lowe 1985]. O

desenvolvimento dos algoritmos do sistema RECTRI partiu do pressuposto que esse tipo de informação já estivesse disponível.

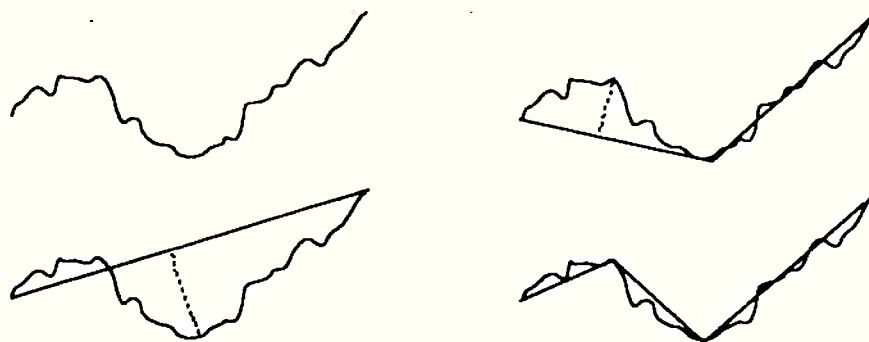


Fig.5.20 Aproximação poligonal.

As figuras 5.21 e 5.22 ilustram os resultados obtidos com a aproximação poligonal descrita para as imagens segmentadas mostradas nas figuras 5.18 e 5.19.

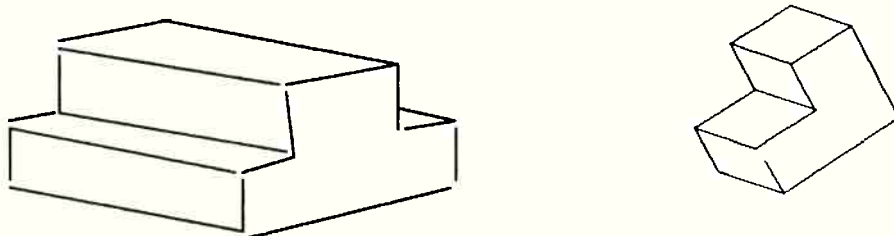


Fig.5.21 Aproximação poligonal das imagens da figura 5.18.

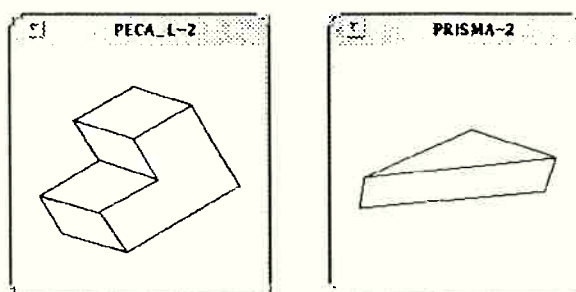


Fig.5.22 Aproximação poligonal das imagens da figura 5.19.

5.6. Extração e Agrupamento de Primitivas

Inicialmente, os segmentos de reta detectados na imagem são ordenados por tamanho, considerando que quanto maiores forem os segmentos, menor a probabilidade de constituírem ruído, portanto, maior sua importância para o sistema RECTRI.

A lista ordenada de segmentos de reta da imagem é percorrida, buscando relações que indiquem colinearidade, paralelismo e conectividade entre os segmentos.

Dois segmentos de reta são considerados colineares quando a quantidade de *superposição* entre eles for *pequena* (se existir), apresentarem *aproximadamente a mesma inclinação* e as *distâncias* entre eles, no sentido transversal e no sentido longitudinal, forem *pequenas* (figura 5.23).

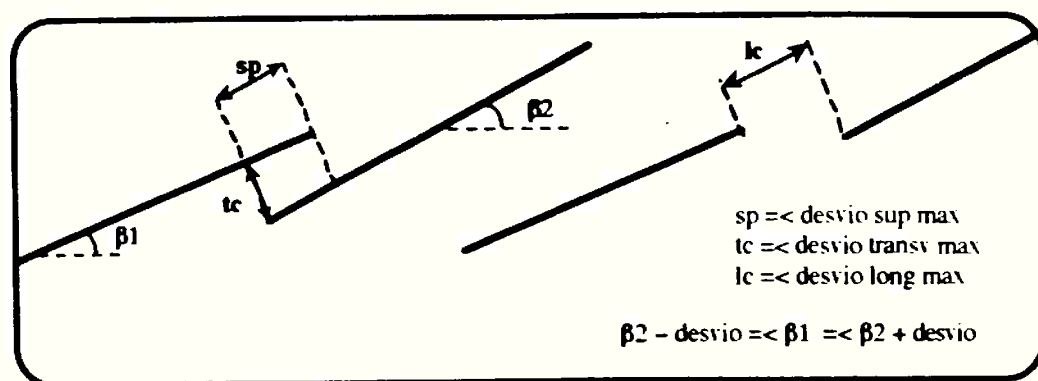


Fig.5.23 Condição de colinearidade.

Conectividade de segmentos é definida através de *junções*. Uma *junção* é um conjunto de segmentos de reta terminando numa

região comum, aqui definida por um círculo de raio R (figura 5.24).

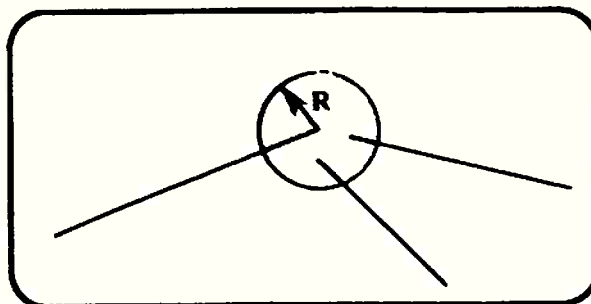


Fig.5.24 Condição de conectividade (junções).

Dois segmentos de reta são considerados paralelos quando a magnitude da *diferença angular* entre suas orientações for *pequena* e eles não estiverem *muito próximos*, na direção transversal, conforme ilustrado na figura 5.25.

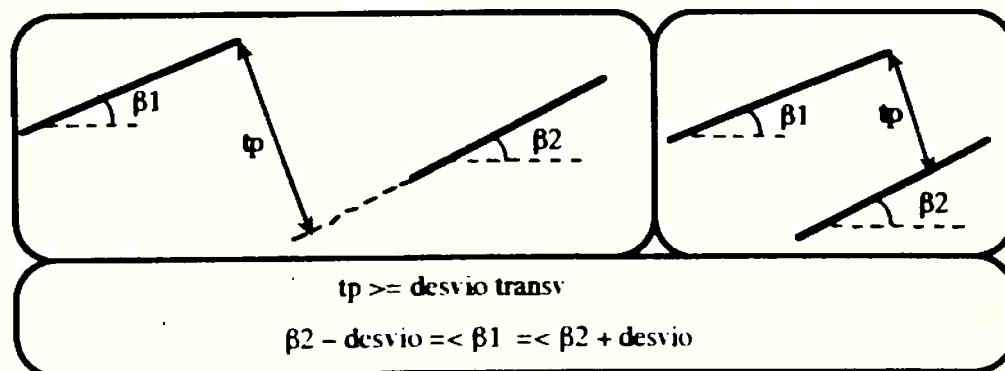


Fig.5.25 Condição de paralelismo.

Todos esses valores limites (grande, pequeno, próximos, R , etc) são definidos por meio de parâmetros, fornecidos pelo usuário. Atualmente, esses parâmetros são definidos empiricamente, considerando os objetos usados na cena, a qualidade das imagens adquiridas e as informações delas retiradas (proporção do tamanho dos objetos na imagem, etc).

A figura 5.26 ilustra a determinação dessas relações primárias (colinearidade, paralelismo, junção) para uma imagem.

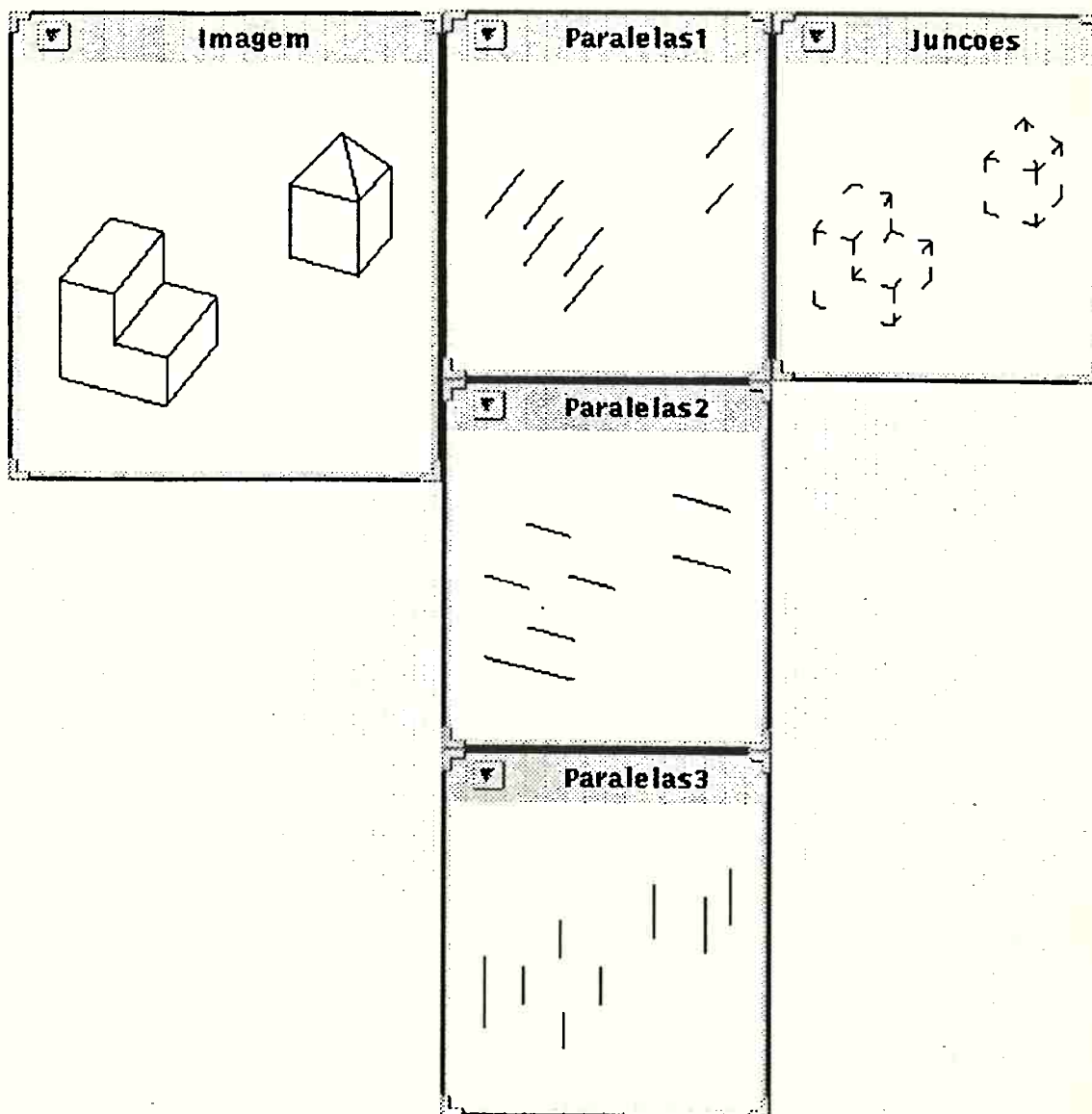


Fig.5.26 Relações primárias detectadas entre segmentos de uma imagem.

Sobre estas relações primárias determinadas, agrupamentos mais expressivos são realizados, buscando conectar segmentos em cadeias maiores e verificando propriedades nestas cadeias, como: se formam ou não cadeias fechadas, de que forma a propriedade de paralelismo entre seus elementos constituintes é mantida, etc. Deste modo, são extraídas figuras como triângulos, trapézios, paralelogramas, entre outras (figura 5.27).

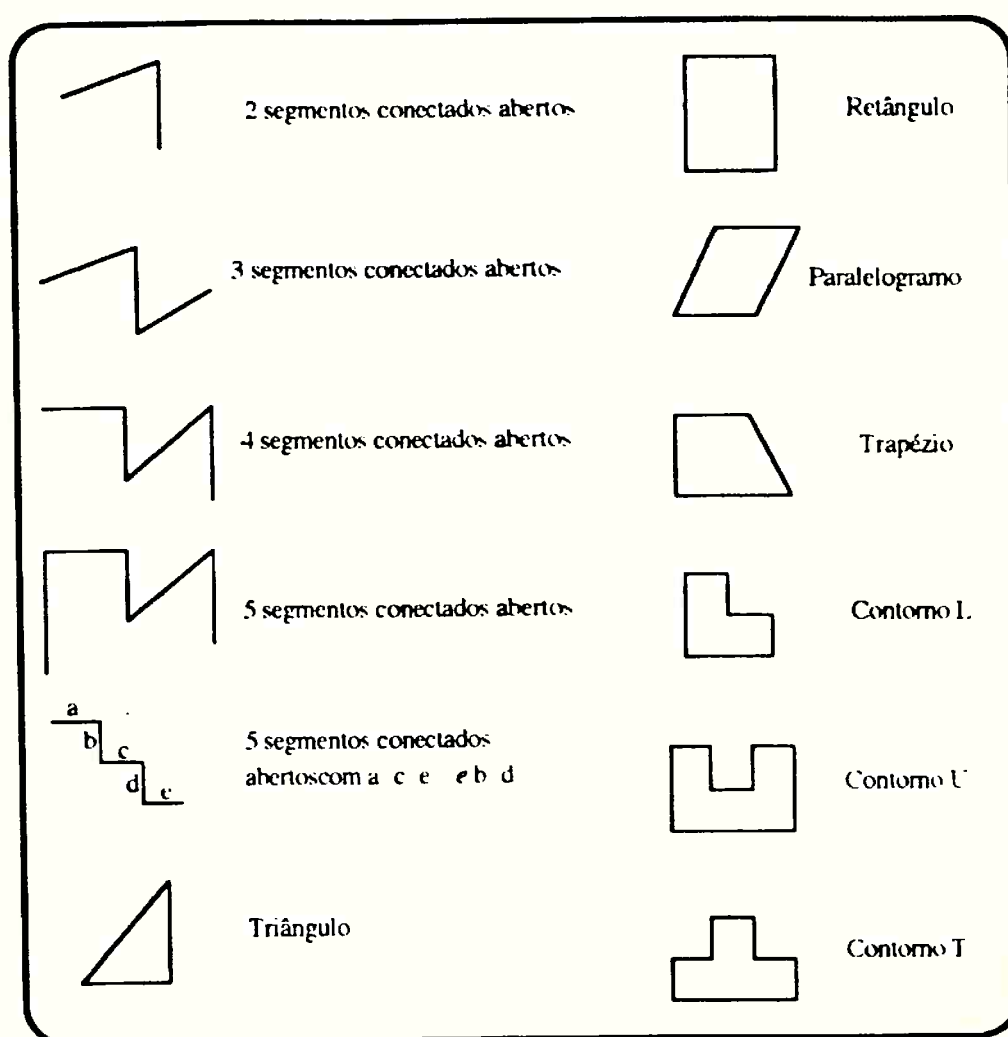


Fig.5.27 Agrupamentos atualmente tratados por RECTRI.

A figura 5.28 ilustra os agrupamentos extraídos do exemplo da figura 5.26.

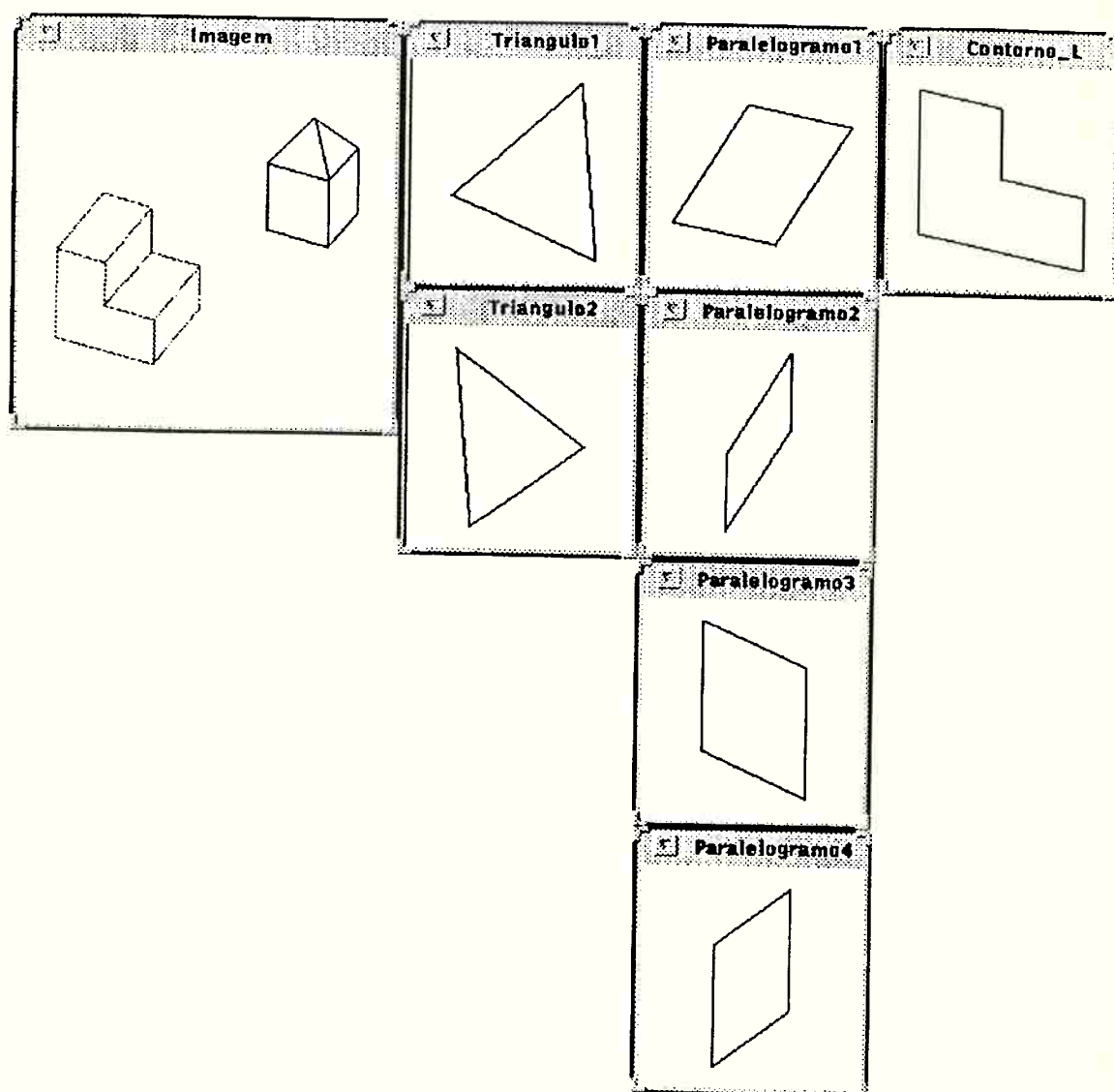


Fig.5.28 Agrupamentos extraídos do exemplo da figura 5.26.

5.7. Busca por Correspondência de Atributos

A partir dos agrupamentos extraídos da imagem, aqui também chamados de atributos, o sistema RECTRI efetua uma busca por correspondência entre estes atributos extraídos da imagem e aqueles, de mesmo tipo, armazenados nos modelos.

No entanto, esta busca é conduzida pela Hierarquia de Estratégias, visando direcioná-la. A hierarquia dita qual o próximo atributo a buscar e como ele deve estar conectado aos atributos já considerados até aquele momento da correspondência.

Deste modo, de acordo com o tipo e conectividade do atributo indicado pela hierarquia, uma lista de atributos candidatos a satisfazer estes requisitos é criada. Os atributos são ordenados de acordo com o tamanho dos seus respectivos maiores segmentos constituintes (aquele atributo de tipo e conectividade conforme requerido que possuir o maior segmento será o primeiro da lista e assim por diante).

Dado que exista pelo menos um atributo que satisfaça os requisitos ditados pelo respectivo nó da hierarquia, um nó *imagem* é construído e posto em correspondência com o respectivo nó da hierarquia. O processo de correspondência entre nós (da imagem e da hierarquia) é exatamente o mesmo descrito para a correspondência entre um novo nó criado no processo de construção da hierarquia e os nós de mesmo número de atributos, já existentes na hierarquia, processo este descrito no item 5.4.

Uma terminação heurística da busca por correspondência foi introduzida, visando agilizar o processo de reconhecimento. Esta

terminação foi estipulada para ser realizada assim que uma folha da hierarquia seja alcançada.

Todos os nós correspondidos durante o processo de busca são armazenados numa estrutura, juntamente com suas respectivas correspondências segmento modelo-segmento imagem.

Caso uma folha seja alcançada, sua respectiva resolução de correspondência é mostrada, indicando quais segmentos da imagem foram correspondidos a quais segmentos de quais aparências (figura 5.29).

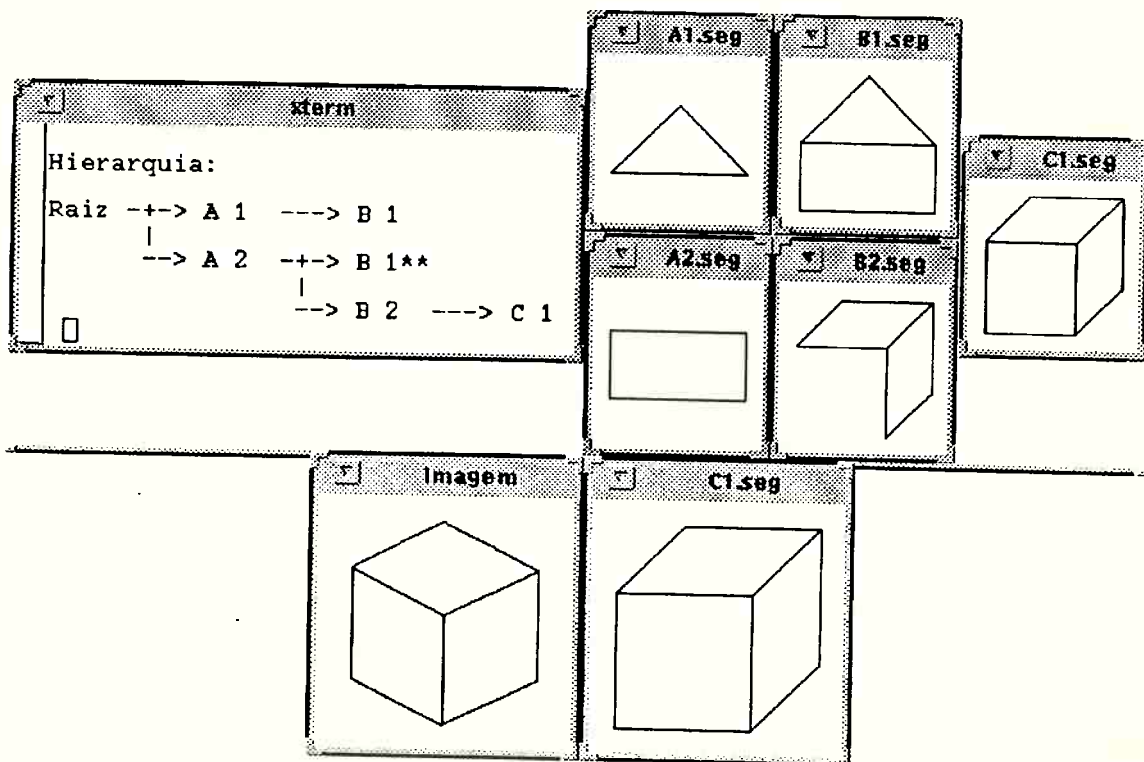


Fig.5.29 Reconhecimento onde uma folha da hierarquia mostrada foi alcançada, para uma imagem do cubo.

Da informação armazenada sobre cada aparência, extrai-se facilmente o objeto ao qual ela faz parte e de que região de visualização ele deve ter sido observado para gerar tal aparência.

Se nenhuma folha for alcançada, e todas as possibilidades de evolução da busca já foram realizadas, o sistema RECTRI deve consultar o espaço de correspondências gerado e mostrar aquela que possuir o maior número de atributos correspondido, indicando quais segmentos da imagem correspondem a quais segmentos de quais aparências incompletas. A figura 5.30 ilustra um caso deste tipo.

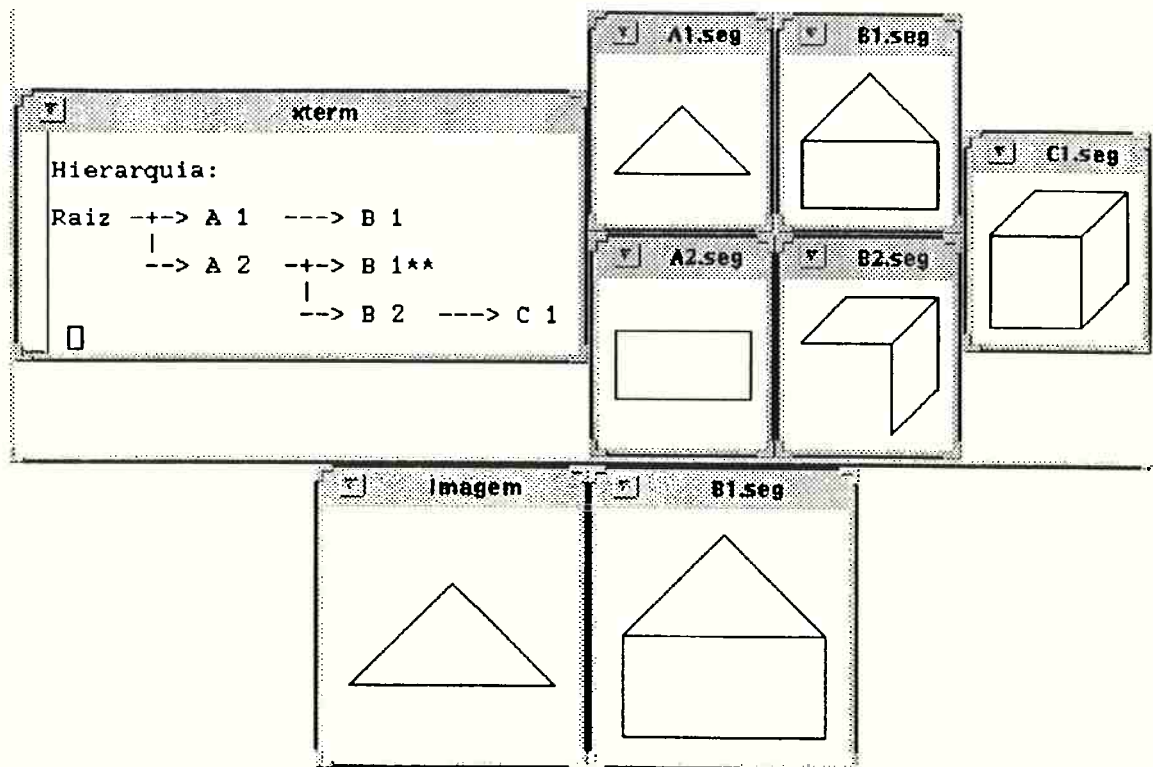


Fig.5.30 Reconhecimento onde nenhuma folha da hierarquia foi alcançada, para a hierarquia aqui ilustrada.

Um processo de verificação deve então ser aplicado sobre os candidatos apresentados pelo reconhecimento, tanto no caso de ter alcançado uma folha quanto naquele em que a busca foi exaustiva, para confirmação ou refuta do reconhecimento.

Antes da aplicação de um processo de localização, alguns processos intermediários podem ser realizados. Uma sugestão seria a introdução de uma heurística fazendo com que, por exemplo, somente os maiores três segmentos não colineares sejam postos em correspondência espacial. A partir deste estágio, o processo de verificação seria aplicado, procurando utilizar todas as evidências disponíveis.

Outra sugestão seria a de procurar correspondências num nível de abstração mais baixo, a partir das hipóteses geradas pelo RECTRI, introduzindo, por exemplo, métricas nos atributos, como proporção entre tamanho dos segmentos constituintes de faces reconhecidas, momentos invariantes, etc.

Caso haja confirmação de uma hipótese, um processo mais apurado de localização deve ser realizado, visando parâmetros de posicionamento mais precisos.

Caso o reconhecimento seja recusado, deve-se prosseguir a busca guiada pela hierarquia. Se a busca já tiver percorrido toda a hierarquia, atributos mais incompletos devem entrar no processo de correspondência, como, por exemplo, cadeias abertas de segmentos; este procedimento, de um certo modo, pode ser entendido como uma diminuição do nível de abstração dos atributos considerados.

Todas estas últimas considerações não se encontram disponíveis na atual versão do sistema RECTRI.

5.8. Apresentação e Análise dos Resultados

Quase a totalidade dos testes do sistema RECTRI foram desenvolvidos para dados simulados. Os dados simulados usados descrevem cenas em termos dos segmentos de reta das bordas das peças constituintes, podendo essa descrição gerar desenho de linhas perfeitos ou não.

O interesse em imagens que não tenham o desenho de linhas perfeito existe para que seja evidenciado o fato de RECTRI aceitar segmentações não perfeitas das imagens.

No entanto, alguns testes foram realizados com imagens reais, mostrando que as descrições geradas por simulação são condizentes com os resultados reais. Os resultados dos processos de segmentação de imagens apresentados nas figuras 5.12 e 5.13 mostram que o desenho de linhas perfeito para o tipo de imagem tratado não é difícil de ser alcançado.

5.8.1. Metodologia da Geração de Dados Simulados

O processo de gerar dados simulados para o sistema RECTRI consiste em aplicar transformações (rotações e translações em x , y e z do sistema de coordenadas de cada objeto) nas descrições 3D de cada objeto constituinte da cena simulada. Se houver mais de um objeto na cena, seus sistemas de coordenadas devem ser forçados a coincidir.

Um plano de visualização é escolhido e a cena é projetada neste plano, tendo eliminadas suas linhas escondidas. Dessa forma, é gerado o desenho de linhas perfeito para a referida cena. Manualmente introduz-se, se desejado, alguma falha na segmentação (geralmente, apagando linhas).

A descrição final gerada neste processo é uma lista de segmentos de reta, onde cada segmento é descrito por quatro coordenadas: x e y dos pontos terminais do segmento, para um sistema de coordenadas qualquer adotado para o plano de visualização.

5.8.2. Apresentação dos Resultados

Durante a implementação do sistema, cada módulo foi testado separadamente, avaliando se as propriedades de interfaceamento estavam conforme o especificado e também se a especificação estava correta até o momento. Assim, vários testes foram realizados para o módulo de extração de relações primárias e agrupamentos de segmentos, para a geração do conjunto de aparências, para a geração da Hierarquia de Estratégias e para o processo de correspondência.

Uma certa dificuldade foi encontrada para a visualização dos resultados de cada módulo, principalmente o da geração da Hierarquia de Estratégias e o da correspondência.

O problema na visualização da hierarquia foi contornado traduzindo os dados numa estrutura gráfica simples de entender, onde seus nós são mostrados através da descrição de seus segmentos constituintes, na forma de desenho. O problema nessa

representação consiste no elevado número de janelas abertas na estação, uma vez que cada nó é uma janela.

Com relação à visualização do resultado da correspondência, optou-se por mostrar a aparência para a qual a correspondência se estabeleceu, do mesmo modo que cada nó da hierarquia foi mostrado. Foi utilizado um mesmo tom de cinza (ou cor) para cada conjunto de segmentos da imagem correspondido a uma determinada aparência. Deve-se ressaltar que as correspondências normalmente não são únicas, mas sim diversas, indicando várias hipóteses. Para os propósitos de visualização, optou-se por mostrar a mais completa (que comportar o maior número de segmentos correspondido) e, se mais de uma satisfizer este requisito, adotou-se a heurística de mostrar a primeira encontrada.

Assim, nas figuras apresentadas a seguir tem-se a imagem (geralmente formada por dados simulados), o nó da hierarquia com o maior número de agrupamentos correspondido à imagem, e as aparências relacionadas àquele nó. Se mais de um nó for correspondido, são atribuídos diferentes tons de cinza para cada conjunto de segmentos da imagem e o respectivo nó a ele correspondido.

5.8.3. Resultados

Além dos resultados já apresentados como exemplos, no decorrer do texto, alguns outros são aqui ilustrados, para mostrar melhor a atuação do RECTRI.

A figura 5.31 mostra a imagem (real) da peça L, o respectivo nó da hierarquia a ele correspondido e as duas aparências

relacionadas a esse nó. A hierarquia usada foi aquela incluindo todas as aparências definidas da figura 5.3 à figura 5.9.

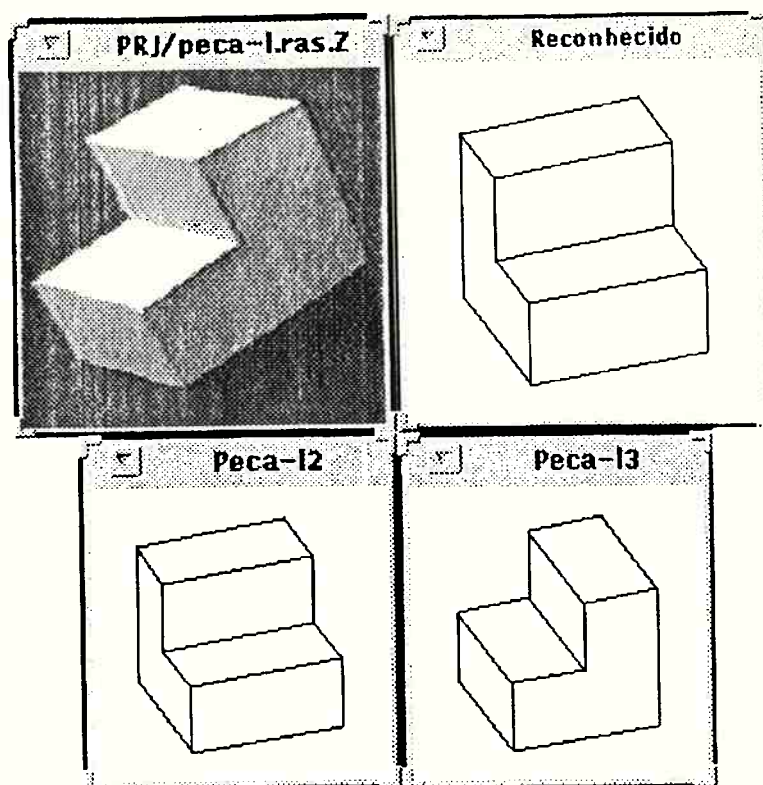


Fig.5.31 Reconhecimento efetuado para a peça L da imagem mostrada, indicando o nó correspondente (Reconhecido) e as respectivas aparências (Peça-12 e Peça-13) a ele associadas.

A figura 5.32 ilustra uma imagem onde a peça L e a torre se encontram isoladas e os respectivos agrupamentos primários apresentados por RECTRI. Na figura 5.33 tem-se os agrupamentos intermediários resultantes.

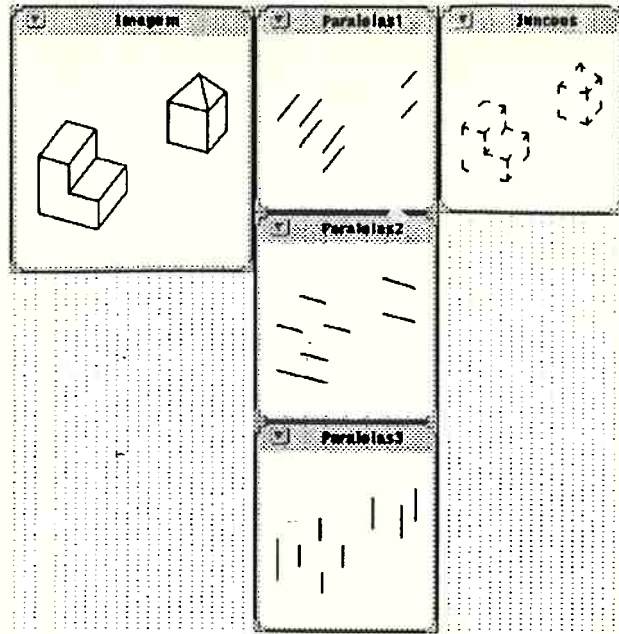


Fig.5.32 Agrupamentos primários detectados para a imagem da peça L e a torre isoladas.

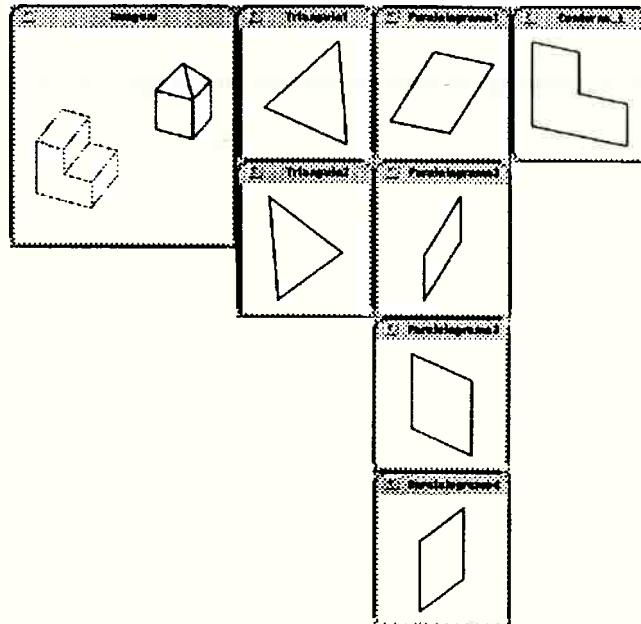


Fig.5.33 Agrupamentos definidos para a imagem apresentada.

A figura 5.34 mostra o resultado do reconhecimento realizado, com a hierarquia completa. Notar que o sistema mostra diferentes cores (níveis de cinza) para cada conjunto de segmentos em correspondência entre imagem e aparência.

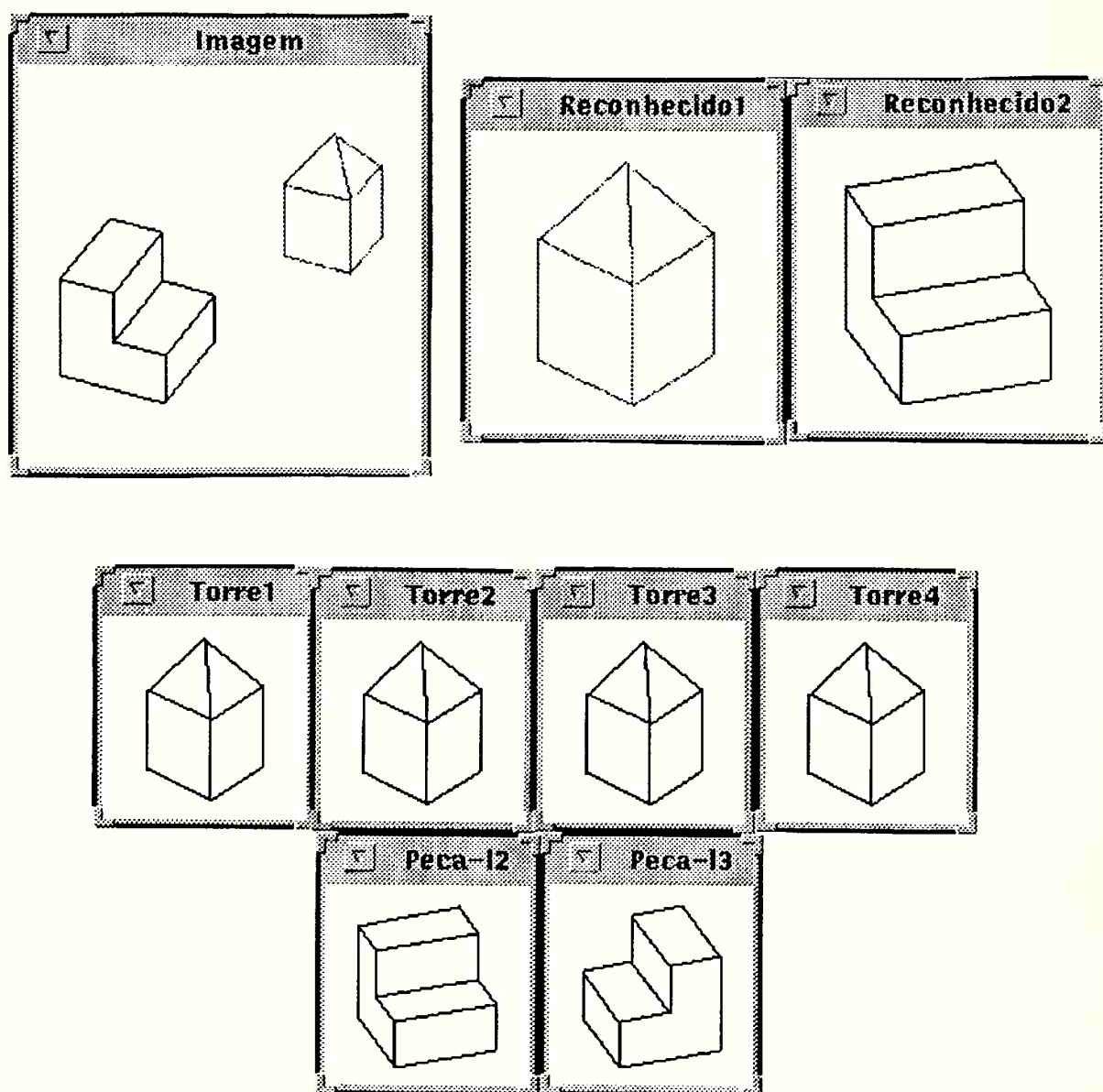


Fig.5.34 Peças L e torre reconhecidas, indicando nós correspondentes da hierarquia e aparências associadas.

A figura 5.35 mostra uma imagem onde se encontram as peças L e cubo, isoladas, com os respectivos resultados do agrupamento primário e, na figura 5.36, com os agrupamentos intermediários.

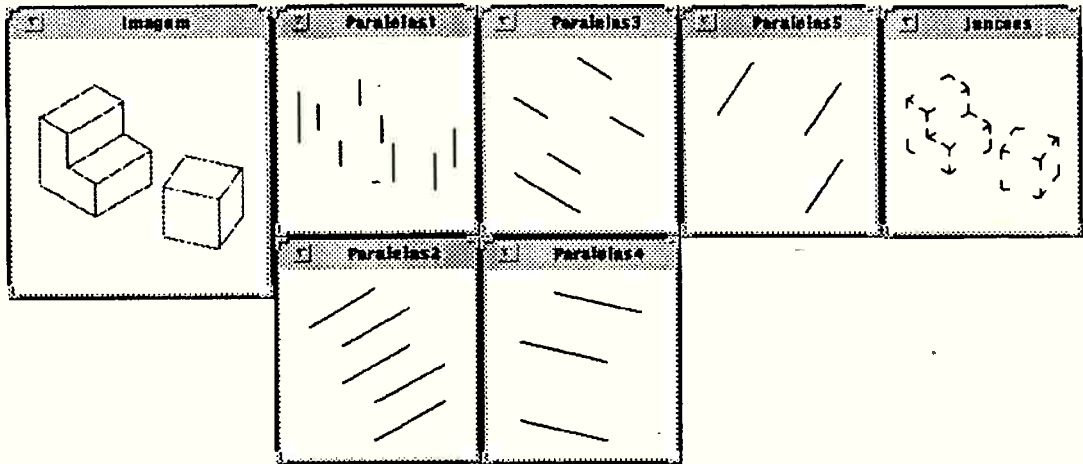


Fig.5.35 Agrupamentos primários para a imagem das peças L e cubo, isoladas.

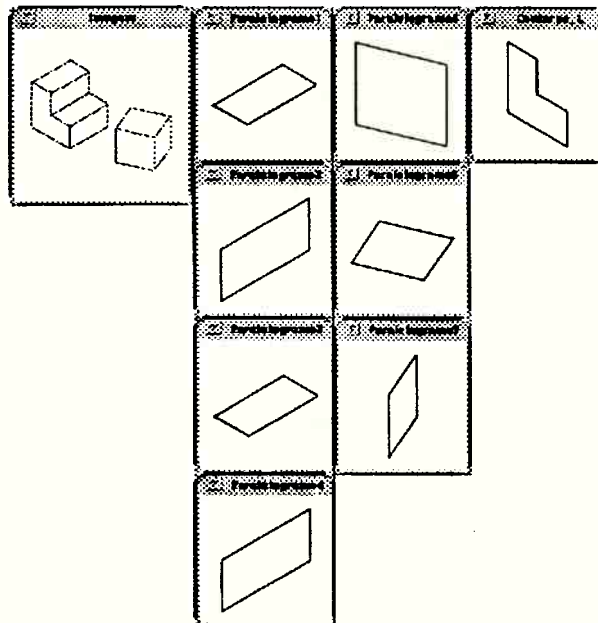


Fig.5.36 Agrupamentos intermediários para a imagem das peças L e cubo.

O resultado do reconhecimento está ilustrado na figura 5.37, para a hierarquia completa.

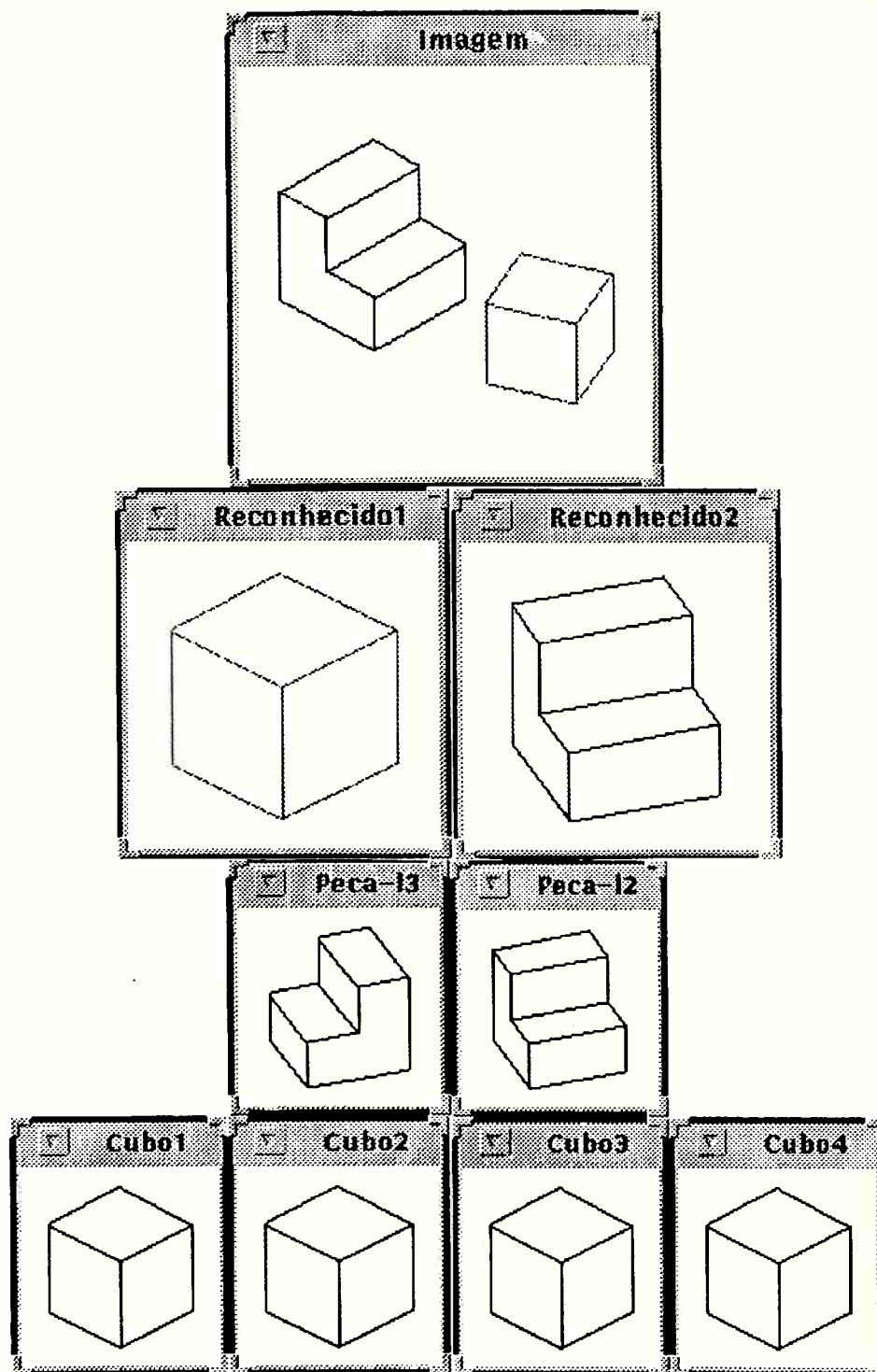


Fig.5.37 Peças L e cubo reconhecidas, indicando nós correspondentes da hierarquia e aparências associadas.

As figuras 5.38 e 5.39 mostram os resultados, para a imagem da peça L e da torre em oclusão parcial, dos agrupamentos primário e intermediário, respectivamente.

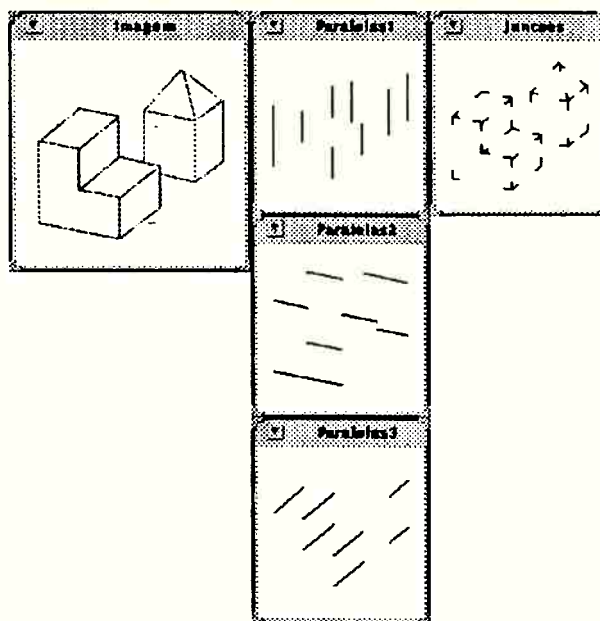


Fig.5.38 Agrupamentos primários para a imagem das peças L e torre, em oclusão.

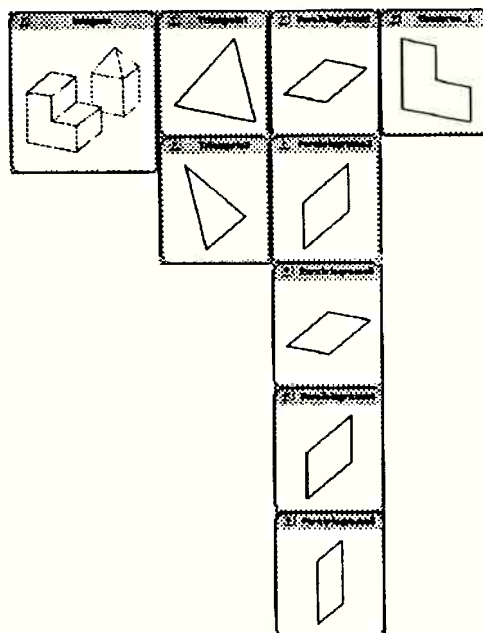


Fig.5.39 Agrupamentos intermediários para a imagem das peças L e torre, em oclusão.

Na figura 5.40 tem-se o resultado do reconhecimento dessa imagem, para a hierarquia completa. A figura 5.41 apresenta as aparências associadas aos nós correspondidos.

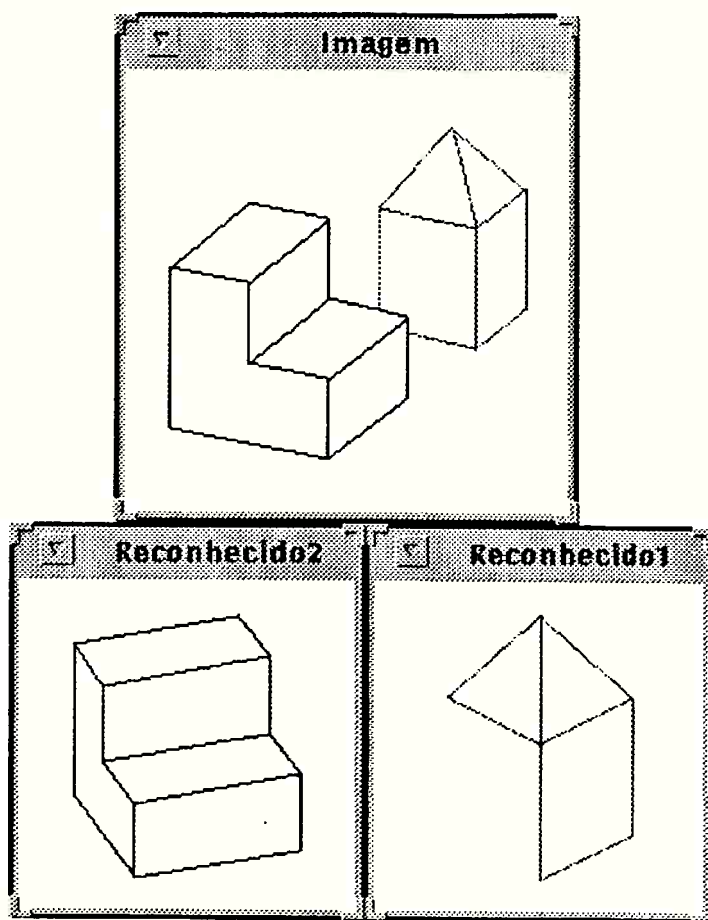


Fig.5.40 Imagem onde a peça L e a torre se encontram em oclusão parcial e os respectivos nós correspondidos da hierarquia.

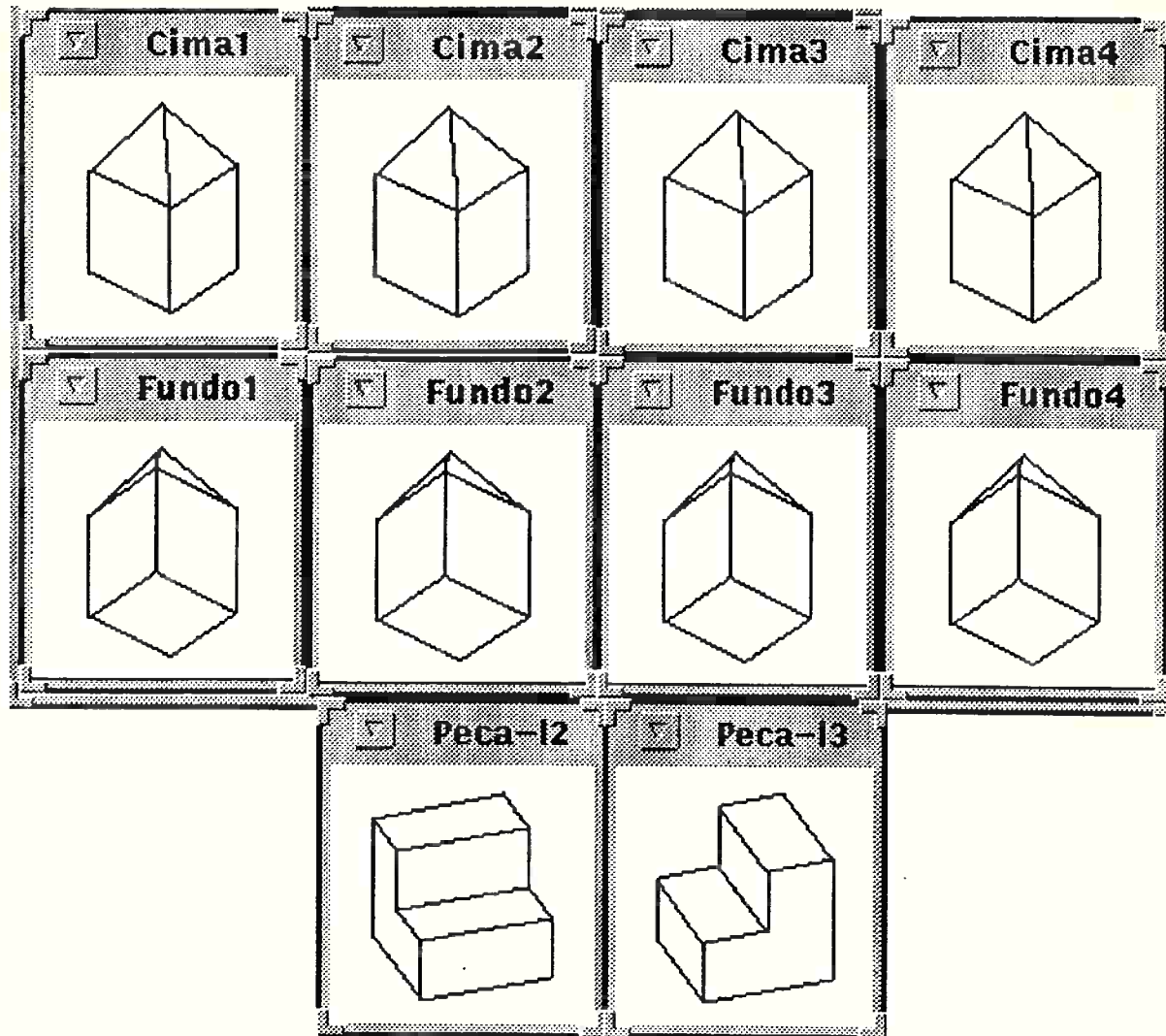


Fig.5.41 Aparências associadas aos nós correspondidos da figura 5.40.

As figuras 5.42 e 5.43 mostram os resultados, para a imagem da peça L e do cubo em oclusão parcial, dos agrupamentos primário e intermediário, respectivamente.

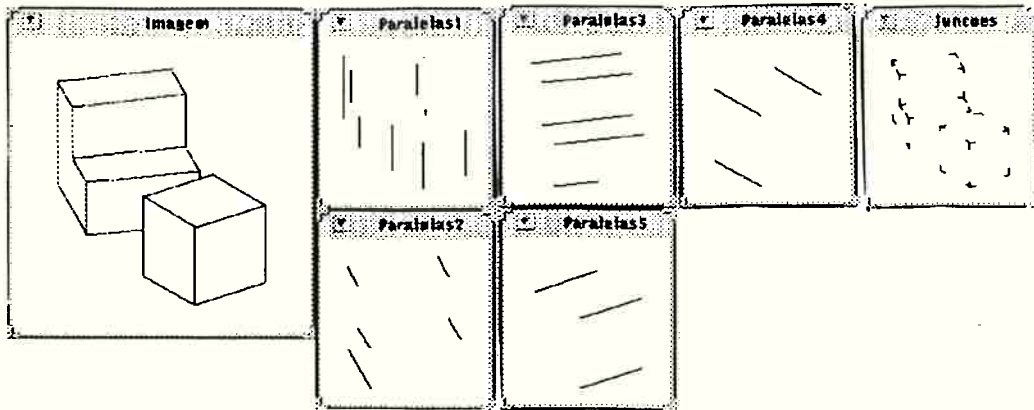


Fig.5.42 Agrupamentos primários para a imagem das peças L e cubo, em oclusão.

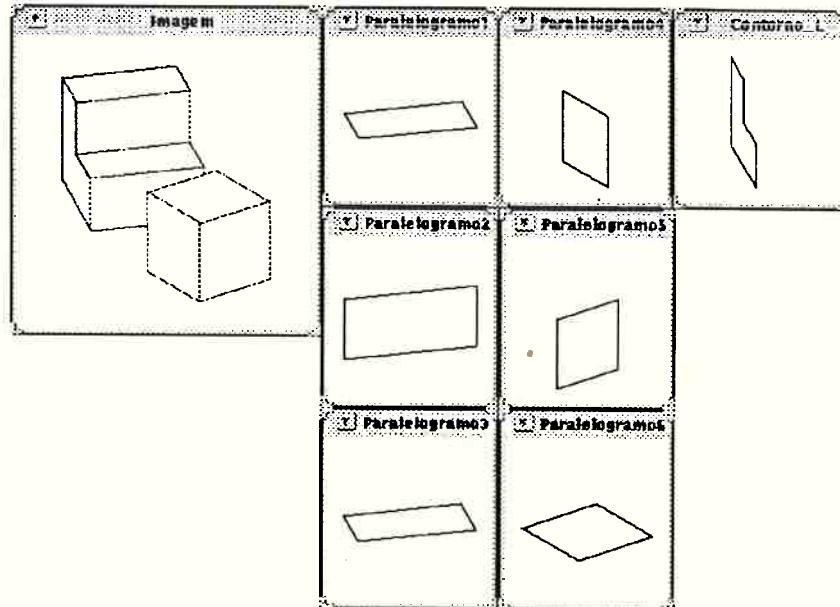


Fig.5.43 Agrupamentos intermediários para a imagem das peças L e cubo, em oclusão.

Na figura 5.44 tem-se o resultado do reconhecimento dessa imagem, para a hierarquia completa, com respectivas aparências associadas aos nós correspondidos.

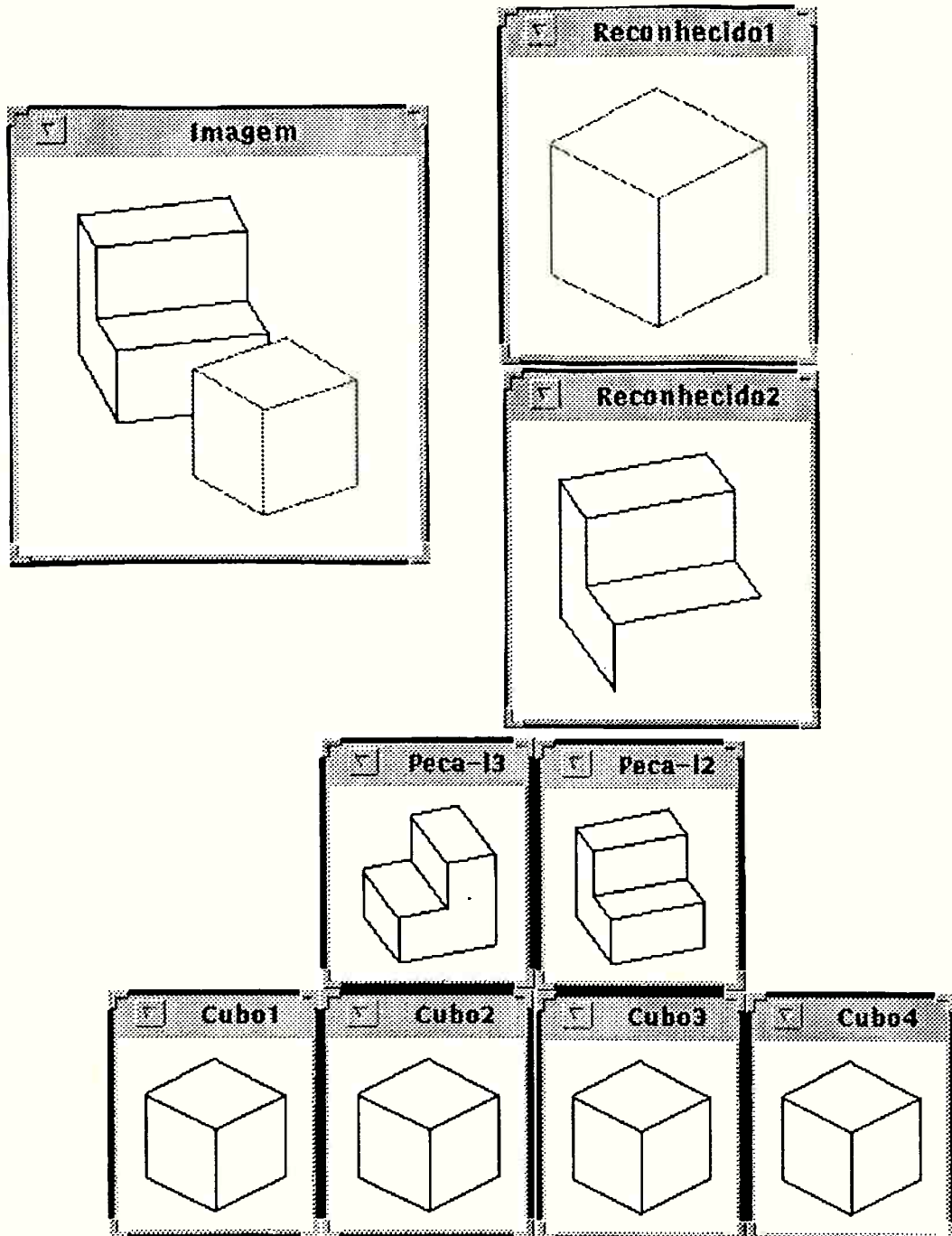


Fig.5.44 Imagem onde a peça L e o cubo se encontram em oclusão parcial e os respectivos nós correspondidos da hierarquia.

As próximas figuras ilustram um caso onde ocorre alinhamento acidental de segmentos, fazendo com que RECTRI detecte um triângulo e tente associá-lo às aparências que contenham triângulo, gerando uma falsa hipótese. Esta falsa hipótese pode ser facilmente eliminada num processo de verificação.

Assim, a figura 5.45 apresenta os agrupamentos intermediários definidos para a imagem da peça U e a figura 5.46 mostra os nós da hierarquia correspondidos à imagem.

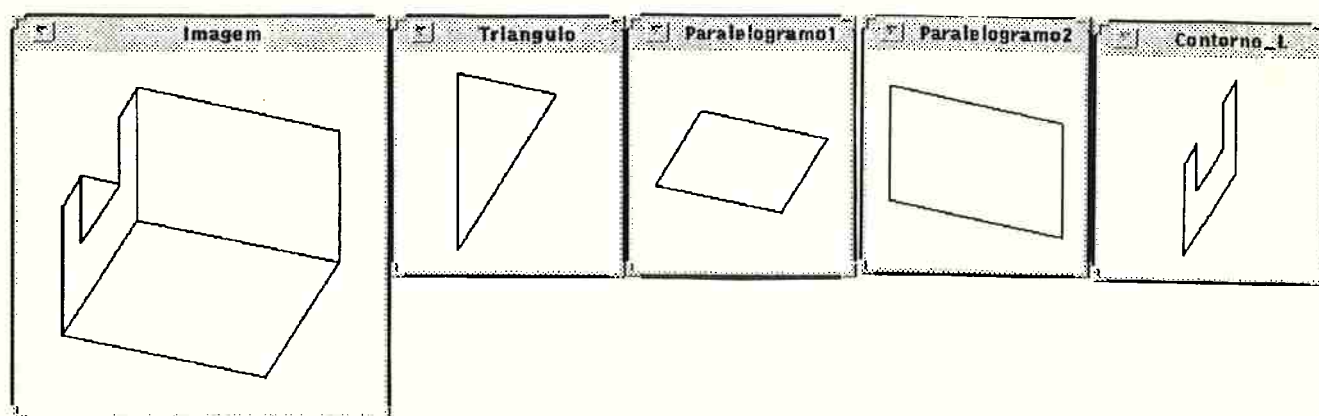


Fig.5.45 Agrupamentos intermediários detectados para a peça U, com alinhamento acidental de segmentos.

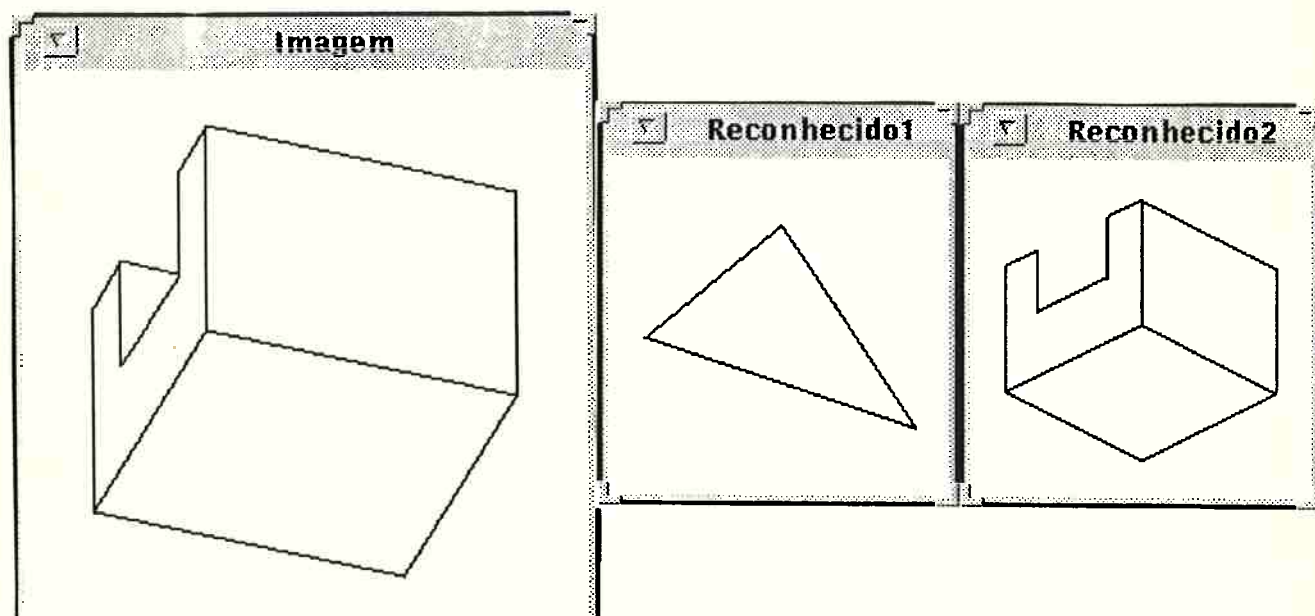


Fig.5.46 Resultado do reconhecimento.

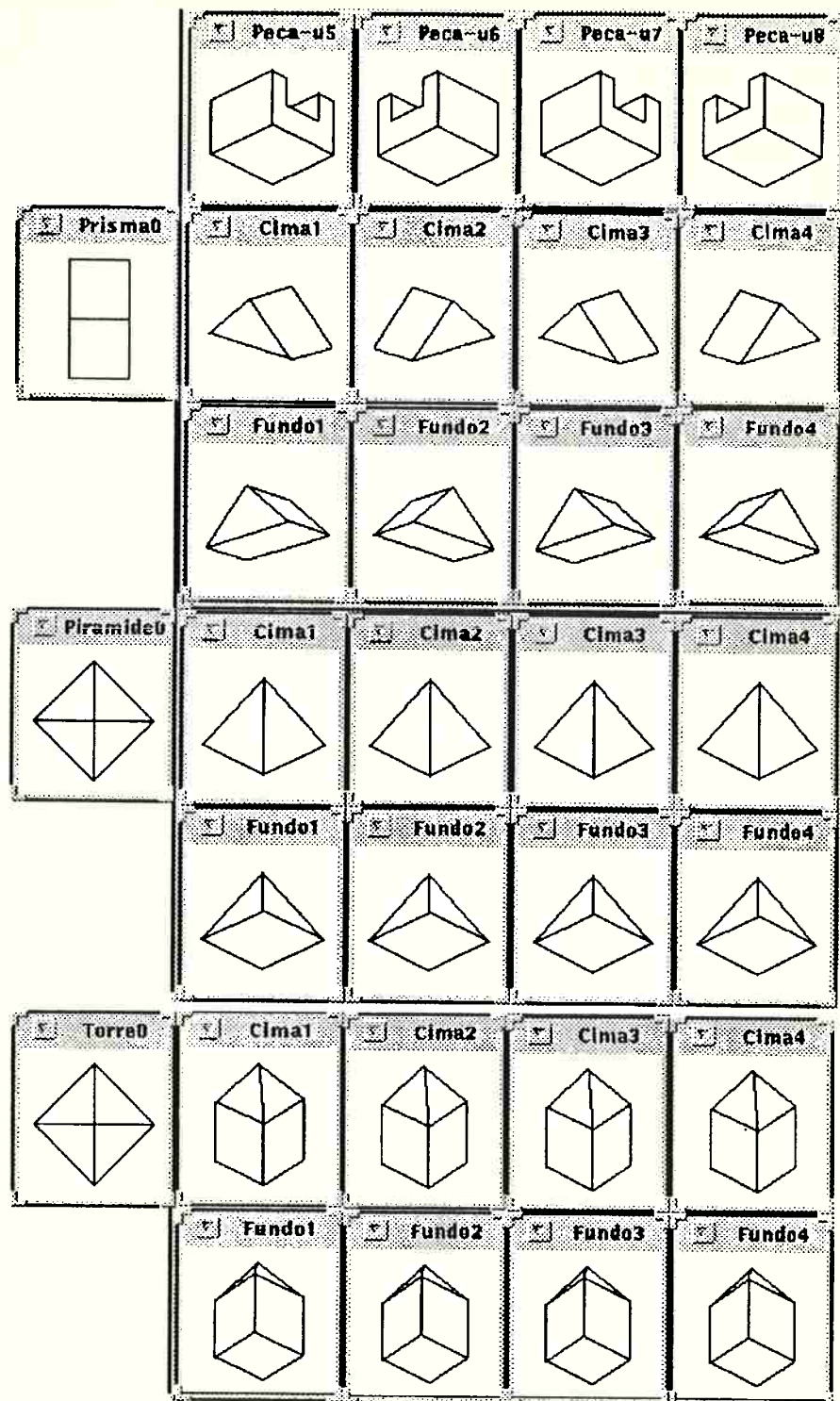


Fig.5.47 Aparências associadas ao reconhecimento apresentado na figura 5.46.

5.8.4. Análise dos Resultados

Os resultados foram bastante motivadores, considerando a dificuldade inerente ao problema abordado. Deve-se ressaltar que a implementação foi bastante trabalhosa, muito mais do que o esperado.

Durante o desenvolvimento do sistema, cada módulo foi testado individualmente, avaliando se as propriedades de interfaceamento estavam conforme o especificado e também se a especificação estava correta até o momento. No entanto, acredita-se que um número maior de testes deve ser executado para o sistema como um todo, integrado.

O propósito da implementação foi o de concluir que o modelo sugerido é um modelo válido e deve ser considerado como uma linha de pesquisa a ser seguida e aperfeiçoada.

Uma dificuldade bastante grande é o ajuste satisfatório de todos os parâmetros, desde aqueles referentes à segmentação de imagens, extração de primitivas (externos à implementação efetuada), até aqueles do RECTRI, indicando desvios permitidos para as diversas medidas, para determinação das primitivas e dos agrupamentos da imagem.

Esses parâmetros são muito sensíveis à forma dos objetos que fazem parte da cena, às condições de iluminação, aos sensores e sua calibração, etc. Uma intensa interação com o usuário é necessária nesta fase de ajuste de parâmetros e o usuário necessita estabelecer determinados critérios para este ajuste, em função do resultado desejado.

Outro problema foi o de determinar maneiras de visualizar eficientemente os resultados, uma vez que a fase de verificação não seria desenvolvida no âmbito desta tese e, portanto, os parâmetros finais de localização e identificação não seriam acessíveis.

Os resultados refletem a correspondência alcançada entre segmentos de uma imagem e segmentos de aparência(s), sendo que o sistema também tem conhecimento do objeto que comporta a(s) referida(s) aparência(s) e de como ele deve ter sido visualizado (região de visualização) para corresponder à tal aparência.

6. CONCLUSÕES E TRABALHOS FUTUROS

6.1. Conclusões e Contribuições

Neste trabalho foi proposto um modelo computacional para o reconhecimento de objetos rígidos tridimensionais, a partir de uma única imagem bidimensional da intensidade luminosa refletida por uma cena. Os resultados obtidos mostraram-se promissores, motivando novas pesquisas e aperfeiçoamentos nesta área.

O argumento central sobre o qual se baseia esta tese consiste em afirmar que existe informação suficiente para um reconhecimento tridimensional, tanto para visão humana quanto computacional, que pode, em muitos casos, ser extraída de representações bidimensionais, sem reconstrução preliminar da profundidade, apesar dessa reconstrução ser um processo muito importante em algumas tarefas visuais.

Este modelo computacional está baseado na seleção e organização de características úteis e discriminatórias dos objetos, as quais são armazenadas numa estrutura chamada Hierarquia de Estratégias.

A Hierarquia de Estratégias reflete uma nova abordagem dada para este problema, onde se encontram representadas estratégias de busca por correspondência entre atributos extraídos das imagens e aqueles selecionados dos objetos modelos.

Essa hierarquia tem a importante função de, além de determinar as estratégias para a busca, reduzir o reconhecimento

a um processo bidimensional. Isso é conseguido através da utilização do conhecimento sobre a aparência visual dos objetos.

Desse modo, pode-se conseguir uma rápida indexação do objeto correto e seu ponto, ou região, de observação. Essa indexação é um componente crítico do sistema, uma vez que seu desempenho determina a quantidade de busca que precisa ser efetuada durante o reconhecimento.

Acredita-se que o reconhecimento não é nem um processo guiado totalmente pelo conhecimento, nem guiado somente pelos dados. Por isso, esta proposta consiste em um sistema de visão que integra ambos os processos: o primeiro, pelo uso da Hierarquia de Estratégias e o segundo, pela utilização de atributos intermediários, também chamados agrupamentos, para acessar esta hierarquia.

Esta proposta foi implementada no sistema de reconhecimento RECTRI, constituído por duas fases: pré-processamento e reconhecimento. Esta divisão é uma característica importante, uma vez que diminui significativamente a complexidade do algoritmo de reconhecimento e seu tempo de processamento.

Na fase de pré-processamento obtém-se, a partir de descrições dos objetos tridimensionais feitas pelo usuário, modelos tridimensionais dos objetos, que podem ser usados num futuro processo de verificação, e a Hierarquia de Estratégias.

A construção da Hierarquia de Estratégias foi feita de modo totalmente automático, incremental e iterativo. Cada nova aparência é introduzida na hierarquia, sem afetar a hierarquia construída até então. O processo parte das folhas da hierarquia até chegar à raiz, num sentido inverso ao utilizado durante o reconhecimento. Cada folha está associada a uma aparência única

de um objeto ou a aparências cujas diferenças em suas projeções não são representáveis e/ou perceptíveis ao sistema.

É importante ressaltar que a condição de rigidez dos objetos está integrada na hierarquia, sem ter o cálculo explícito da transformação rígida aplicada ao objeto, a cada passo da correspondência.

A fase de reconhecimento consiste na extração dos atributos primitivos da imagem, agrupamento perceptual desses atributos, correspondência bidimensional entre esses agrupamentos e aqueles extraídos dos modelos e um processo de verificação e localização, este último não implementado na versão atual do RECTRI.

O extrator de atributos aplica processamento de segmentação e detecção de bordas numa imagem, obtendo atributos primitivos, os quais consistem em segmentos de retas representando as bordas extraídas. Esse processo não requer segmentação perfeita da imagem.

O uso de atributos intermediários, também chamados agrupamentos, possibilita que tanto objetos genericamente descritos como modelos exatos possam ser utilizados. A idéia consiste em um sistema de visão no qual estruturas genéricas são reconhecidas inicialmente, e depois os detalhes dos objetos são examinados num processo de verificação ou ainda de correspondência, porém num nível de abstração mais baixo.

O processo de agrupamento consiste em criar uma estrutura relacional que evidencie as relações (colinearidade, paralelismo, partes de um paralelograma, etc) entre os atributos de uma imagem ou uma aparência. Esses agrupamentos oferecem diversas vantagens; entre elas, cita-se:

- contornos lineares e curvos, junções e simetrias locais são bons candidatos para descrever formas tridimensionais;
- grupos de atributos de uma imagem são intrinsecamente menos ambíguos que atributos isolados;
- a complexidade associada à interpretação da imagem é consideravelmente reduzida;
- o processo de agrupamento elimina atributos isolados indesejados.

Após o agrupamento, o processo de correspondência é efetuado, onde são geradas hipóteses de correspondência entre conjuntos de atributos da imagem e dos modelos. Este processo de busca por correspondência entre atributos é guiado pela Hierarquia de Estratégias, criada no pré-processamento do sistema.

Uma vez que se obtém uma estimativa inicial para um objeto e seu ponto de observação, um processo de verificação deve ser efetuado para resolver a correspondência espacial.

6.2. Perspectivas Futuras

Os esforços atuais consistem em um desenvolvimento mais profundo do sistema RECTRI. Há muito espaço ainda para pesquisas voltadas para segmentação e vetorização da imagem, com ênfase em algoritmos mais robustos e eficientes. No entanto, deve-se ressaltar que resultados animadores têm sido adquiridos por algumas abordagens, para determinado tipo de imagem, como, por exemplo, a abordagem por morfologia matemática.

Acredita-se também que outras heurísticas devam ser testadas para a detecção de relações primárias e agrupamento de atributos.

Também procura-se direcionar esforços para alguns aperfeiçoamentos, tais como organização global da hierarquia, heurísticas de terminação da busca, controle mais eficiente da busca, adoção do modelo perspectivo para o processo de imageamento já na fase de construção da hierarquia, introdução de mecanismos de entrada de restrições durante o processo de determinação das aparências, visando diminuir o número de aparências, etc.

Na atual versão do sistema RECTRI, alguns critérios de organização da hierarquia foram aplicados; porém, para que se alcance uma organização melhor, que agilize ao máximo a busca por correspondência durante o reconhecimento, deve-se aplicar uma organização global ao final da construção da hierarquia ou então o processo de seleção e organização inicial dos atributos deve ter conhecimento global.

Pode-se pensar na aplicação destes critérios globais de organização como sendo filtros que se aplicam durante todo o processo de organização do conhecimento e de correspondência entre os dados e o conhecimento armazenado. É importante notar que a ordem de aplicação destes filtros influencia na função de controle final do sistema de reconhecimento. Por isso, deve-se não só analisar o efeito da aplicação de cada filtro isoladamente, mas também o resultante da interação entre eles.

É interessante, também, que seja feita a integração de funções de custo na busca na hierarquia, de tal modo que, caso o custo da evolução de um determinado nó (geração da hipótese)

seja maior do que o custo da aplicação do processo de verificação das hipóteses levantadas até então, prossegue-se a busca; caso contrário, aborta-se a busca, efetuando a verificação das hipóteses. Para este procedimento, torna-se imprescindível a exata definição das funções de custo para os processos de geração das hipóteses e de verificação destas hipóteses, o que não é uma tarefa simples.

Outra proposta interessante seria a de introduzir uma heurística para atuar dinamicamente no controle da busca. Uma maneira de se efetuar este controle seria através da indexação inicial não por uma, mas por diversas propriedades disponíveis para acesso à hierarquia. Na versão atual do sistema RECTRI, os agrupamentos determinados na imagem são ordenados segundo os critérios já discutidos e somente o primeiro agrupamento resultante desta ordenação é que acessa a hierarquia. A heurística aqui proposta consiste em utilizar mais de um agrupamento para isso, determinando as aparências mais prováveis e direcionar a busca para os ramos das aparências mais prováveis. A busca pode ser efetuada por processamento paralelo, seguindo pelos ramos indicados, ou mesmo seqüencial, porém acrescido da informação das aparências mais prováveis.

Outra importante área de investigação consiste no aprendizado a partir de cenas corretamente interpretadas, uma vez que a ordenação da hierarquia pode ser continuamente ajustada, procurando uma classificação mais eficiente para o tipo de cena interpretada, com o ganho de experiência do processo visual.

Esta proposta ofereceu uma solução para o problema de reconhecimento 3D baseado em modelos, a partir de informação 2D.

Acredita-se que outros tipos de representação devam ser explorados, tanto para as aplicações às quais este sistema está voltado quanto para outras mais gerais, uma vez que o uso de sofisticadas estruturas de dados é um importante componente para visão computacional.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGGARWAL, J.K.; CHIEN, C.H. 3D structures from 2D images. In SANZ, J.L.C., ed **Advances in machine vision**. New York, Springer, 1989. p.64-121.
- ALOIMONOS, J.Y. Unification and integration of visual modules: an extension of the Marr paradigm. In: **DARPA IMAGE UNDERSTANDING WORKSHOP**, Palo Alto, 1989. **Proceedings**. Palo Alto, DARPA, 1989. p.507-51.
- AOYAMA, H.; KAWAGOE, M. A piecewise linear approximation method preserving visual feature points of original figures. **CVGIP: Graphical Models and Image Processing**, v.53, n.5, p.435-46, Sept. 1991.
- ARKIN, E.M.; CHEW, L.P.; HUTTENLOCHER, D.P.; KEDEM, K.; MITCHEL, J.S.B. An efficiently computable metric for comparing polygonal shapes. **IEEE PAMI: Transactions on Pattern Analysis and Machine Intelligence**, v.13, n.3, p.209-16, Mar. 1991.
- AYACHE, N. **Artificial vision for mobile robots: stereo vision and multisensory perception**. Trad. de Peter T. Sander. Cambridge, The MIT Press, 1991.
- BAESSMANN, H.; BESSLICH, P.W. On the recognition-by-components approach applied to computer vision. **Proceedings of the SPIE**. The International Society for Optical Engineering, v.1192, p.52-62, 1989.
- _____. **Bildverarbeitung**. Berlin, Springer, 1991.
- BALLARD, D.H.; BROWN, C.M. **Computer vision**. Englewood Cliffs, Prentice-Hall, 1982.
- BARRERA, J.; BANON, J.F.; LOTUFO, R.A. A mathematical morphology toolbox for the KHOROS system. **IMAGE ALGEBRA AND MORPHOLOGICAL IMAGE PROCESSING**, 5º, San Diego, July 1994. **Proceedings of the SPIE**. 1994.
- BERGEVIN, R.; LEVINE, M.D. Part-based description and recognition of objects in line drawings. **Proceedings of the SPIE**. The International Society for Optical Engineering, v.1192, p.63-74, 1989.
- _____. Part decomposition of objects from single view line drawings. **CVGIP: Image Understanding**, v.55, n.1, p.73-83, Jan. 1992.
- _____. Extraction of line drawing features for object recognition. **Pattern Recognition**, v.25, n.3, p.319-34, 1992.
- _____. Generic object recognition: building and matching coarse descriptions from line drawings. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.15, n.1, p.19-36, Jan. 1993.
- BESL, P.J.; JAIN, R.C. Three-dimensional object recognition. **ACM Computing Surveys**, v.17, n.1, p.75-145, 1985.

- BEUCHEUR, S.; VINCENT, L. Introduction aux outils morphologiques de segmentation. *Traitement d'images en microscopie à balayage et en microanalyse par sonde électronique* ANRT ed., p F1-F43, Paris, 1990.
- BHANU, B.; HO, C.C. CAD-based 3D object representation for robot vision. *IEEE Computer*, v.20, n.8, p.19-35, Aug. 1987.
- BIEDERMAN, I. Human image understanding: recent research and a theory. *Computer Vision, Graphics, and Image Processing*, v.32, p.29-73, 1985.
- BOLLES, R.C.; HORAUD, P. 3DPO: a three-dimensional part orientation system. *The International Journal of Robotics Research*, v.5, n.3, p.3-26, 1986.
- BROOKS, R.A. Model-based 3D interpretations of 2D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.5, p.140-150, 1983.
- BURNS, J.B.; KITCHEN, L.J. Rapid object recognition from a large model base using prediction hierarchies. In: *DARPA IMAGE UNDERSTANDING WORKSHOP*, Cambridge, 1988. *Proceedings*, DARPA, 1988, p.711-9.
- CANNY, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.8, n.6, p.679-98, Nov. 1986.
- CHEN, C.H.; KAK, A.C. A robot vision system for recognizing 3D objects in low-order polynomial time. *IEEE Transactions on Systems, Man, and Cybernetics*, v.19, n.6, p.1535-63, Nov./Dec. 1989.
- CHIN, R.T.; DYER, C.R. Model-based recognition in robot vision. *ACM Computing Surveys*, v.18, n.1, p.67-108, 1986.
- CLEMENS, D.T.; JACOBS, D.W. Space and time bounds on indexing 3D models from 2D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.13, n.10, p.1007-17, Oct. 1991.
- DAVIES, E.R. *Machine vision: theory, algorithms, practicalities*. London, Academic Press, 1990.
- DICKINSON, S.J.; PENTLAND, A.P.; ROSENFELD, A. 3D shape recovery using distributed aspect matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.14, n.2, p.174-98, Feb. 1992.
- DOLAN, J.; WEISS, R. Perceptual grouping of curved lines. In: *DARPA IMAGE UNDERSTANDING WORKSHOP*, Palo Alto, 1989. *Proceedings*, DARPA, 1989, p.1135-45.
- DRAPER, B.A.; RISEMAN, E.M. Learning 3D object recognition strategies. In: *IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION*, 3^o, Osaka, 1990. *Proceedings*, Los Alamitos, IEEE Computer Society Press, 1990, p.320-4.
- DU, L.; SULLIVAN, G.D.; BAKER, K.D. 3D grouping by viewpoint consistency ascent. In: *BRITISH MACHINE VISION CONFERENCE*, Glasgow, 1991. *Proceedings*, London, Springer, 1991, p.45-53.
- EDELMAN, S. On learning to recognize 3D objects from examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.15, n.8, p.833-7, Aug. 1993.

- EGGERT, D.; BOWYER, K. Computing the perspective projection aspect graph of solids. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.15, n.2, p.109-28, Feb. 1993.
- FAN, T.J.; MEDIONI, G.; NEVATIA, R. 3D object recognition using surface descriptions. In: **DARPA IMAGE UNDERSTANDING WORKSHOP**, Cambridge, 1988. **Proceedings**. DARPA, 1988. p.383-97.
- FAUGERAS, O.D.; HEBERT, M. The representation, recognition, and locating of 3D objects. **The International Journal of Robotics Research**, v.5, n.3, p.27-52, 1986.
- FLYNN, P.J. **CAD-based computer vision: modeling and recognition strategies**. Michigan, 1990. PhD (Dissertation) - Michigan State University, Department of Computer Science.
- FLYNN, P.J.; JAIN, A.K. CAD-based computer vision: from CAD models to relational graphs. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.13, n.2, p.114-32, Feb. 1991.
- _____. BONSAI: 3D object recognition using constrained search. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.13, n.10, p.1066-75, Oct. 1991.
- FORSYTH, D. A novel approach to color constancy. In: **IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION**, 2^o, Tampa, 1988. **Proceedings**. Los Alamitos, IEEE Computer Society Press, 1988. p.9-18.
- FREEMAN, H. Computer Processing of Line-Drawing Images. **Computing Surveys**, v.6, n.1, p.57-97, Mar. 1974.
- _____, ed. **Machine vision for three-dimensional scenes**. San Diego, Academic Press, 1990.
- GEIGER, D.; YUILLE, A. A common framework for image segmentation. **International Journal of Computer Vision**, v.6, n.3, p.227-43, 1991.
- GEIGER, D.; GIROSI, F. Parallel and deterministic algorithms for MRFs: surface reconstruction and integration. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.13, p.401-12, 1991.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.6, p.721-41, 1984.
- GIGUS, Z.; MALIK, J. Computing the aspect graph for line drawings of polyhedral objects. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.12, n.2, p.113-22, Feb. 1990.
- GONZALEZ, R.C.; WINTZ, P. **Digital image processing**. Reading, Addison-Wesley, 1987.
- GOODRICH, M.T. A polygonal approach to hidden-line and hidden-surface elimination. **CVGIP: Graphical Models and Image Processing**, v.54, n.1, p.1-12, Jan. 1992.
- GRIMSON, W.E.L.; LOZANO-PÉREZ, T. Recognition and localization of overlapping parts from sparse data in two and three dimensions. In: **IEEE INTERNATIONAL**

- CONFERENCE ON ROBOTICS AND AUTOMATION, St. Louis, 1985
Proceedings. Washington, IEEE Computer Society Press, 1985, p.61-6
- GRIMSON, W.E.L. **Object Recognition by Computer: the role of geometric constraints**. Cambridge, The MIT Press, 1990
- GUZMAN, A. Decomposition of a visual scene into three dimensional bodies. AFIPS Conference. **Proceedings**. Montvale, v.33, n.2, p.291-304
- HANSEN, C.; HENDERSON, T.C. CADG-based computer vision. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.11, n.11, p.1181-93, Nov. 1989.
- HELLER, A.J.; STENSTROM, J.R. Verification of recognition and alignment hypotheses by means edge verification statistics. In DARPA IMAGE UNDERSTANDING WORKSHOP, Palo Alto, 1989. **Proceedings**. Palo Alto, DARPA, 1989. p.957-66.
- HELMKE, H.; JANSSEN, R.; SAUR, G. Automatische Erzeugung dreidimensionaler Kantenmodelle aus mehreren zweidimensionalen Objektansichte. **Informatik-Fachberichte 254**, p.617-24, 1990.
- HORAUD, R. New methods for matching 3D objects with single perspective views. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.9, n.3, p.401-12, May 1987.
- HORAUD, R.; BRADY, M. On the geometric interpretation of image contours. In: KAPUR, D.; MUNDY, J.L., ed. **Geometric reasoning**. Cambridge, MIT Press, 1989. p.333-53.
- HORAUD, R.; VEILLON, F.; SKORDAS, T. Finding geometric and relational structures in an image. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 1°, Antibes, 1990. **Proceedings**. Berlin, Springer, 1990. p.374-84.
- HORN, B.K.P. **Robot vision**. Cambridge, MIT Press, 1986.
- HUTTENLOCHER, D.P.; ULLMAN, S. Recognizing solid objects by alignment. In: DARPA IMAGE UNDERSTANDING WORKSHOP, Cambridge, 1988. **Proceedings**. DARPA, 1988. p.1114-22.
- HUTTENLOCHER, D.P.; CASS, T.A. Measuring the quality of hypotheses in model-based recognition. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV'92), Santa Margherita Ligure, 1992. **Proceedings**. Berlin, Springer, 1992. p.773-7
- IKEUCHI, K.; KANADE, T. Modeling sensors and applying sensor model to automatic generation of object recognition program. In: DARPA IMAGE UNDERSTANDING WORKSHOP, Cambridge, 1988. **Proceedings**. DARPA, 1988. p.697-710.
- _____. Automatic generation of object recognition programs. **Proceedings of the IEEE**, v.76, n.8, p.1016-35, Aug. 1988
- JACOBS, D.W. **The use of grouping in visual object recognition**. Massachusetts Institut of Technology, 1988. MIT Artificial Intelligence Laboratory, USA.
- KANADE, T. Recovery of the three-dimensional shape of an object from a single view. **Artificial Intelligence**, v.17, p.409-60, 1981
- KANATANI, K. **Group-theoretical methods in image understanding**. Berlin, Springer, 1990

- KORN, A.F. Toward a symbolic representation of intensity changes in images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.10, n.5, p.610-25, Sep 1988.
- KUNO, Y.; OKAMOTO, Y.; OKADA, S. Robot vision using a feature search strategy generated from a 3D object model. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.13, n.10, p.1085-97, Oct. 1991.
- LAMDAN, Y.; SCHWARTZ, J.T. & WOLFSON, H.J. On recognition of 3D objects from 2D images. In **IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 5^o**, Philadelphia, 1988. **Proceedings**. Washington, IEEE Computer Society Press, 1988, p.1407-13.
- LINDENBAUM, M.; BRUCKSTEIN, A. On recursive $O(N)$ partitioning of digitized curve into digital straight segments. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.15, n.9, p.949-53, Sept. 1993.
- LIU, Y.; HUANG, T.S.; FAUGERAS, O.D. Determination of camera location from 2D to 3D line and point correspondences. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.12, n.1, p.28-37, 1990.
- LOWE, D.G. **Perceptual organization and visual recognition**. Hingham, Kluwer Academic, 1985.
- _____. Three-dimensional object recognition from single two-dimensional images. **Artificial Intelligence**, v.31, p.355-95, 1987.
- _____. Fitting parameterized three-dimensional models to images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.13, n.5, p.441-50, May 1991.
- MAJUMDAR, J.; LEVI, P.; REMBOLD, U. 3D model-based robot vision by matching scene description with the object model from a CAD modeller. **Robotics and Autonomous Systems**, v.5, p.69-83, 1989.
- MALIK, J.; PERONA, P. A computational model of texture segmentation. In: **IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, San Diego, 1989. CVPR: proceedings**. Los Alamitos, IEEE Computer Society Press, 1989, p.326-332.
- MARR, D. **Vision: a computational investigation into the human representation and processing of visual information**. San Francisco, W.H. Freeman, 1982.
- MARTELLI, A. An application of heuristic search methods to edge and contour detection. **Communications of the ACM**, v.19, n.2, p.73-83, Feb. 1976.
- McANDREW, P.; WALLACE, A.M. Rapid invocation and matching of 3D models to 2D images using curvilinear data. In **IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING AND ITS APPLICATIONS, 3^o**, Warwick, 1989. **Proceedings**. London, IEE, 1989, p.83-7.
- McLEAN, S.; HORAN, P.; CAELLI, T. A data-driven indexing mechanism for the recognition of polyedral objects. **Proceedings of the SPIE**. The International Society for Optical Engineering, v.1609, p.121-33, 1992.
- MOHAN, R.; NEVATIA, R. Perceptual organization for segmentation and description. In: **DARPA IMAGE UNDERSTANDING WORKSHOP, Palo Alto, 1989. Proceedings**. DARPA, 1989 p.415-24.

- MORIMOTO, C.H. **Sistema de visão robótica para reconhecimento de objetos poliédricos a partir de uma única imagem.** São Paulo, 1991. p.156. Dissertação (Mestrado) - Escola Politécnica, Universidade de São Paulo
- MOSES, Y.; ULLMAN, S. Limitations of non model-based recognition schemes. In EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV'92), Santa Margherita Ligure, 1992. **Proceedings.** Berlin, Springer, 1992, p.820-8
- MUMFORD, D.; SHAH, J. Boundary detection by minimizing functionals. In IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 3º, San Francisco, 1985. **CVPR: proceedings.** Silver Spring, IEEE Computer Society Press, 1985, p.22-6.
- OSHIMA, M.; SHIRAI, Y. Object recognition using three-dimensional information. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (IJCAI-81), 7º, Vancouver, 1981. **Proceedings.** s.L., IJCAI, 1981. v.2, p.601-6.
- OVERINGTON, I. **Computer vision: a unified, biologically-inspired approach.** Amsterdam, Elsevier, 1992.
- PAMPAGNIN, L.H. **Reconnaissance d'objets tridimensionnels en perception monoculaire et multisensorielle : application a la robotique spatiale.** Toulouse, 1990. Thèse (Doctorat). Laboratoire d'Automatique et d'Analyse des Systèmes, Université Paul Sabatier.
- REQUICHA, A.A.G. Representations for rigid solids: theory, methods, and systems. **ACM Computing Surveys**, v.12, n.4, p.437-64, Dec. 1980
- RILLO, A.H.R.C. **Sistema de visão binária para peças parcialmente oclusas.** São Bernardo do Campo, 1989. p.137. Dissertação (Mestrado) - Faculdade de Engenharia Industrial, Fundação de Ciências Aplicadas (FEI - FCA).
- _____. Um sistema de visão computacional para reconhecimento de peças industriais. In CONGRESSO NACIONAL DE AUTOMAÇÃO INDUSTRIAL, 4º, São Paulo, 1990. **CONAI: anais.** São Paulo, SUCESU, 1990. p.220-9.
- _____. Um sistema de visão computacional para reconhecimento de peças isoladas e parcialmente oclusas. In: SIMPOSIO BRASILEIRO DE COMPUTAÇÃO GRAFICA E PROCESSAMENTO DE IMAGENS, 3º, Gramado, 1990. **SIBGRAPI: anais.** Porto Alegre, UFRGS, 1990. p.236-45.
- _____. Sistema de reconhecimento de peças industriais baseado em agrupamento de características. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 9º, Vitória, 1992. **CBA: anais.** Vitória, SBA, 1992. p.234-9.
- _____. Reconhecimento tridimensional a partir de uma única imagem bidimensional de intensidade luminosa. In: CONGRESSO NACIONAL DE AUTOMAÇÃO INDUSTRIAL, 5º, São Paulo, 1992. **CONAI: anais.** São Paulo, SUCESU, 1992. p.263-71.
- _____. Grouping-based recognition system. **Proceedings of the SPIE.** The International Society for Optical Engineering, v. 1619, p.274-82, 1992
- _____. Model-based 3D object recognition from 2D. In PORTUGUESE CONFERENCE ON PATTERN RECOGNITION, 4º, Coimbra, 1992. **RECPAD: proceedings.** Coimbra, 1992.

- _____. 3D object recognition using a decision hierarchy. **Proceedings of the SPIE.** The International Society for Optical Engineering, v 1771, 1993
- RILLO, M.; RILLO, A.H.R.C.; COSTA, L.A.R. The LSI assembly cell. In: IFAC/IFIP/IFORS/IMACS/ISPE SYMPOSIUM ON INFORMATION CONTROL PROBLEMS IN MANUFACTURING TECHNOLOGY, 7^o, Toronto, 1992. **Proceedings.** IFAC, 1992. p. 361-5.
- _____. A célula de montagem do LSI. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 9^o, Vitória, 1992. **CBA: anais.** Vitória, SBA, 1992. p.952-7.
- RILLO, M.; RILLO, A.H.R.C.; CIRELLI, S.; ZILINSKAS, W. Controlador programável por grafos de comando etapa-transição (GRAF CET). In: CONGRESSO NACIONAL DE AUTOMAÇÃO INDUSTRIAL, 1^o, São Paulo, 1983. **CONAI: anais.** São Paulo, SUCESU, 1983. p.246-52
- ROBERTS, L.G. Machine perception of three-dimensional solids. In: TIPPET, J.T. et al., eds. **Optical and electro optical information processing.** Cambridge, MIT Press, 1965. p.159-97.
- SAROLDI, A. Successive groupings: adding knowledge to improve segmentation. In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, 8^o, Munich, 1988. **ECAI: proceedings.** London, Pitman, 1988. p.669-71.
- SHIRAI, Y. **Three-dimensional computer vision.** Berlin, Springer, 1987.
- SUGIHARA, K. **Machine interpretation of line drawings.** Cambridge, MIT Press, 1986.
- SWAIN, M.J. Object recognition from large database using a decision tree. In: DARPA IMAGE UNDERSTANDING WORKSHOP, Cambridge, 1988. **Proceedings.** DARPA, 1988. p.690-6.
- _____. **Color indexing.** Rochester, 1990. Thesis (PhD). University of Rochester.
- THOMPSON, D.W.; MUNDY, J.L. Three dimensional model matching from an unconstrained viewpoint. In: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 1987. **Proceedings.** Washington, IEEE Computer Society Press, 1987. p.208-20.
- TOMMASELLI, A.M.G. **Um método recursivo aplicado ao problema de localização em visão de máquina.** Campinas, 1993. p.142. Tese (Doutorado). Departamento de Engenharia da Computação e Automação Industrial, Faculdade de Engenharia Elétrica, Universidade Estadual de Campinas
- TOMITA, F.; KANADE, T. A 3D vision system: generating and matching shape descriptions in range images. In: IEEE CONFERENCE ON ARTIFICIAL INTELLIGENCE APPLICATIONS, 1^o, Denver, 1984. **CAIA: conference record.** New York, IEEE, 1984. p.186-91.
- ULLMAN, S.; BASRI, R. **Recognition by linear combinations of models.** Cambridge, MIT, 1989. (MIT A.I. Memo n.1152)
- VAYDA, A.J.; KAK, A.C. A robot vision system for recognition of generic shaped objects. **CVGIP: Image Understanding,** v.54, n 1, p 1-46. July 1991

- WALTZ, D. Generating semantic descriptions from drawings of scenes with shadows. In WINSTON, P.H., ed. **The psychology of computer vision**. New York, McGraw-Hill, 1975. p.19-91
- WOLFE, W.J.; MATHIS, D.; SKLAIR, C.W.; MAGEE, M. The perspective view of three points. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.13, n.1, p.66-73, Jan. 1991.
- WOLFSON, H.J. Model-based object recognition by geometric hashing. In EUROPEAN CONFERENCE ON COMPUTER VISION, 1^o, Antibes, 1990. **Proceedings**. Berlin, Springer, 1990. p.526-36
- ZHANG, S.; SULLIVAN, G.D.; BAKER, K.D. Automatic construction of a relational model for recognition of a 3D object. **Proceedings of the SPIE**. The International Society for Optical Engineering, v.1609, p.161-72, 1992.
- _____. Using automatically constructed view-independent relational model in 3D object recognition. In EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV'92), Santa Margherita Ligure, 1992. **Proceedings**. Berlin, Springer, 1992. p.778-86.
- ZHAO, F. **Machine recognition as representation and search**. Cambridge, MIT, 1989. (MIT AI Memo 1189)

APÊNDICE I - DICIONÁRIO

Atributo	Feature
Manufatura Auxiliada por Computador (CAM)	Computer Aided Manufacturing
Primitiva	Feature
Projeto Auxiliado por Computador (CAD)	Computer Aided Design
Reconhecimento Baseado em Modelos	Model-based Recognition
Imagem de Intensidade Luminosa	Intensity Image
Correspondência	Matching
Estruturas Aramadas	Wire-frame
Homotetia	Escaling
Suavidade	Smoothness
Alinhamento	Alignment
Rótulo	Label
Objeto Retirado do Topo de um Amontoado Formado por Objetos Idênticos Empilhados	Bin Picking
Atributos Salientes	Salient Features

Desenho de Linhas	Line Drawing
Representação por Borda	Boundary Representation
Cilindros Generalizados	Generalized Cylinders
Cones Generalizados	Generalized Cones
Detecção de Bordas	Edge Detection
Estruturas por Varredura	Sweep Representation
Imagem Gaussiana Estendida	Extended Gaussian Image
Pontos da Imagem	Picture Element (pixel)
Linhas Escondidas	Hidden Lines
Recozimento por Campo Médio	Mean Field Annealling
Modelo de Membrana Fraca	Weak Membrane Model
Partição de Águas	Watershed