

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E AUTOMAÇÃO
INDUSTRIAL

**Imagens de Profundidade: Segmentação e Representação
por Superfícies Planares**

Autora: **Olga Regina Pereira Bellon**

Orientador: **Clésio Luis Tozzi**

Tese submetida à Faculdade de Engenharia Elétrica e de
Computação da Universidade Estadual de Campinas,
como como requisito parcial para obtenção do Título de
Doutor em Engenharia Elétrica.

08 de setembro de 1997

Este exemplar corresponde à redação final da tese
defendida por: Olga Regina Pereira
Bellon e aprovada pela Comissão
Julgada em 08 / 09 / 1997.
Orientador: Clésio Luis Tozzi

06
29/02
B417i

32740/BC

UNICAMP
BIBLIOTECA CENTRAL

UNIDADE BC
 N.º CHAMADA: UNICAMP
B417i
 V. Es.
 TOMBO BC/ 30840
 PROC. 395/98
 C D
 PREÇO 34,00
 DATA 28/03/98
 N.º CPD _____

CM-00105143-1

FICHA CATALOGRÁFICA ELABORADA PELA
 BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

B417i Bellon, Olga Regina Pereira
 Imagens de profundidade: segmentação e
 representação por superfícies planares / Olga Regina
 Pereira Bellon.--Campinas, SP: [s.n.], 1997.

Orientador: Clésio Luis Tozzi.
 Tese (doutorado) - Universidade Estadual de
 Campinas, Faculdade de Engenharia Elétrica e de
 Computação.

1. Processamento de imagens. 2. Visão por
 computador. I. Tozzi, Clésio Luis. II. Universidade
 Estadual de Campinas. Faculdade de Engenharia Elétrica
 e de Computação. III. Título.

Capítulo 1: Introdução	1
Capítulo 2: Conceitos sobre imagens de profundidade	
2.1. Introdução	15
2.2. Definição de imagem de profundidade	15
2.3. Princípios básicos de obtenção	18
2.3.1. Triangulação	19
2.3.2. Tempo de reflexão	20
2.3.2.1. Ultra-som	21
2.3.2.2. <i>Laser</i>	22
2.4. Transformações de coordenadas	24
2.4.1. Os índices i e j são incrementos de distância	24
2.4.2. Os índices i e j são incrementos angulares	25
2.5. Estado da arte: Câmera 3D arco-íris	26
2.6. Sumário	28
Capítulo 3: Extração de características	
3.1. Introdução	29
3.2. Tipos de bordas	30
3.2.1. Bordas de profundidade	31
3.2.2. Bordas de orientação	32
3.2.3. Bordas suaves	32
3.2.4. Considerações sobre valores de limiar	34
3.3. Ajuste de superfícies	34
3.3.1. Superfícies planares	34
3.3.2. Superfícies de maior grau	36
3.4. Curvaturas de superfícies	36
3.4.1. Curvaturas Principais, Média, e Gaussiana	37
3.5. Cálculo das características	41
3.5.1. Descrição das características	42
3.5.2. Cálculo das derivadas	44
3.6 Sumário	45

Capítulo 4: Segmentação de imagens

4.1. Introdução	46
4.2. Definição formal	47
4.3. Descrição das abordagens	48
4.3.1. Análise de histograma	48
4.3.2. Crescimento de regiões	50
4.3.3. Divisão-e-união	51
4.3.4. Agrupamento	52
4.4. Considerações sobre super-segmentação	54
4.5. Considerações sobre pré-processamento	55
4.6. Trabalhos relacionados	56
4.6.1. Besl&Jain	56
4.6.2. Taylor et al.	58
4.6.3. Hoffman&Jain	59
4.6.4. Abdelmalek	61
4.6.5. Krishnapuram&Munshi	62
4.6.6. Sabata et al.	64
4.6.7. Bhandarkar&Siebert	67
4.6.8. Yokoya&Levine	68
4.6.9. Gil et al.	69
4.6.10. Yu et al.	70
4.7. Comentários	72
4.8. Sumário	77

Capítulo 5: O método de segmentação

5.1. Introdução	78
5.2. Definição de segmentação/reconstrução	79
5.3. Descrição da imagem	81
5.4. Descrição sumária	81
5.5. Detecção de bordas e cálculo das normais	83

5.6. Cálculo da estimativa do número de regiões	85
5.6.1. Construção do mapa de bordas	85
5.6.2. Contagem do número de regiões	87
5.6.3. Objetos poliédricos x não-poliédricos	88
5.7. Cálculo do erro de aproximação	90
5.8. Agrupamento	92
5.9. Refinamento da imagem segmentada	94
5.10. Cálculo das equações dos planos	95
5.11. Cálculo do erro de reconstrução	96
5.12. Comparação com outros métodos	97
5.12.1. Técnicas utilizadas	97
5.12.2. Tipos de segmentos resultantes	98
5.12.3. Principal contribuição	99
5.12.4. Principal limitação	100
5.13. Sumário	102

Capítulo 6: O método de representação

6.1. Introdução	103
6.2. Trabalhos relacionados	104
6.3. A estrutura utilizada	107
6.4. Determinação dos vértices e arestas	111
6.5. Principais problemas associados	120
6.5.1. Limiar para proximidade entre duas entidades	120
6.5.2. Inconsistências geométicas	122
6.6. Medida da qualidade da representação	125
6.7. Sumário	126

Capítulo 7: Discussão sobre a construção de modelos 3-D completos

7.1. Introdução	127
7.2. Casamento de vistas	127
7.3. Trabalhos relacionados	128
7.3.1. Seales&Faugueras	128
7.3.2. Higuchi et al.	129
7.3.3. Masuda&Yokoya	130
7.3.4. Soucy&Laurendeau	130
7.3.5. Chen&Medioni	131
7.3.6. Sapidis&Besl	132
7.4. Usando a representação desenvolvida	132
7.5. Sumário	135

Capítulo 8: Resultados experimentais

8.1. Introdução	136
8.2. Detalhes de implementação	137
8.3. Resultados visuais	139
8.4. Tabela comparativa	150
8.5. Exemplo de estrutura obtida	152
8.6. Um exemplo mais complexo	153
8.7. Um exemplo real	156
8.8. Resultados visuais para a reconstrução 3-D	160

Capítulo 9: Conclusão e trabalhos futuros

9.1. Conclusão	163
9.2. Trabalhos futuros	167

Referências	170
--------------------	------------

Dedico este trabalho aos meus pais.

Resumo

Este trabalho apresenta um método integrado para segmentação e representação de imagens de profundidade por superfícies planares. O processo de segmentação utiliza uma associação das técnicas de agrupamento e detecção de bordas. A informação sobre bordas fornece ao algoritmo de agrupamento uma estimativa inicial sobre o número de regiões que compõem a imagem, e permite separar superfícies vizinhas distintas de mesma orientação. O resultado da segmentação é um conjunto de segmentos aproximados por planos, cujos coeficientes são conhecidos. Com a informação obtida na segmentação, obtém-se a reconstrução das superfícies observadas na imagem. O erro de reconstrução obtido, que é a diferença entre a superfície original e a reconstruída, é utilizado como medida de qualidade da segmentação. O processo de representação recebe os dados da segmentação e define uma representação poliédrica para a superfície segmentada, baseada na representação por fronteira (B-Rep). Resultados experimentais, usando imagens de profundidade sintéticas e reais, são apresentados. Embora a representação final seja poliédrica, o método é aplicado não apenas a superfícies poliédricas, mas também a superfícies não-poliédricas.

Abstract

This work presents an integrated method for segmentation and representation of range images through planar surfaces. The segmentation process associates clustering and edge detection techniques. The edge information supplies an initial estimation to the number of regions that compound the image. That estimation is needed to initialize the clustering algorithm. The edge information also allows to distinguish those neighbour regions with the same orientation. The segmentation result is a set of segments approximated by planes with known coefficients. The observed surfaces are reconstructed by using the segmentation process information. The reconstruction error, which is a difference between the original surfaces and the reconstructed surfaces, is used as a segmentation quality measure. The representation process takes the segmentation information and defines a polyhedral representation to the segmented surfaces. This polyhedral representation is based on the B-Rep representation. Experimental results, using both synthetic and real images, are presented. The representation is polyhedral, but the method can be applied also to non-polyhedral surfaces.

Capítulo 1: Introdução

Nos últimos vinte e cinco anos, a análise de imagens a partir de imagens de intensidade luminosa tem tomado muito mais atenção do que por qualquer outro tipo de imagem, em parte pela qualidade dos equipamentos desenvolvidos para sua obtenção. No entanto, estas imagens têm como grande limitação a falta de informação tridimensional (3-D) explícita sobre a cena observada, o que dificulta a sua interpretação automática. Isto fez com que diversos métodos para obtenção desta informação fossem desenvolvidos para imagens de intensidade luminosa, entre eles: estéreo-visão, *shape from shading*, *shape from motion*, etc (Aggarwal e Chien, 1989; Ballard e Brown, 1982; Nitzan, 1988).

Paralelamente, os sensores de profundidade têm apresentado melhoramentos significativos ao longo dos últimos anos (Beraldin et al., 1992; Besl, 1989A; Geng, 1996; Jarvis, 1983A; Jarvis, 1983B). As imagens geradas por estes sensores, aqui denominadas imagens de profundidade (*range images*), fornecem a informação 3-D diretamente, eliminando-se a necessidade de extrair esta informação quando do seu processamento e favorecendo a análise da imagem.

Com o objetivo de aumentar a qualidade dos resultados obtidos na análise de imagens, diversos trabalhos (Aggarwal, 1990; Asar et al., 1990; Gil et al., 1983; Hackett e Shah, 1989; Luo e Kay, 1989; Mitiche e Aggarwal, 1986; Zhang e Wallace, 1993) têm associado imagens de intensidade luminosa e imagens de profundidade. A integração das informações ocorre, principalmente, na etapa de segmentação das

imagens, e a idéia básica é avaliar, para cada região da imagem, qual sensor possui informação mais confiável. No entanto, esta abordagem é muito custosa e exige complexas técnicas estatísticas no processo de integração dos dados (Abidi e Gonzalez, 1992).

Independente do tipo de imagem utilizada, o processo de análise de imagens, pode ser dividido em três etapas principais (Gonzalez e Woods, 1992):

- (1) Processamento de baixo nível;
- (2) Processamento de nível intermediário;
- (3) Processamento de alto nível.

Embora as imagens de profundidade simplifiquem a tarefa de análise de imagens, por fornecerem a informação 3-D diretamente, questões importantes são mantidas em suas diversas etapas.

O processamento de baixo nível consiste na **aquisição** e **pré-processamento** da imagem. A aquisição engloba todo o processo de obtenção da imagem, e os principais problemas associados estão relacionados à baixa resolução e quantidade de ruído da imagem resultante, conforme será discutido na Seção 2.3. O pré-processamento tem como função fornecer uma imagem melhorada à próxima etapa, por exemplo, reduzindo ruídos da imagem.

O processamento de nível intermediário é composto pela **segmentação** mais a **representação** e **descrição** da imagem. A etapa de segmentação deve identificar e

isolar partes homogêneas significativas da imagem. Este processo é crucial para uma boa análise da imagem, e muitos esforços vem sendo aplicados para tentar solucionar seus problemas. Entretanto, existem ainda diversos pontos importantes a serem solucionados, entre eles podem ser citados:

- (1) A escolha do método mais adequado à aplicação em questão;
- (2) A definição do critério de homogeneidade a ser utilizado na técnica de segmentação escolhida;
- (3) A determinação do número correto de regiões que compõem a imagem;
- (4) A obtenção de técnicas robustas, ou seja, pouco sensíveis à existência de ruídos na imagem;
- (5) A determinação de medidas de qualidade significativas para avaliar os resultados obtidos;
- (6) A redução de restrições sobre a classe dos objetos a serem observados na imagem.

A etapa de representação, seguinte à etapa de representação, deve tomar as informações decorrentes desta e determinar uma estrutura para descrever o conteúdo da imagem, de modo a reduzir as informações contidas na mesma a um conjunto discriminatório e facilitar sua posterior interpretação. Dentre as dificuldades encontradas neste processo, destacam-se:

- (1) A escolha da representação mais adequada para a imagem, de acordo com os dados fornecidos pela segmentação e com a aplicação em questão;
- (2) A determinação de representações que preservem informações importantes sobre as características das superfícies observadas;
- (3) A obtenção de estruturas coerentes mesmo com resultados de segmentação imprecisos;
- (4) A determinação de medidas de qualidade significativas para avaliar os resultados obtidos.

A última etapa da análise de imagens, o processamento de alto nível, consiste no **reconhecimento e interpretação** das informações geradas pelas etapas anteriores. Os principais problemas relacionados a esta etapa são:

- (1) A necessidade de se lidar com informações possivelmente imprecisas das etapas anteriores;
- (2) O custo computacional do processo de correspondência entre as representações obtidas e os modelos previamente armazenados.

Ao longo de todo o processo de análise de imagens, uma **base de dados** recebe as informações intermediárias que são extraídas da imagem e pode ser consultada a qualquer momento. Esta base de dados é inicializada com informações relevantes sobre a cena observada como, por exemplo, modelos de objetos para

reconhecimento. Um esquema ilustrando as etapas descritas é mostrado na Figura 1.1.

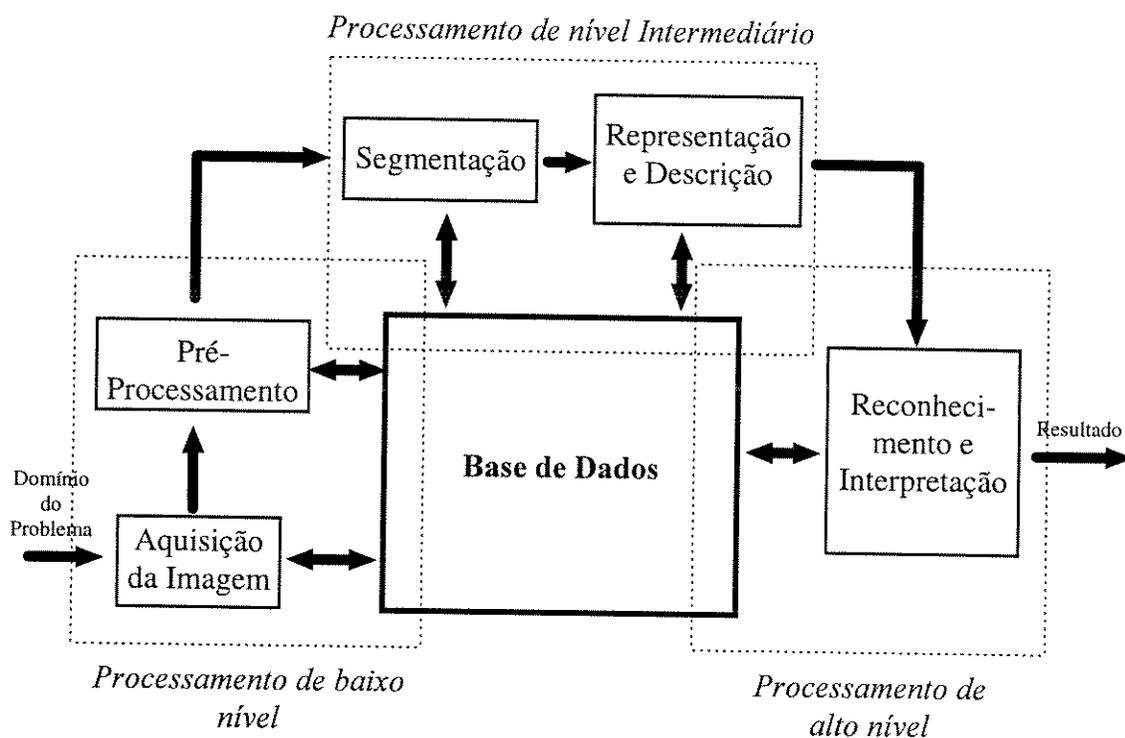


Figura 1.1. Esquema geral com as principais etapas da análise de imagens.

(Gonzalez e Woods, 1992)

Este trabalho é uma contribuição à solução dos problemas de segmentação e representação a partir de imagens de profundidade.

A segmentação de imagens pode ser definida como o processo de particionar uma imagem em segmentos significativos que correspondam a objetos, partes de objetos, ou grupos de objetos da cena observada. Uma definição formal sobre segmentação de imagens é apresentada na Seção 4.2 deste trabalho.

As principais técnicas utilizadas para solucionar o problema da segmentação de imagens são (Bellon, 1996; Pal e Pal, 1993):

- (1) Análise de histograma (Sahoo et al., 1988);
- (2) Crescimento de regiões (Besl e Jain, 1988; Sapidis e Besl, 1995);
- (3) Divisão-e-união (Grosky e Jain, 1986; Sabata et al., 1993; Taylor et al., 1989);
- (4) Agrupamento (Bhandarkar e Siebert, 1992; Coleman e Andrews, 1979; Hoffman e Jain, 1987; Krishnapuram e Munshi, 1991).

A dificuldade maior normalmente encontrada nos algoritmos que utilizam análise de histograma, crescimento de regiões, ou divisão-e-união é a necessidade de se definir valores rígidos de limiar. Os algoritmos que utilizam agrupamento são mais independentes em relação a este tipo de problema mas, apesar desta vantagem, demandam um outro tipo de consideração: a determinação do número ideal de regiões desejadas para a imagem. Existem duas soluções para esta questão:

- (1) Ter uma boa estimativa do número de regiões que compõem a imagem;
- (2) Gerar diversos agrupamentos, cada um segmentando a imagem em um número diferente de regiões, e desenvolver uma metodologia para determinar qual agrupamento melhor representa a imagem, ou seja, qual o número correto de segmentos para a imagem.

A segunda opção é ineficiente em muitas aplicações, devido ao alto custo computacional exigido para gerar diversos agrupamentos, além do que as metodologias normalmente utilizadas para seleção do melhor agrupamento nem sempre levam ao resultado ótimo. A primeira solução, embora restritiva sob o aspecto de que se deve ter alguma informação sobre o número de regiões visíveis na cena observada, é uma boa abordagem desde que seja possível obter esta estimativa através dos dados da própria imagem e realizar uma atualização, com base nos resultados de segmentação obtidos, para a estimativa inicial aplicada. Assim, para aplicar eficientemente a técnica de agrupamento, é desejável iniciar o processo com alguma estimativa prévia, não necessariamente precisa, sobre o número de regiões que compõem a imagem, e desenvolver um método que identifique a necessidade ou não de modificar este número.

O método de segmentação desenvolvido neste trabalho utiliza um algoritmo de agrupamento (*K-Means*). Para ajudar a solucionar as limitações decorrentes desta técnica, realiza-se sua associação com técnicas de detecção de bordas.

A detecção de bordas em imagens de profundidade (Bhandarkar e Siebert, 1992; Faugueras e Herbert, 1987; Gil et al., 1983; Hoffman e Jain, 1987; Mitiche e Aggarwal, 1983; Parvin e Medioni, 1989; Wani e Batchelor, 1994; Yokoya e Levine, 1989) gera informações geométricas mais ricas sobre a cena observada, se comparada à detecção de bordas em imagens de intensidade luminosa, porque a informação tridimensional explícita contida nas imagens de profundidade permite a identificação de diversos tipos de bordas: de profundidade, de orientação, e suave, conforme

descrito na Seção 3.2. Desta forma, a identificação de bordas em imagens de profundidade pode contribuir para a determinação da representação ou modelo de objetos, além de fornecer uma indicação significativa do número de regiões homogêneas que compõem a imagem.

Dentre as opções de abordagens a serem utilizadas para a segmentação, ainda é possível tomar a decisão entre segmentação por planos ou por superfícies de maior grau. A segmentação por superfícies de maior grau produz resultados altamente discriminatórios e adequados ao reconhecimento 3-D de objetos. Entretanto, o custo computacional para sua obtenção é muito alto, em comparação à segmentação por planos, pois é necessário utilizar um número maior de parâmetros para especificar a superfície. Além disso, os resultados são menos confiáveis, pois uma pequena variação (ou erro) no cálculo destes parâmetros pode alterar significativamente o tipo da superfície em questão. A segmentação em regiões planares é de fácil computação, baseada nos valores das normais, além do que pedaços de regiões planares podem servir como base para um posterior processo de união em superfícies de maior grau, caso necessário. Uma discussão sobre ajuste de superfícies planares e de maior grau é apresentada na Seção 3.3.

O processo de segmentação desenvolvido neste trabalho utiliza como característica de entrada a informação sobre os coeficientes da normal a cada *pixel* da imagem. O resultado obtido é um conjunto de regiões e as equações dos planos mais bem ajustados a cada uma delas. Com as informações resultantes da segmentação, obtém-se a reconstrução das superfícies que compõem a cena observada. O erro de

reconstrução obtido, que é a diferença entre a imagem original e a reconstruída, é utilizado como medida de qualidade da segmentação.

Após realizado o processo de segmentação na imagem, é necessário produzir uma representação para o conjunto de superfícies segmentadas. O objetivo desta representação é transformar a informação obtida na segmentação para uma forma mais adequada a aplicações de visão computacional. A representação construída deve ressaltar características discriminatórias das superfícies, por exemplo, de modo que possa ser comparada com modelos previamente armazenados em uma base de dados.

O processo de representação é fortemente relacionado aos resultados da segmentação, pois só podemos representar aquilo que foi identificado e uma identificação incorreta leva a uma representação errônea. Isto sugere que a solução dos dois problemas seja realizada de forma integrada, de modo que discrepâncias obtidas na representação atuem como uma realimentação ao processo de segmentação (Bellon e Tozzi, 1997). No entanto, a despeito de sua importância, esta integração tem sido pouco contemplada na literatura, em especial do ponto de vista topológico.

Alguns trabalhos realizam os processos de segmentação e descrição de imagens de forma integrada (Besl e Jain, 1988; Hoffman e Jain, 1987; Newman et al., 1993; Yokoya e Levine, 1989). Entretanto, estes não tratam o problema de descrição em termos da topologia da cena. Bhandarkar e Siebert, 1992, criam uma

representação topológica para objetos segmentados em imagens de profundidade, mas seu trabalho é restrito a objetos poliédricos. Faugueras e Herbert, 1987, não impõem a restrição de objetos poliédricos e solucionam o problema da representação por um processo de triangulação dos pontos que compõem a superfície dos objetos. A representação por triangulação tem boa aplicação em visualização e reconstrução de objetos, pois pode levar a uma reconstrução com pouco erro de aproximação em relação ao objeto original. Entretanto, esta abordagem tem alto custo computacional, tanto em relação à definição da estrutura que irá comportar a representação da cena quanto na etapa de correspondência entre representações.

Neste trabalho foi desenvolvido um método para definir uma representação poliédrica, dirigida à topologia, para imagens de profundidade segmentadas em superfícies planares. A representação pode ser utilizada não apenas para objetos poliédricos, mas também para objetos não-poliédricos. Neste último caso, o objeto deve sofrer uma aproximação poliédrica.

A idéia básica da representação é tomar as K regiões resultantes da segmentação (R_1, R_2, \dots, R_k) e construir uma representação poliédrica para este conjunto R de regiões, onde $R = \cup_{i=1}^K R_i$. A questão principal aqui é encaixar os segmentos de planos obtidos de modo a formar uma superfície poliédrica, onde tanto as bordas de profundidade quanto de orientação são preservadas. O resultado da representação criada é uma estrutura para R contendo: um conjunto de regiões

componentes; as coordenadas 3-D dos vértices de cada região, ordenados de modo a formarem um polígono; e a equação do plano associado a cada região.

A seguir, são apresentadas as principais vantagens e contribuições dos métodos de segmentação e de representação desenvolvidos

- (1) Solucionam os problemas de segmentação e de representação de forma integrada: Esta integração permite a realimentação dos resultados da representação no processo de segmentação, melhorando o desempenho global dos métodos.
- (2) Preservam tanto as bordas de profundidade quanto as de orientação: Estas características geométricas são muito importantes para a obtenção de um modelo completo e reconhecimento 3-D.
- (3) Permite a determinação de medidas de qualidade significativas: Muitos trabalhos solucionam os problemas apresentados sem definição de medidas para qualificar seus resultados. O trabalho desenvolvido fornece medidas de qualidade para todas as suas etapas.
- (4) Não depende do conhecimento prévio do número de regiões da imagem: A associação da técnica de detecção de bordas à técnica de agrupamento permite determinar uma estimativa inicial para este número, e sua atualização depende apenas das medidas de qualidade obtidas na segmentação.

- (5) Apresenta uma nova abordagem para a determinação do número ótimo de regiões da imagem: Através da informação sobre a reconstrução da imagem, pode-se identificar a necessidade de dividir ou agrupar regiões segmentadas.
- (6) Pode ser aplicado a imagens contendo objetos poliédricos e/ou não-poliédricos: Apesar da representação final ser poliédrica, esta pode ser aplicada sobre objetos não-poliédricos, gerando boas aproximações dos mesmos. Estas aproximações podem ser refinadas pelo aumento no número de segmentos utilizado como entrada no algoritmo de agrupamento.
- (7) Superfícies de maior grau podem ser obtidas: Apesar da segmentação e representação por planos, superfícies de maior grau podem ser obtidas pela união de segmentos de planos.
- (8) É adequado à interpretação da imagem: Tomando-se diferentes vistas de uma mesma cena, pode-se realizar a correspondência (*matching*) das representações obtidas, com base em seus polígonos componentes. Com isso, é possível construir um modelo 3-D completo para um objeto e utilizar a informação obtida para seu reconhecimento 3-D.

Este trabalho é composto por nove capítulos, cujas descrições sumárias são apresentadas a seguir.

Capítulo 1: Introdução. Apresenta o problema a ser solucionado, os objetivos do trabalho, e suas principais contribuições.

Capítulo 2: Conceitos sobre imagens de profundidade. Apresenta os conceitos básicos sobre imagens de profundidade. Aborda a definição de imagens de profundidade, os princípios básicos para sua obtenção, os métodos de transformações de coordenadas e, finalmente, uma breve discussão sobre o estado da arte em câmeras para obtenção de dados de profundidade.

Capítulo 3: Extração de características. Mostra como algumas características (p.ex. bordas, ajuste de superfícies, curvaturas de superfícies) podem ser extraídas das imagens de profundidade e utilizadas como atributos para segmentação e classificação de regiões na imagem.

Capítulo 4: Segmentação de imagens. Apresenta o problema da segmentação propriamente dito, onde são descritas as abordagens mais utilizadas. Ainda neste capítulo, são discutidos diversos trabalhos relacionados, encontrados na literatura, e são apresentadas conclusões sobre o assunto.

Capítulo 5: O método de segmentação. Apresenta o método desenvolvido para segmentação de imagens de profundidade, abordando também a questão da reconstrução de suas superfícies componentes.

Capítulo 6: O método de representação. Mostra como criar uma representação poliédrica robusta para conjuntos de regiões conexas ajustadas por planos, fornecidas pela segmentação.

Capítulo 7: Discussão sobre a construção de modelos 3-D completos. Apresenta uma discussão sobre soluções para o problema de construção de um modelo 3-D completo, mais adequado ao reconhecimento 3-D de objetos. Indica como aplicar os resultados obtidos no método de segmentação e representação desenvolvido para a solução deste problema.

Capítulo 8: Resultados experimentais. Neste capítulo estão concentrados os resultados experimentais obtidos, que mostram a viabilidade do método proposto, além de alguns detalhes de implementação.

Capítulo 9: Conclusão e trabalhos futuros. Apresenta as conclusões obtidas e indica trabalhos futuros.

Capítulo 2: Conceitos sobre imagens de profundidade

2.1. Introdução

Este capítulo apresenta os conceitos básicos sobre as imagens de profundidade, necessários para uma boa compreensão do conteúdo deste trabalho. Além de definições, discute-se os princípios básicos para a obtenção das imagens de profundidade, apresenta-se a metodologia para realizar as transformações de coordenadas, e realiza-se uma breve discussão sobre o estado da arte em câmeras para obtenção de dados de profundidade.

2.2. Definição de imagens de profundidade

Uma imagem de profundidade é um conjunto de medidas de distância da origem de um sistema de coordenadas de referência até pontos na superfície dos objetos da cena observada.

Se estas medidas de distância são tomadas em relação a três eixos de coordenadas ortogonais na forma de pontos esparsos, de modo irregular, diz-se que a imagem está na forma xyz . Se as medidas de distância indicam a profundidade ao longo de vetores tridimensionais, indexados por dois inteiros (i,j) , numa forma regular ou matricial, diz-se que a imagem está na forma r_{ij} . Deste modo, qualquer imagem na forma r_{ij} pode ser convertida na forma xyz , mas não o inverso.

Se o intervalo de amostragem dos pontos da superfície de uma imagem de profundidade na forma xyz são consistentes com as direções vertical e horizontal, a imagem pode ser representada na forma de uma matriz de valores de profundidade r_{ij} (quantizados e escalados), onde os valores correspondentes das coordenadas xyz são determinados pela posição da linha e coluna na matriz e pelo valor de profundidade, respectivamente.

Considera-se uma imagem de profundidade uma amostra com mais do que 100 pontos (Besl, 1989A). Um conjunto menor de pontos é denominado dados esparsos de profundidade, em comparação com dados densos de profundidade das imagens de profundidade. Por exemplo, a matriz 20x20, mostrada na Figura 2.1, relativa a distâncias de pontos de uma superfície representa uma imagem de profundidade.

A Figura 2.2 é relativa à mesma imagem de profundidade, mostrada como uma matriz na Figura 2.1. Na Figura 2.2, os dados de profundidade foram mostrados de modo a permitir a visualização da superfície imageada.

171	160	163	163	166	166	168	166	168	166	163	160	163	163	160	163	166	163	166	163
168	166	166	163	166	163	168	166	166	166	163	163	166	163	166	163	166	160	163	163
168	168	166	166	166	163	160	166	166	171	168	168	168	166	180	163	166	160	160	166
168	163	166	166	163	163	160	163	176	174	185	177	185	179	212	196	185	204	166	186
163	166	160	166	163	163	166	190	174	168	168	182	185	190	201	196	199	182	196	199
166	163	163	163	168	160	163	166	166	163	168	177	190	188	199	188	190	196	193	185
163	166	166	157	160	160	160	171	160	168	168	182	199	199	199	193	199	188	193	193
160	160	160	166	157	160	168	166	166	163	163	182	201	199	190	188	190	190	193	193
163	166	157	163	160	157	160	177	166	160	171	201	215	199	196	201	190	190	188	188
155	160	160	163	160	163	160	166	166	163	163	204	207	207	190	185	193	190	196	196
157	155	163	160	157	157	168	166	168	163	177	188	201	199	196	196	201	182	210	196
157	157	155	157	160	157	163	171	163	157	155	204	185	196	193	188	198	188	193	201
157	160	155	155	157	157	168	168	163	168	168	190	201	201	196	188	190	193	185	193
157	155	160	160	157	157	163	157	157	160	157	182	204	190	185	190	190	188	185	188
157	157	157	160	157	157	152	166	160	163	166	193	196	193	199	190	190	165	190	185
155	157	160	160	160	152	168	152	163	152	188	171	212	212	193	190	188	182	188	185
152	157	155	155	152	155	149	163	160	155	157	185	210	210	212	215	210	185	204	193
155	155	157	152	152	155	155	171	174	168	171	188	188	199	188	204	188	185	215	207
155	157	152	157	149	157	157	168	179	204	182	221	174	193	182	179	212	188	201	182
155	155	155	155	152	149	146	174	188	193	188	185	168	170	171	190	190	193	190	179

Figura 2.1: Matriz 20x20 de medidas de profundidade: forma r_{ij} da imagem de profundidade. (Besl, 1989A)

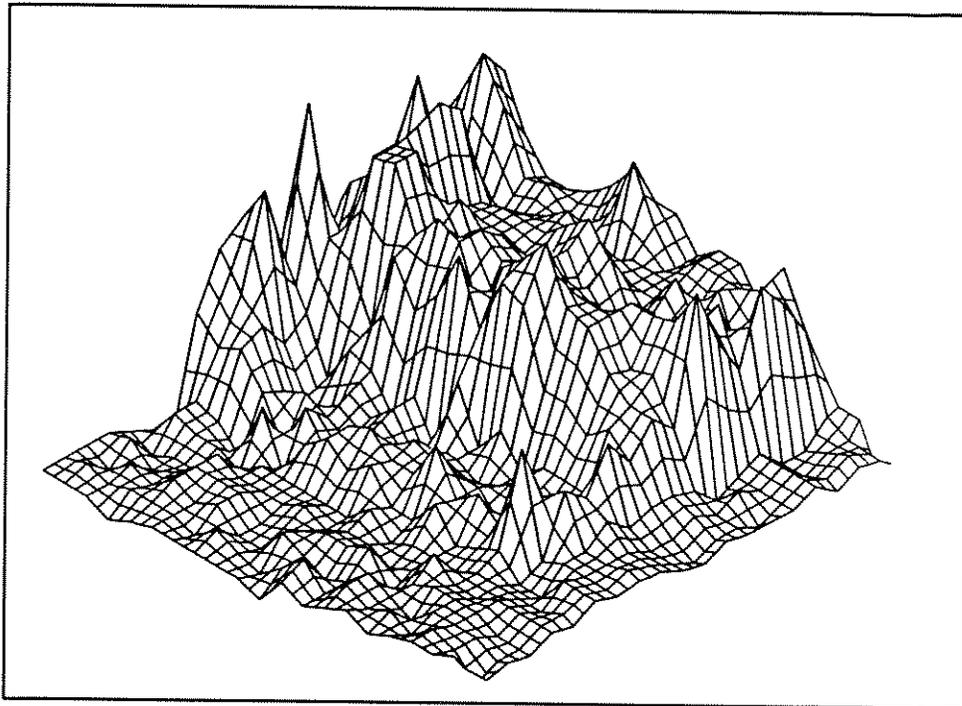


Figura 2.2. A matriz de profundidade da Figura 2.1 mostrada como uma superfície.

(Besl, 1989A)

2.3. Princípios básicos de obtenção

A primeira geração de sensores para obtenção direta de imagens de profundidade (Besl, 1989A; Jarvis, 1983A; Jarvis, 1983B) apresentavam problemas tais como: tempo de aquisição inadequado para a maioria das aplicações de visão computacional, alto custo, imprecisão de medidas, etc. Com a evolução das tecnologias associadas, a grande maioria dos sensores de profundidade da geração atual (Arman e Aggarwal, 1993; Beraldin et al., 1992; Geng, 1996; Kweon e Kanade, 1992) não contêm os problemas citados.

Nesta seção, apresentam-se os princípios básicos para a obtenção direta de imagens de profundidade. Os aspectos tecnológicos específicos sobre os equipamentos utilizados, tais como tempo de aquisição, precisão, custo, etc, só serão discutidos quando representarem características qualitativas inerentes ao método de obtenção. Métodos para obtenção indireta de imagens de profundidade são descritos em Besl, 1989A e Jarvis, 1983A.

Existem duas metodologias para se obter diretamente imagens de profundidade (Arman e Aggarwal, 1993; Besl, 1989A; Jarvis, 1983A). Uma é baseada em operações de triangulação, e a outra é baseada no tempo de reflexão do sinal emitido. As idéias principais relacionadas a cada uma das abordagens são mostradas a seguir.

2.3.1. Triangulação

Nesta abordagem, a configuração utilizada pelo sensor consiste em uma fonte pontual emissora de energia e um receptor, onde este tem localização conhecida em relação ao emissor, como é mostrado na Figura 2.3.

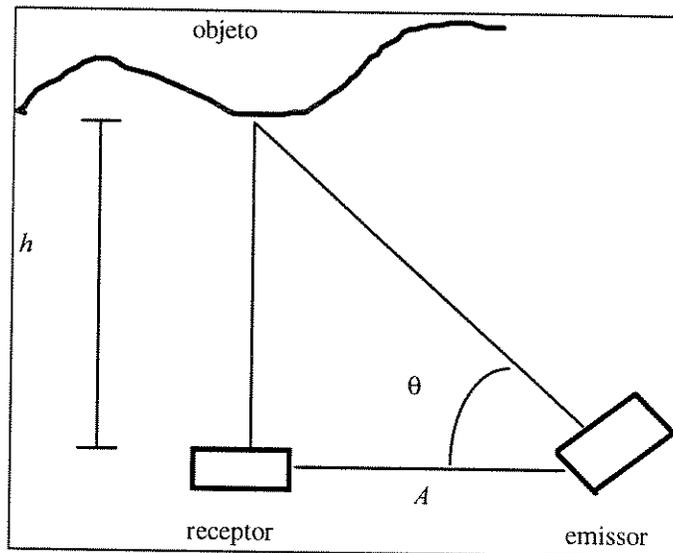


Figura 2.3: Geometria para cálculo de profundidade por triangulação.

(Jarvis, 1983A)

A informação de profundidade (h) de cada ponto é obtida por um processo de triangulação, conhecendo-se a distância fonte-receptor (A) e o ângulo entre o feixe de luz emitido e a base (θ), de acordo com a Equação 2.1. A imagem de profundidade pode ser obtida tomando-se diversas medidas através de um deslocamento rotacional da fonte de emissão de energia, para ângulos θ conhecidos.

$$h = A \tan \theta \quad (2.1)$$

Neste método, como mostra a Equação 2.1, quanto maior for a distância entre o receptor e o emissor, maior será a precisão dos resultados. No entanto, o aumento desta distância poderá acarretar problemas de oclusão, como por exemplo partes do objeto serem alcançadas pelo emissor mas não refletirem energia na direção do receptor. Este é um problema típico das técnicas que utilizam triangulação.

Apesar de terem um menor custo associado, os equipamentos baseados em triangulação exigem muito mais tempo para obtenção total de uma imagem de profundidade do que os baseados em tempo de reflexão (Jarvis, 1983A).

2.3.2. Tempo de reflexão

Neste caso, o emissor e o receptor são coaxiais, eliminando os problemas de oclusão.

A distância absoluta entre sensor e objeto é disponível diretamente, através de uma relação com o tempo decorrido entre emissão e recepção de sinal, baseada na equação:

$$v\tau = 2r = \textit{distância percorrida} \quad (2.2)$$

onde v é a velocidade de propagação do sinal, r é a distância entre a fonte e o ponto do objeto, e τ é o tempo que o sinal leva para ir do transmissor ao objeto e retornar.

Os equipamentos que medem distâncias por tempo de reflexão têm características diferentes, dependendo se utilizam ultra-som ou *laser*, onde a

velocidade do som e a velocidade da luz são, respectivamente, os parâmetros utilizados para se obter a medida.

2.3.2.1. Ultra-som

As imagens obtidas a partir deste equipamento apresentam pouca resolução, devido à dificuldade de obter um padrão de ultra-som bem focado (Jarvis, 1983A). Isto faz com que o ultra-som seja mais utilizado na obtenção de dados esparsos.

Como exemplo de aplicação do ultra-som para obtenção de dados esparsos, Brown, 1985, apresenta um sistema de visão, aplicado a robôs autoguiados, que reconhece com sucesso obstáculos formados por objetos poliédricos.

Uma outra desvantagem do ultra-som é que, para alguns ângulos de incidência do pulso em relação à superfície do objeto, tem-se muito pouca energia refletida, como pode ser visto na Figura 2.4. O valor máximo aceitável para o ângulo de incidência do sinal de ultra-som é de 40° .

Deste modo, pode-se afirmar que o ultra-som é pouco adequado para aplicações onde se exige grande precisão na extração de características da imagem.

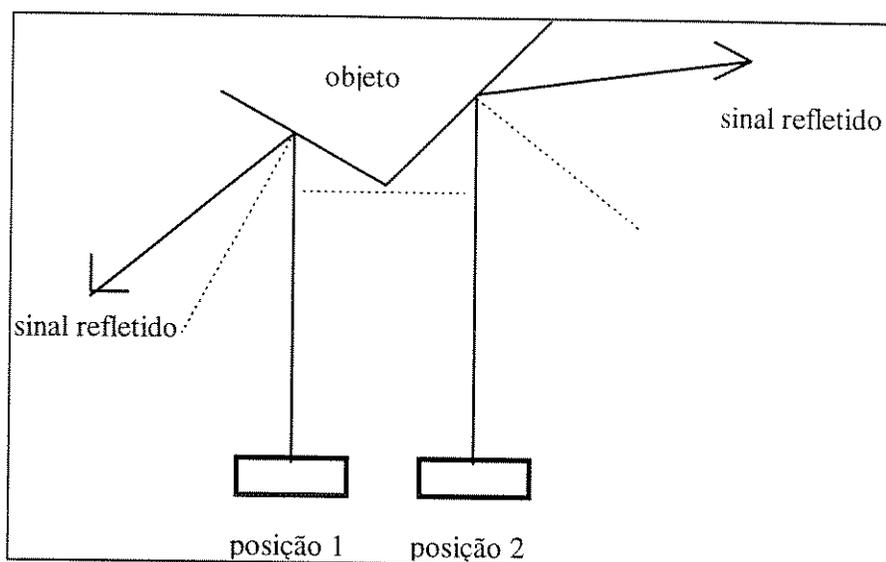


Figura 2.4: Reflexão do sinal por ultrassom. (Jarvis, 1983A)

2.3.2.2. Laser

Existem duas maneiras de obter distâncias a partir de fontes *lasers*. A primeira utiliza a medida do deslocamento de fase de uma onda contínua de um fluxo *laser* modulado, entre sua partida da fonte e seu retorno ao detector coaxial. A segunda mede o tempo que o pulso *laser* leva para deixar a fonte, atingir o objeto alvo, e retornar coaxialmente ao detector. No primeiro caso pode-se obter diretamente informações de profundidade e de intensidade luminosa dos objetos da cena (Besl, 1989A; Beraldin et al., 1992; Gil et al., 1983; Jarvis, 1983A; Nitzan et al., 1977).

O sistema mostrado na Figura 2.5 utiliza uma fonte *laser* modulada para fornecer duas imagens em correspondência, uma de intensidade luminosa e outra de profundidade. O equipamento tem três componentes principais: um transmissor, um

espelho móvel para gerar o rastreamento, e um receptor. O feixe *laser* emitido pelo transmissor é modulado em amplitude com uma onda senoidal. Este feixe modulado é desviado pelo espelho de modo a cobrir todo o campo de interesse. Quando o feixe atinge o objeto, uma porção da luz é refletida de volta ao longo da direção coaxial com o feixe incidente, capturada pelo espelho, e desviada para o receptor. Um filtro de interferência rejeita a luz ambiente e transmite o sinal retornado para um tubo fotomultiplicador. O fotomultiplicador converte a luz recebida em uma corrente senoidal na frequência modulada. A amplitude desta corrente é proporcional à reflectância do objeto. A fase da corrente, em relação ao sinal transmitido, é proporcional ao tempo que a luz leva para ir do instrumento ao objeto e retornar, possibilitando a determinação do valor de profundidade.

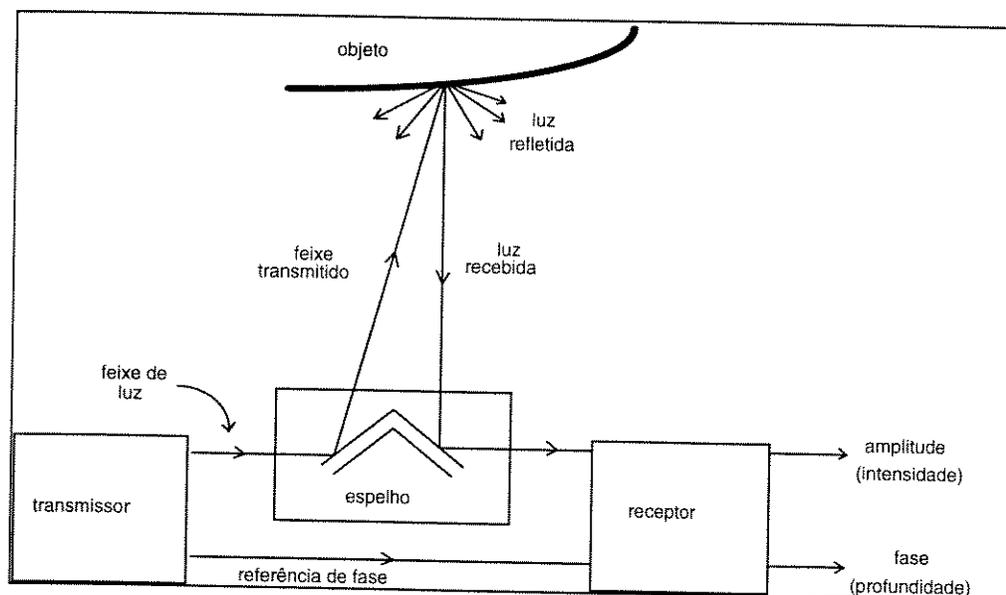


Figura 2.5. Diagrama de bloco simplificado do sensor de intensidade luminosa/profundidade. (Nitzan et al., 1977).

Como o sinal emitido pelo *laser* é melhor focado em relação ao ultra-som, pode-se obter melhores resoluções nas imagens geradas e maior precisão nos resultados. No entanto, a precisão obtida por equipamentos que utilizam *laser* é diretamente dependente da potência do sinal transmitido, sendo que potências muito altas entram em conflito com critérios de segurança e custo (Besl, 1989A; Jarvis, 1983A).

2.4. Transformações de coordenadas

Em algumas aplicações, é necessário transformar a informação da forma r_{ij} para a forma xyz . Duas maneiras comuns de realizar esta transformação são mostradas a seguir, dependendo do que representam os índices i e j na imagem original.

2.4.1. Os índices i e j são incrementos de distância

Se r_{ij} é o valor do pixel na i -ésima linha e j -ésima coluna da matriz, então suas coordenadas 3-D Cartesianas podem ser obtidas pelas seguintes equações:

$$x = a_x + s_x i \quad (2.3)$$

$$y = a_y + s_y j \quad (2.4)$$

$$z = a_z + s_z r_{ij} \quad (2.5)$$

onde os valores de s são incrementos de distância (fatores de escala) e os valores de a são os *offsets* das coordenadas.

A transformação mostrada nas equações 2.3-2.5 é apropriada para imagens de profundidade r_{ij} ortográficas, onde o termo ortográfico significa que as distâncias são medidas ao longo de raios paralelos ortogonais ao plano de imagem.

2.4.2. O índices i e j são incrementos angulares

No sistema de coordenadas esféricas os índices (i,j) correspondem a incrementos angulares nas direções vertical (latitude ou elevação) e horizontal (longitude ou azimute). Neste caso, a transformação de coordenadas é dada por:

$$x = a_x + s_r r_{ij} \cos(is_\phi) \sin(js_\theta) \quad (2.6)$$

$$y = a_y + s_r r_{ij} \sin(is_\phi) \quad (2.7)$$

$$z = a_z + s_r r_{ij} \cos(is_\phi) \cos(js_\theta) \quad (2.8)$$

onde os valores s_r , s_ϕ , e s_θ são os fatores de escala na profundidade, elevação e azimute, e os valores de a são os *offsets*. A Figura 2.6 ilustra a relação espacial entre os sistemas de coordenadas esférico e cartesiano.

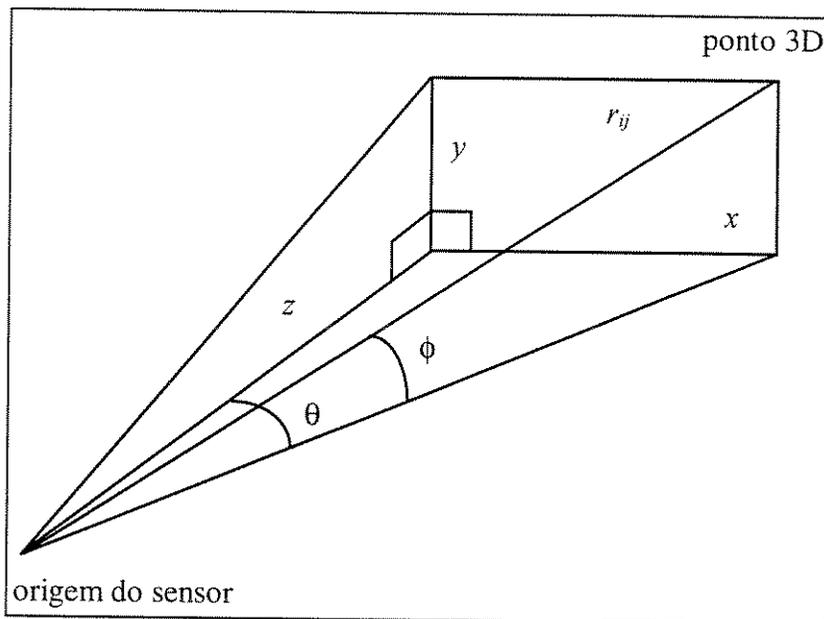


Figura 2.6. Relação entre coordenadas cartesianas e esféricas. (Besl, 1989A)

2.5. Estado da arte: Câmera 3-D arco-íris

O trabalho de Geng, 1996, apresenta um sistema, denominado câmera 3-D arco-íris (*Rainbow 3D camera*), que utiliza luz estruturada e triangulação para obter medidas de distância. O princípio da obtenção dos dados 3-D é baseado nas seguintes informações:

- (1) Um projetor de luz colorida ilumina a cena e uma imagem é tomada por uma câmera CCD colorida;
- (2) A luz projetada é codificada com a informação do ângulo de projeção correspondente, ou seja, para cada ângulo é projetada uma cor correspondente;

(3) Cada *pixel* da câmera é associado a um único raio que passa pelo ponto focal da câmera, de modo que o ângulo entre o eixo da câmera e o raio é conhecido;

(4) A distância entre a câmera e o projetor é conhecida.

A diferença básica entre o método de Geng e os similares convencionais é que cada quadro capturado pela câmera fornece um quadro completo de dados tridimensionais. A captura de um quadro completo de uma imagem 3-D (256x256, 512x512, 1024x1024 *pixels* por quadro) é dada a uma taxa de 60 quadros/s ou mais rápida. Outras características importantes da câmera 3-D arco-íris, segundo o autor, são:

(1) Não existem partes mecânicas móveis;

(2) Não existem problemas relacionados à segurança;

(3) O sistema é portátil, robusto, e de fácil instalação;

(4) São geradas duas imagens em correspondência: uma de profundidade e outra de intensidade luminosa;

(5) Seu custo é muito baixo.

2.6. Sumário

Este capítulo apresentou um conjunto de conceitos básicos para a compreensão das imagens de profundidade. A seguir, no Capítulo 3, serão apresentadas uma série de características, que podem ser extraídas a partir de imagens de profundidade, de grande aplicação no processamento destas imagens.

Capítulo 3: Extração de características

3.1. Introdução

Existe um grande número de características que podem ser extraídas, a partir de uma imagem de profundidade, de modo a estabelecer critérios de similaridade entre *pixels* e permitir a segmentação e classificação de regiões da imagem. Dentro deste contexto, é muito comum o uso de características relacionadas às propriedades das superfícies observadas na imagem. Neste caso, as características mais utilizadas são:

- (1) Medidas de curvatura (Gaussiana, Média, etc...);
- (2) Vetor normal à superfície;
- (3) Coeficientes da superfície mais adequada.

Estes valores são calculados para cada *pixel*, considerando-se uma determinada vizinhança do mesmo. É possível ainda utilizar informações de descontinuidades de superfícies, tais como bordas.

As características utilizadas num processo de segmentação devem ser adequadas à aplicação em questão (p.ex., não é interessante utilizar medidas de curvatura em cenas que contém apenas objetos poliédricos), e devem atender a dois critérios:

- (1) Ser de fácil computação;
- (2) Permitir a diferenciação entre vários tipos de superfícies.

Este capítulo apresenta conceitos relacionados às características mais utilizadas na segmentação de imagens de profundidade. Inicialmente, apresentam-se os tipos de bordas que podem ser identificadas em imagens de profundidade. Em seguida, discute-se brevemente uma metodologia para a determinação dos coeficientes de adequação para superfícies planares e de maior grau. Na sequência, definem-se medidas de curvatura para superfícies e mostra-se como a associação destas características permite a classificação de superfícies. Finalmente, algumas das características citadas são descritas em termos de derivadas parciais, apresentando-se um método computacional para calcular as mesmas a partir dos valores obtidos para estas derivadas.

3.2. Tipos de bordas

Em imagens de profundidade, é muito comum a obtenção de valores discrepantes para os *pixels* que se encontram em regiões de descontinuidades entre superfícies. Por este motivo, é importante que estes *pixels* sejam corretamente identificados. Além disso, as informações de bordas também podem contribuir para a determinação da representação ou modelo de objetos, necessários ao reconhecimento 3-D.

Numa imagem de profundidade, pode-se identificar três diferentes tipos de bordas, como mostram Hoffman e Jain, 1987, a saber:

- (1) Bordas de profundidade (*step* ou *jump edge*);
- (2) Bordas de orientação (*roof* ou *crease edge*);
- (3) Bordas suaves (*smooth edges*);

3.2.1. Bordas de profundidade

Este tipo de borda surge a partir de uma descontinuidade significativa nos valores de profundidade entre *pixels* vizinhos, como mostra a Figura 3.1(a).

Para identificar bordas de profundidade, pode-se calcular a máxima diferença de profundidade entre cada *pixel* da imagem e seus vizinhos (janela 3x3).

A magnitude deste valor é dada por:

$$M_{\text{profundidade}}(x, y) = \max \{ |z(x, y) - z(x + k, y + l)| : -1 \leq k, l \leq 1 \} \quad (3.1)$$

Se este valor é maior do que um limiar estipulado, o *pixel* em questão é identificado como de borda.

3.2.2. Bordas de orientação

Essas bordas são formadas por *pixels* onde se verifica uma variação significativa na orientação de suas normais em relação às normais de seus vizinhos, como pode ser visto na Figura 3.1(b).

As bordas de orientação são identificadas à partir do cálculo da máxima diferença angular entre normais à superfície para cada *pixel* da imagem em relação a seus vizinhos. A magnitude deste valor é dada por:

$$M_{orientação}(x, y) = \max \left\{ \cos^{-1}(\mathbf{n}(x, y) \bullet \mathbf{n}(x+k, y+l)); -1 \leq k, l \leq 1 \right\} \quad (3.2)$$

Novamente, se este valor é maior do que o limiar estipulado, o *pixel* em questão é identificado como de borda.

3.2.3. Bordas suaves

Os *pixels* componentes das bordas suaves se caracterizam por uma variação gradual na orientação de suas normais em relação às normais de seus vizinhos, como mostra a Figura 3.1(c). Sua identificação é bem mais complexa do que no caso das bordas de profundidade e orientação. São pouco utilizadas em visão computacional.

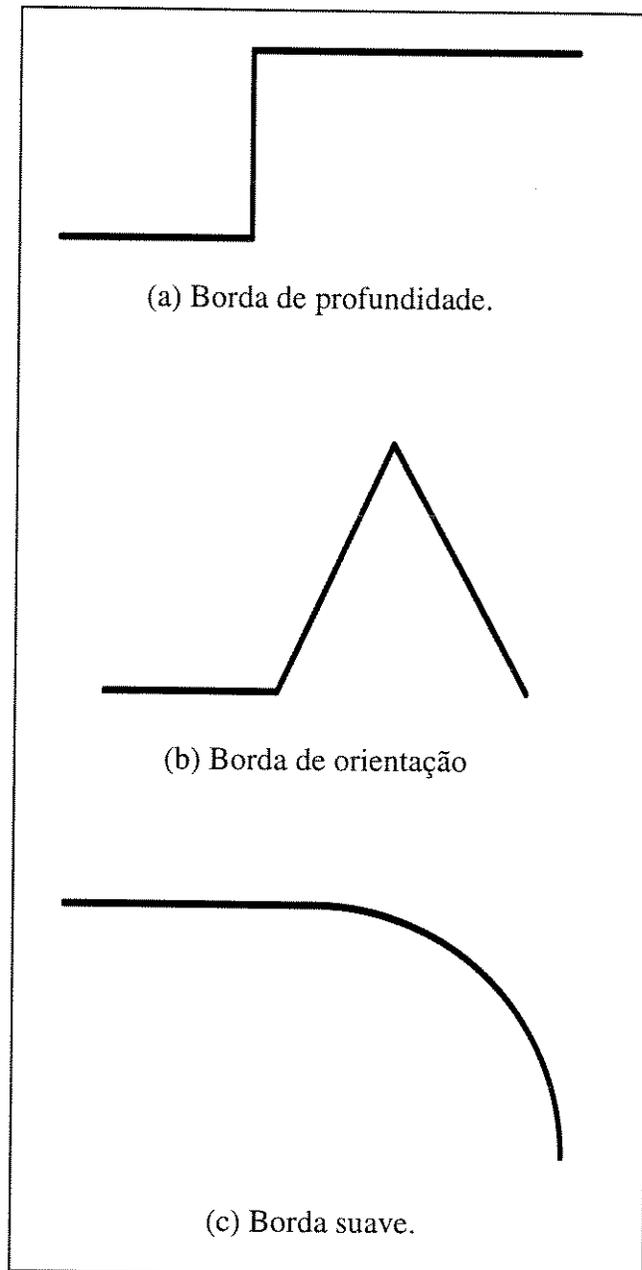


Figura 3.1: Tipos de bordas.

3.2.4. Considerações sobre valores de limiar

Os valores de limiar que vão decidir se um *pixel* pertence ou não a uma determinada borda podem ser obtidos com base, por exemplo, no conhecimento prévio sobre a cena em questão, ou nos dados (estatísticos) da própria imagem, tal como a média mais o desvio padrão do valor da magnitude de cada caso (Abdelmalek, 1990; Haddon, 1988). Entretanto, estes valores de limiar são obtidos na sua maioria por estimativas baseadas em experimentos, usando tentativa e erro (Hoover et al., 1996).

3.3. Ajuste de superfícies

Técnicas de ajuste de superfícies vêm sendo amplamente utilizadas em segmentação de imagens de profundidade. Esta seção mostra como realizar o ajuste de um plano ou de superfícies de maior grau, a partir de um conjunto de pontos de coordenadas 3-D conhecidas. Maiores informações sobre ajuste de superfícies podem ser encontradas em Lancaster e Salkauskas, 1986.

3.3.1. Superfícies planares

Para encontrar os coeficientes da equação do plano que melhor se ajusta a um conjunto de pontos, uma solução usual é aplicar o Método dos Quadrados Mínimos (Noble e Daniel, 1988). O objetivo é obter os coeficientes do plano para o

qual o somatório das distâncias de todos os pontos em relação ao plano ajustado seja mínima.

Dada a equação de um plano na forma:

$$ax + by + cz + d = 0 \quad (3.3)$$

o que se quer é obter os parâmetros do plano (a,b,c,d) dadas as coordenadas 3-D (x,y,z) de um conjunto de n pontos, $n \geq 3$.

A solução procurada pode ser obtida pela resolução do seguinte sistema de equações lineares:

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -d \\ -d \\ \vdots \\ -d \end{bmatrix} \quad (3.4)$$

que pode ser denotado por:

$$A X = B \quad (3.5)$$

Para $n = 3$, a solução é única para três pontos não-colineares. No entanto, para $n > 3$ e pontos não coplanares, devemos utilizar um artifício denominado Método da Pseudo-Inversa. Neste último caso, a solução do sistema é dada por:

$$X = (A^T A)^{-1} A^T B \quad (3.6)$$

desde que $\exists (A^T A)^{-1}$

3.3.2. Superfícies de maior grau

A idéia de se utilizar o Método dos Quadrados Mínimos para adequação de planos pode ser expandida para superfícies de maior grau. Por exemplo, para uma superfície biquadrática, a equação 3.3 deve ser modificada para:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0 \quad (3.7)$$

Embora os métodos de adequação sejam semelhantes, a solução da equação 3.7 é mais afetada pelo ruído ou erros numéricos de computação do que a equação 3.3. Isto acontece porque os valores dos coeficientes (a,b,c,d,e,f,g,h,i,j) podem ser muito pequenos, e pequenas variações dos seus sinais podem modificar completamente o tipo da superfície que está sendo ajustada.

Outras estimativas diferentes, seguindo a idéia de minimização do erro quadrático, podem ser utilizadas para se chegar aos coeficientes desejados.

3.4. Curvaturas de superfícies

As curvaturas de superfícies, a partir do trabalho de Besl, 1986, vêm sendo cada vez mais utilizadas para segmentação e classificação de superfícies em imagens de profundidade. Esta seção apresenta alguns conceitos básicos e mostra como estes são aplicados ao problema de segmentação. Maiores informações sobre curvaturas de superfícies podem ser encontradas em Arman e Aggarwal, 1993, e Do Carmo, 1976.

3.4.1. Curvaturas Principais, Média, e Gaussiana

As direções principais de uma superfície para um dado ponto são aquelas onde ocorrem os extremos da curvatura normal (curvatura máxima k_1 e curvatura mínima k_2), que são denominadas curvaturas principais (Besl, 1986). A Figura 3.2 ilustra espacialmente as curvaturas máxima e mínima.

Utilizando-se k_1 e k_2 , pode-se definir outros dois tipos de curvaturas muito utilizadas, as curvaturas Gaussiana (K) e Média (H), dadas pelo produto das curvaturas principais e pela média das curvaturas principais, respectivamente, de acordo com as Equações 3.8 e 3.9.

$$K = k_1 k_2 \quad (3.8)$$

$$H = \frac{1}{2}(k_1 + k_2) \quad (3.9)$$

Como a curvatura Média é dada pela média entre k_1 e k_2 , seu valor é menos sensível a ruídos do que o valor da curvatura Gaussiana.

As curvaturas principais e K e H são relacionadas por:

$$k^2 - 2Hk + K = 0 \quad (3.10)$$

onde k_1 e k_2 são as raízes da Equação 3.10.

Nas Tabelas 3.1 e 3.2 pode-se verificar, respectivamente, como os valores de k_1 e k_2 podem ser utilizados para classificar localmente superfícies em seis tipos, e como os valores de K e H podem classificar localmente as superfícies em oito tipos diferentes (Besl e Jain, 1986). Os tipos de superfícies obtidos a partir da combinação dos valores K e H podem ser vistos na Figura 3.3.

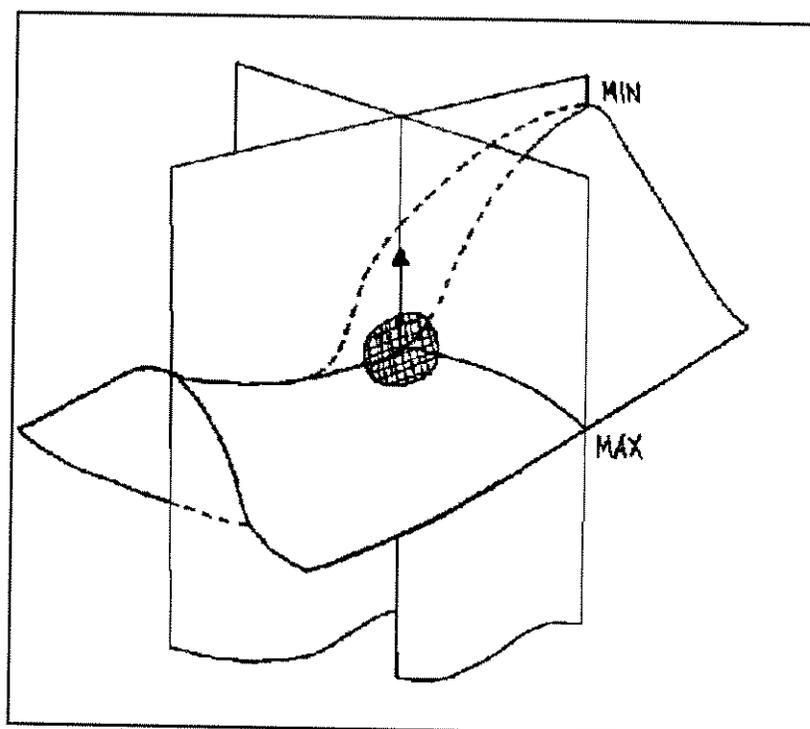


Figura 3.2: Curvaturas máxima e mínima.

**Tabela 3.1: Classificação de tipos de superfícies
utilizando curvaturas principais. (Besl, 1989)**

k_2 / k_1	$k_1 < 0$	$k_1 = 0$	$k_1 > 0$
$k_2 < 0$	peak	ridge	saddle
$k_2 = 0$	ridge	flat	valley
$k_2 > 0$	saddle	valley	pit

**Tabela 3.2: Classificação de tipos de superfícies
utilizando curvaturas K e H . (Besl, 1989)**

H / K	$K > 0$	$K = 0$	$K < 0$
$H < 0$	peak	ridge	saddle ridge
$H = 0$	(none)	flat	minimal
$H > 0$	pit	valley	saddle valley

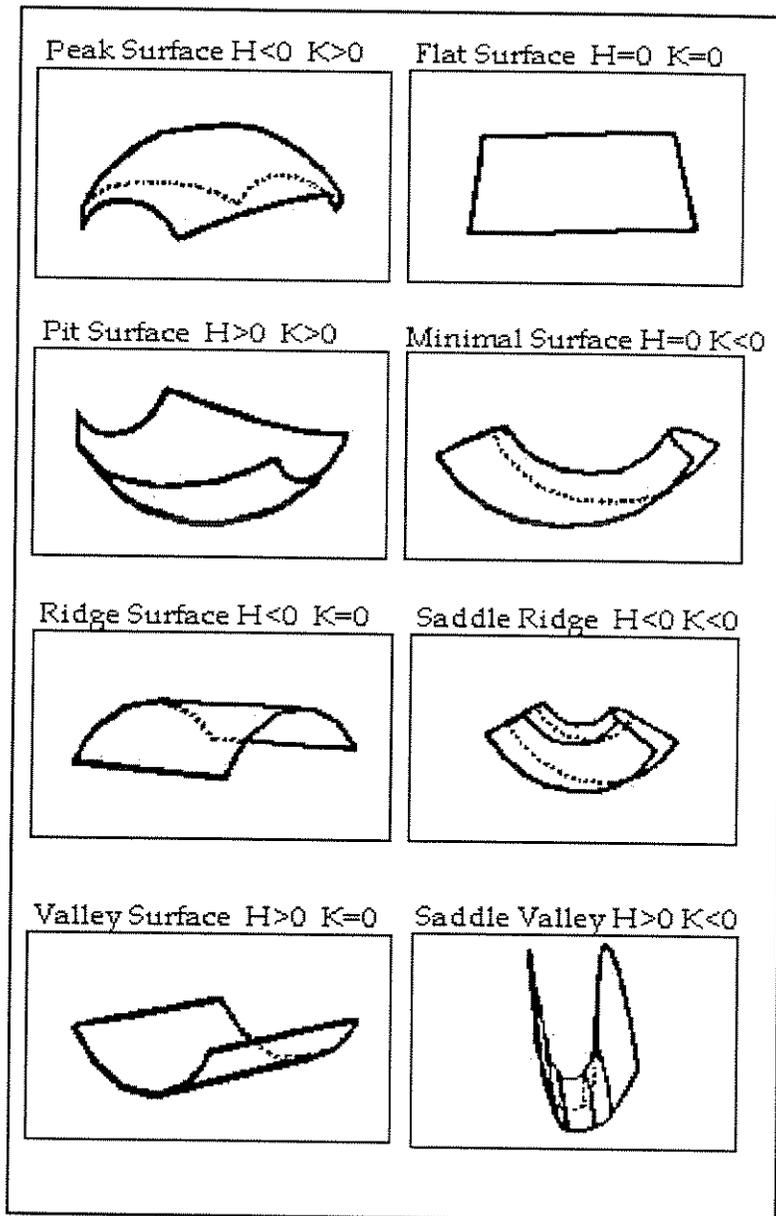


Figura 3.3: Tipos de superfícies definidas pelas curvaturas K e H . (Besl, 1986)

O par $\{k_1, k_2\}$ contém a mesma informação sobre a curvatura da superfície que o par $\{K, H\}$. No entanto, em termos de uma classificação mais refinada de superfícies numa imagem de profundidade, o uso dos sinais das curvaturas

Gaussiana e Média é mais vantajoso do que o uso dos sinais de k_1 e k_2 , pois permite a classificação de um maior número de tipos.

A classificação de superfícies a partir da análise de curvaturas não é adequada para regiões planas, pois os erros numéricos nos cálculos das curvaturas dificilmente permitirão que se obtenha valores iguais a zero, sendo necessário definir critérios de limiar para o qual as curvaturas k_1 , k_2 , K , e H são consideradas como zero. A determinação destes critérios não é tarefa trivial, principalmente em imagens ruidosas. Assim, é muito comum em técnicas de segmentação que utilizam valores de curvatura realizar-se uma identificação inicial das regiões planas por meio de outros critérios e, em seguida, utilizar as características de curvatura para obter uma classificação final.

3.5. Cálculo das características

Neste item, descreve-se de forma organizada e em termos de derivadas parciais algumas das características mais utilizadas na segmentação de imagens de profundidade, a saber:

- (1) Normal à superfície;
- (2) Ângulo de orientação da normal;
- (3) Ângulo de inclinação da normal;

(4) Curvatura Gaussiana;

(5) Curvatura média.

Em seguida, apresenta-se um método computacional simples para calcular estas derivadas parciais e, conseqüentemente, as características citadas.

3.5.1. Descrição das características

Para a descrição das características (ver Besl, 1989B, para maiores detalhes), assume-se que os pontos na superfície podem ser representados pela tripla $[u, v, f(u,v)]$, onde $f(u,v)$ corresponde ao valor de profundidade na coordenada (u,v) . Além disso, denomina-se:

f_u = derivada parcial de $f(u,v)$ em relação a u ;

f_v = derivada parcial de $f(u,v)$ em relação a v ;

f_{uu} = derivada parcial de f_u em relação a u ;

f_{vv} = derivada parcial de f_v em relação a v ;

f_{uv} = derivada parcial de f_u em relação a v ;

f_{vu} = derivada parcial de f_v em relação a u .

Com base na notação apresentada acima, as seguintes características podem ser definidas:

1. Normal à superfície

$$\bar{n}(u, v) = \frac{(-f_u, -f_v, 1)^T}{(1 + f_u^2 + f_v^2)^{1/2}} \quad (3.11)$$

2. Ângulo de orientação da normal

$$\alpha(u, v) = \tan^{-1}\left(\frac{f_u}{f_v}\right) \quad (3.12)$$

3. Ângulo de inclinação da normal

$$\theta(u, v) = \tan^{-1}\left[(f_u^2 + f_v^2)^{1/2}\right] \quad (3.13)$$

4. Curvatura Gaussiana

$$K(u, v) = \frac{f_{uu}f_{vv} - f_{uv}^2}{(1 + f_u^2 + f_v^2)^2} \quad (3.14)$$

5. Curvatura média

$$H(u, v) = \frac{f_{uu} + f_{vv} + f_{uu}f_v^2 + f_{vv}f_u^2 - 2f_u f_v f_{uv}}{2(1 + f_u^2 + f_v^2)^{3/2}} \quad (3.15)$$

3.5.2. Cálculo das derivadas

As estimativas para as derivadas parciais são computadas aplicando-se máscaras apropriadas na imagem. Uma vez computadas estas estimativas, pode-se calcular os valores das características descritas para cada *pixel* da imagem.

Para calcular as derivadas usando uma vizinhança 7x7 do *pixel* pode-se utilizar as seguintes máscaras (Besl e Jain, 1988):

$$\mathbf{D}_u = \mathbf{d}_0 \mathbf{d}_1^T$$

$$\mathbf{D}_v = \mathbf{d}_1 \mathbf{d}_0^T$$

$$\mathbf{D}_{uu} = \mathbf{d}_0 \mathbf{d}_2^T$$

$$\mathbf{D}_{vv} = \mathbf{d}_2 \mathbf{d}_0^T$$

$$\mathbf{D}_{uv} = \mathbf{d}_1 \mathbf{d}_1^T$$

onde

$$\mathbf{d}_0 = \frac{1}{7} [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]^T$$

$$\mathbf{d}_1 = \frac{1}{28} [-3 \ -2 \ -1 \ 0 \ 1 \ 2 \ 3]^T$$

$$\mathbf{d}_2 = \frac{1}{84} [5 \ 0 \ -3 \ -4 \ -3 \ 0 \ 5]^T$$

As derivadas parciais são, então, aproximadas pelas seguintes convoluções:

$$f_u(i, j) = D_u * \bar{z}(i, j)$$

$$f_v(i, j) = D_v * \bar{z}(i, j)$$

$$f_{uu}(i, j) = D_{uu} * \bar{z}(i, j)$$

$$f_{vv}(i, j) = D_{vv} * \bar{z}(i, j)$$

$$f_{uv}(i, j) = D_{uv} * \bar{z}(i, j)$$

onde $\bar{z}(i, j)$ é o valor de profundidade no ponto (i, j) , após um processo de suavização dado por um filtro Gaussiano de dimensões 7×7 (Besl e Jain, 1988).

Estas máscaras foram desenvolvidas de modo a aproximar a superfície por uma quádrlica, de acordo com os objetivos de Besl e Jain. Outras aproximações podem ser utilizadas (Barsky, 1984; Newman et al., 1993), gerando máscaras diferentes. Abdelmalek, 1990, apresenta uma breve discussão sobre o desenvolvimento deste tipo de máscaras, que são baseadas em polinômios ortogonais (Beckman, 1973).

3.6. Sumário

Este capítulo apresentou conceitos relativos às características de maior aplicação na segmentação e classificação de superfícies a partir de imagens de profundidade. Na seqüência, No Capítulo 4, será apresentada uma discussão sobre segmentação em imagens de profundidade.

Capítulo 4: Segmentação de imagens

4.1. Introdução

A segmentação é uma etapa essencial na análise de imagens. É nesta fase que são identificadas as regiões de interesse na imagem, o que torna possível sua interpretação. O processo de segmentação de imagens de profundidade segue os mesmos conceitos utilizados para imagens de intensidade luminosa. A diferença básica está na adequação das técnicas de segmentação ao tipo de informação a ser tratada.

Este capítulo apresenta, inicialmente, uma definição formal do problema de segmentação. Em seguida, é realizada uma descrição sobre cada uma das quatro principais abordagens utilizadas para segmentação de imagens com base na identificação de regiões homogêneas: análise de histograma, crescimento de regiões, divisão-e-união, e agrupamento. Na sequência, é apresentada uma breve discussão sobre a questão da super-segmentação e sobre o pré-processamento das imagens. Finalmente, discutem-se diversos trabalhos relacionados ao tema e, com base nestes trabalhos, apresenta-se um breve sumário sobre o problema da segmentação e uma tabela comparativa.

4.2. Definição formal

Gonzalez e Woods, 1992, fazem a seguinte definição formal para o problema da segmentação de imagens:

Considere R a representação da imagem como um todo. A segmentação pode ser vista como o processo que particiona R em k sub-regiões, R_1, R_2, \dots, R_k , tal que

$$(1) \cup_{i=1}^k R_i = R,$$

$$(2) R_i \text{ é uma região conexa, } i = 1, 2, \dots, k,$$

$$(3) R_i \cap R_j = \emptyset \text{ para todo } i \text{ e } j, i \neq j,$$

$$(4) P(R_i) = \text{VERDADEIRO para } i = 1, 2, \dots, n, \text{ e}$$

$$(5) P(R_i \cup R_j) = \text{FALSO para } i \neq j,$$

onde $P(R_i)$ é um predicado lógico sobre os pontos no conjunto R_i e \emptyset é o conjunto vazio.

Apesar desta ser uma definição simples, algumas inconsistências podem ser encontradas em situações práticas. Por exemplo, se duas regiões não são vizinhas, em algumas situações elas podem ter a mesma propriedade, o que viola o Item 5. Uma discussão sobre estas inconsistências é apresentada em Hoover et al., 1996.

4.3. Descrição das abordagens

4.3.1. Análise de histograma

Nesta abordagem, a idéia básica é analisar o histograma da imagem para identificar um agrupamento de *pixels* em um mesmo valor ou dentro de uma pequena faixa de variação, de acordo com um valor de limiar estipulado. A segmentação é dada então pela atribuição de um mesmo rótulo para todos os *pixels* que compõem o agrupamento, ou seja, eles irão pertencer a uma mesma região.

Para imagens de intensidade luminosa, o histograma fornece o número de *pixels* encontrados na imagem para cada intensidade, dentro da variação de intensidade luminosa característica da imagem. Já o histograma de imagens de profundidade fornece o número de *pixels* da imagem para cada nível de profundidade permitido. Um exemplo de histograma é mostrado na Figura 4.1.

Quando a imagem é composta por diversas regiões distintas, normalmente o seu histograma apresenta múltiplos picos, como mostra a Figura 4.1. Cada pico deve corresponder a uma região específica, e picos adjacentes são separados por vales que indicam um valor de limiar a ser usado para segmentação.

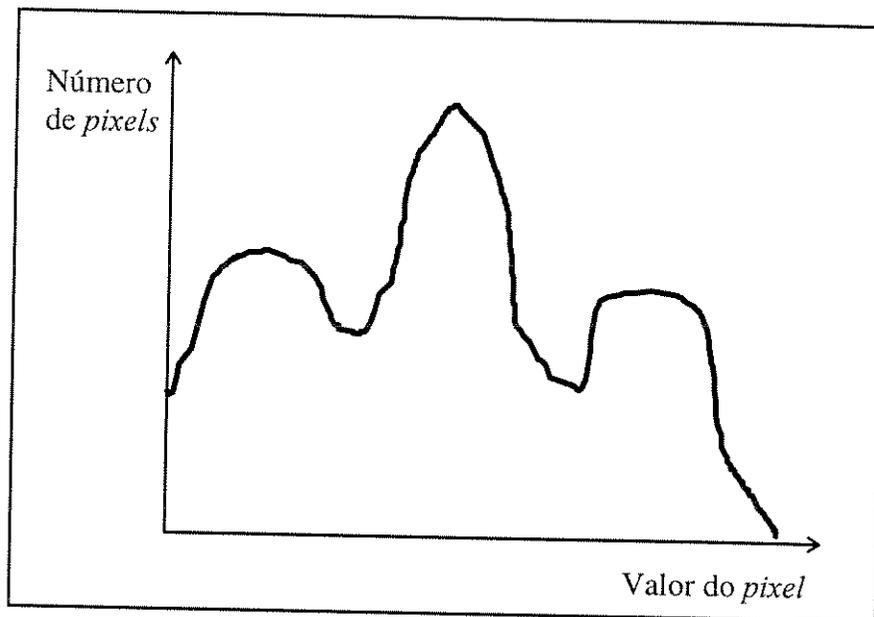


Figura 4.1: Exemplo de histograma.

Se o histograma de uma imagem tem vales profundos (significativos), a seleção de limiares pode ser uma tarefa trivial. No entanto, a situação contrária é mais provável de ser encontrada, principalmente em imagens reais. Para contornar este problema, algumas técnicas aplicam a teoria probabilística de Bayes (Gonzalez e Woods, 1992) para determinar os valores ótimos de limiar. Também são muito aplicadas as técnicas de determinação de valores de limiar dinâmicos ou locais (Gonzalez e Woods, 1992). Outras técnicas procuram levar também em consideração informações espaciais sobre a imagem, tais como proximidade e conectividade entre *pixels* (Hacket e Shah, 1989). Deste modo, fica mais fácil tratar situações em que os vales nos histogramas não são muito significativos, obtendo-se melhores resultados na segmentação. Uma boa revisão sobre técnicas de análise de histograma pode ser encontrada em Sahoo et al., 1988.

4.3.2. Crescimento de regiões

As técnicas que utilizam crescimento de regiões (Besl e Jain, 1989; Gonzalez e Woods, 1992; Haralick e Shapiro, 1992; Jain et al., 1996; Pratt, 1991) são implementadas em dois passos principais:

- (1) Extração de regiões-semente, formadas por apenas um *pixel* ou por pequenos conjuntos conexos de *pixels*, de acordo com a definição de conexidade utilizada (;Gonzalez e Woods, 1992);
- (2) A partir dos *pixels* das regiões-semente ir incorporando *pixels* vizinhos que possuam características similares, ou seja, que atendam ao critério de homogeneidade definido.

As regiões-semente devem ter a propriedade de representar da melhor maneira possível as regiões a que pertencem, visto que todas as consequentes atribuições de *pixels* à uma região vão estar relacionadas aos valores iniciais associados aos *pixels* da sua região-semente.

Uma desvantagem dos métodos que utilizam a abordagem de segmentação por crescimento de regiões é a necessidade de definir valores rígidos de limiar durante todo o processo. Uma técnica para amenizar este problema é definir os valores de limiar com base nos dados a serem tratados, ao invés de valores de limiar absolutos. Assim, é comum se definir um critério de qualidade, baseado por exemplo no ruído previsto para a imagem, e criar relações entre este critério e os valores de limiar necessários (e.g., Besl e Jain, 1988). Além deste problema, a determinação de

regiões-semente adequadas também representa um fator de dificuldade desta abordagem.

4.3.3. Divisão-e-união

A abordagem de segmentação por divisão-e-união (Faugeras e Herbert, 1987; Haralick e Shapiro, 1992; Jain et al., 1996; Pratt, 1991) consiste em duas etapas:

- (1) Na etapa de divisão, a imagem é subdividida hierarquicamente até que os *pixels* correspondentes a cada pedaço de imagem sejam homogêneos em relação a um determinado critério;
- (2) Na etapa de união, aquelas regiões adjacentes que possuem a mesma propriedade ou propriedades semelhantes são agrupadas de modo a formar uma só região.

É comum que o processo de divisão seja baseado numa estrutura tipo *quadtree*. Este método também tem como desvantagem a necessidade de valores rígidos de limiar, tanto na etapa de divisão quanto na de união. Estes valores podem ser iguais, mas é muito comum utilizar um valor de limiar diferente na etapa de união, de modo a relaxar o processo de segmentação.

4.3.4. Agrupamento

A segmentação por técnicas de agrupamento é baseada em algoritmos que, a partir de uma dada imagem, geram o particionamento da mesma através da formação de grupos de *pixels* de acordo com a proximidade entre suas características observadas. Dentre as abordagens existentes para os algoritmos de agrupamento, a saber, por minimização do erro (custo), hierárquica, e grafo-teórica (Dubes e Jain, 1976; Duda e Hart, 1973), a primeira é a mais utilizada, principalmente por ter menor custo associado e ser de simples implementação.

Dentre os algoritmos que minimizam o custo, destacam-se o *K-Means*, o *ISODATA*, e o *CLUSTER*, bem como suas versões *Fuzzy*. A idéia básica destes algoritmos é a seguinte: Dado o número K de segmentos desejados para a imagem, bem como os valores de um conjunto inicial de K pontos da imagem, escolhidos para representar os centros destes segmentos (centros de grupos), atribuir cada *pixel* da imagem ao grupo cujo centro possua valores mais próximas às suas. Os algoritmos diferem, basicamente, no modo com que os centros dos grupos iniciais são obtidos, na forma de atualização dos valores destes centros, e no critério para determinar o número ótimo de grupos. Todos os algoritmos apresentam como um problema associado o fato de que seu resultado pode convergir para um mínimo local (Selim e Ismail, 1984).

No caso de se ter uma boa estimativa sobre o número de regiões que compõem a imagem, o número K inicial de grupos é fornecido pelo operador e, ao

final do algoritmo, é criada uma imagem particionada em K regiões. O algoritmo *K-Means* toma valores de centros de grupos indicados pelo usuário, ou aleatoriamente. Uma outra abordagem (*ISODATA*) é tomar o valor médio dos pixels da imagem como o primeiro centro de grupo e o valor mais distante como sendo o segundo centro de grupo, e assim sucessivamente, até que o número de grupos indicado pelo usuário seja obtido.

Para casos onde não é conhecido o número K mais adequado de grupos, pode-se escolher um número M que represente o máximo de grupos possíveis para a imagem. Neste caso, o algoritmo (*CLUSTER*) irá gerar agrupamentos ou partições com dois até M grupos, e deve-se desenvolver um critério para definir qual partição melhor representa a cena observada.

De qualquer forma, ao final do processo, cada *pixel* pertencente a uma região possui valores tais que o tornam mais semelhante ao centro de grupo desta região do que dos outros centros, sendo que todos os centros têm seus valores inicializados no começo do processo e podem sofrer atualizações.

Como a análise de similaridade entre *pixels* é dada por meio da comparação de seus vetores de características, esta abordagem pode ter um custo computacional muito alto quando é necessário utilizar vetores de características de dimensão elevada.

A diferença mais significativa entre as técnicas de segmentação anteriormente apresentadas e a segmentação por agrupamento é que esta é muito

menos dependente de valores de limiar. Por exemplo, para se anexar um *pixel* a uma região, na abordagem de crescimento de regiões, é necessário que as propriedades deste estejam dentro de uma determinada faixa de tolerância (limiar) em relação às propriedades dos *pixels* componentes da região-semente em questão. Já para se anexar um *pixel* a um agrupamento, é necessário apenas verificar qual centro de agrupamento tem propriedades mais próximas a este *pixel*, o que independe de qualquer valor de limiar. Entretanto, a segmentação por agrupamento exige o conhecimento ou determinação do número correto de regiões da imagem.

4.4. Considerações sobre super-segmentação

É comum a aplicação das técnicas de segmentação com o objetivo de gerar imagens super-segmentadas, criando a necessidade de se analisar a semelhança entre segmentos vizinhos e proceder a união dos mesmos, se for adequado. Este interesse pela super-segmentação pode ser analisado sob dois aspectos. Primeiro, é desejável que nenhuma região distinta pertencente à imagem deixe de ser identificada, isto é, seja agrupada incorretamente à outra região. Segundo, segmentos pequenos podem ser ajustados com maior precisão, o que colabora para resultados satisfatórios numa provável reconstrução das superfícies.

O uso de super-segmentação pode ser facilmente identificado em muitos dos trabalhos discutidos na Seção 4.6. No entanto, o custo de se realizar esta super-segmentação e da posterior união de pequenos segmentos em regiões mais

significativas pode ser elevado. Em se tratando de aplicações tais como reconhecimento 3-D, pode ser mais importante identificar características significativas do objeto do que permitir sua reconstrução com pequeno erro.

4.5. Considerações sobre pré-processamento

Na maioria dos processos de segmentação de imagens de profundidade, assim como em imagens de intensidade luminosa, é necessária uma etapa de pré-processamento dos dados. Isto acontece por vários motivos. O principal deles é que as medidas fornecidas pelos sensores têm, geralmente, problemas de imprecisão devido a ruídos do detector, quantização, erros de calibração, etc, (Besl, 1989A; Jarvis, 1983A; Tamura et al., 1994). Outro motivo importante é que algumas técnicas, por exemplo as que utilizam o cálculo de curvaturas, envolvem derivadas parciais de segunda ordem, que amplificam os efeitos dos ruídos presentes na imagem. No geral, este pré-processamento consiste simplesmente em aplicar um suavizamento na imagem (Besl e Jain, 1988; Yokoya e Levine, 1989), com o objetivo de atenuar o ruído.

Neste trabalho não se discute detalhes de nenhuma técnica para pré-processamento de imagens, mencionando-se apenas o uso de algumas técnicas durante a discussão dos trabalhos sobre segmentação. Pode-se encontrar um conjunto de referências sobre técnicas de pré-processamento para tratamento de imagens de profundidade em Arman e Aggarwal, 1993.

4.6. Trabalhos relacionados

Esta seção se destina a discutir uma série de trabalhos escolhidos da literatura por estarem relacionados ao tema tratado. Cada trabalho segue pelo menos uma das abordagens de segmentação descritas na seção anterior. Alguns destes trabalhos, denominados híbridos, seguem mais de uma abordagem. Existem, também, aqueles que utilizam integração de imagens.

Devido à diversidade de possibilidades para a classificação das técnicas, e devido ao fato destas técnicas terem diferentes tipos de interseções, optou-se por não inserí-las em nenhum tipo de classificação. No entanto, no decorrer da discussão, procurar-se-á manter as técnicas que estão mais relacionadas entre si próximas umas das outras. Ao final da seção, apresenta-se uma tabela comparativa englobando os trabalhos sobre segmentação discutidos.

4.6.1. Besl&Jain

Besl e Jain, 1988, apresentam um método para segmentação e reconstrução de imagens de profundidade por crescimento de regiões, com base em informações sobre as curvaturas das superfícies observadas, e posterior ajuste de superfícies de grau variável.

A primeira etapa é gerar uma imagem cujo rótulo atribuído a cada *pixel*, ou seja, a sua classificação, é baseado numa combinação dos sinais de curvaturas Gaussiana (K) e Média (H), valores estes calculados para uma vizinhança 7×7 do

pixel. Em seguida, é realizada a segmentação da imagem classificada considerando como segmento todo conjunto conectado de *pixels* que possuam o mesmo rótulo. Cada região assim criada sofre um processo de erosão por um elemento estruturante de tamanho 3x3, até atingir um limite mínimo de *pixels* (estipulado pelos autores) que irá formar a região-semente. Desta forma, para formar a semente, procura-se utilizar os *pixels* centrais às regiões, considerando que estes tenham valores mais confiáveis por estarem livres de influências causadas pelos valores dos pixels de borda, que têm valores de K e H pouco confiáveis.

Uma vez definidas as regiões-semente, é realizado o ajuste de uma superfície aos seus *pixels* componentes. O grau desta superfície pode ser diferente para cada região-semente, sendo determinado através de um processo de hipótese e verificação. A primeira tentativa é considerar a função como um plano. Não sendo aprovada a hipótese inicial, ou seja, o conjunto de pixels da região-semente não atendendo as restrições de planaridade, procura-se utilizar uma função de maior grau, e assim sucessivamente, até atingir um limite estipulado como sendo uma superfície de quarto grau. Em seguida, o crescimento de regiões irá tentar acrescentar à cada região os *pixels* vizinhos que atendam à função ajustada para representar a superfície composta pela região-semente. Assim, obtém-se como resultado final um conjunto de regiões onde cada uma delas representa uma superfície de primeiro a quarto grau cujos parâmetros são conhecidos.

Como limitações deste método pode-se citar:

- (1) As estimativas dos valores de curvatura são muito sensíveis a ruídos;
- (2) Necessita de vários valores de limiar de difícil obtenção.
- (3) Não é garantida a separação de regiões vizinhas distintas, mas de mesmo tipo, que recebem o mesmo rótulo.

A maior importância do trabalho de Besl e Jain é seu carácter inovador na exploração de diversos aspectos dentro do problema da segmentação/reconstrução de imagens de profundidade, a saber:

- (1) Uso da geometria diferencial para classificação em oito tipos de superfícies;
- (2) Ajuste de superfícies por ordem variável.

4.6.2. Taylor et al.

Taylor et al., 1989, apresentam um método do tipo divisão-e-união para segmentação de imagens de profundidade compostas de superfícies planares. Os autores afirmam que o método pode ser estendido a superfícies de maior grau, desde que estas possam ser consideradas como compostas por pequenas regiões aproximadas por planos. Entretanto, esta extensão para superfícies não planares não é discutida no trabalho.

O critério de homogeneidade utilizado no método é baseado na similaridade dos seguintes parâmetros: os dois ângulos que definem a orientação da normal à

superfície, calculados para cada *pixel* pela adequação de um plano em uma vizinhança 3x3 do *pixel*, e o valor de profundidade associado.

A segmentação é realizada pela divisão-e-união da imagem baseada, inicialmente, apenas na similaridade dos ângulos de orientação, o que fornece rapidez ao método. Para as regiões que atendem ao critério de semelhança de orientação é realizado um teste de continuidade, verificando-se os valores de profundidade. Se o teste de continuidade falhar, a região será novamente dividida.

Embora seja um método de segmentação rápida, os resultados experimentais mostram que existe muito pouca definição nas regiões de borda da imagem, ou seja, estas regiões se apresentam muito fragmentadas.

4.6.3. Hoffman&Jain

Em Hoffman e Jain, 1987, obtém-se como resultado da segmentação de uma imagem de profundidade um conjunto de superfícies homogêneas classificadas entre planares, côncavas e convexas.

Inicialmente, a imagem é particionada em pedaços de superfícies através de um algoritmo de agrupamento, usando um vetor de características de dimensão 6, composto pelas coordenadas 3-D de cada ponto (x,y,z) da imagem e os coeficientes (a,b,c) da normal à superfície para cada ponto. No algoritmo utilizado, dado um número N de grupos, que representa o número máximo de regiões aceitável para compor a imagem, são fornecidas $N-1$ imagens particionadas em conjuntos com 2 até

N grupos. A decisão de qual conjunto representa mais corretamente a cena é baseada em critérios que irão privilegiar aqueles conjuntos cujos grupos têm menores distâncias médias entre seus *pixels* componentes (maior densidade de grupo), e cujos centros de grupos estão mais afastados (melhor distinção entre grupos).

A classificação dos segmentos obtidos é realizada com base em critérios (teste de inclinação, medidas de curvatura, e análise de auto-valores) que procuram identificar se a superfície é planar ou curva, e ainda se é côncava ou convexa. Para a reconstrução final das superfícies, analisa-se as bordas entre segmentos com a mesma classificação, verificando-se a existência ou não de bordas de orientação entre elas, de modo a se realizar a união entre segmentos adjacentes compatíveis.

Como limitação deste método está o seu alto custo computacional relacionado às etapas de:

- (1) Agrupamento com vetor de características de dimensão 6;
- (2) Testes de curvatura;
- (3) União dos segmentos.

Sua importância é dada pela alta confiabilidade dos testes para classificação.

4.6.4. Abdelmalek

No trabalho de Abdelmalek, 1990, realiza-se uma análise relativa ao erro, para os valores calculados das curvaturas Gaussiana (K) e Média (H) em imagens de profundidade, para identificação de bordas e superfícies homogêneas da mesma.

De acordo com as estimativas obtidas, concluiu-se que os valores de K são mais suscetíveis a ruídos do que os de H . Assim, a segmentação e classificação das superfícies é realizada com base apenas na medida da curvatura Média, sendo possíveis os seguintes rótulos: planar ($H=0$), côncava ($H<0$), ou convexa ($H>0$).

Como em regiões planares ou próximas a planares os valores de K e H são pouco confiáveis, devido à dificuldade de se definir valores de limiar para $K=0$ e $H=0$, após a segmentação baseada nos valores de H é realizado um teste nos pontos de borda de cada segmento. Se o ponto de borda estiver numa região que pode ser bem aproximada por um plano, este será desconsiderado para o restante do processamento. Os pontos pertencentes às bordas de profundidade são obtidos anteriormente, de modo que seus valores não influenciem no processo, causando desvios nos cálculos das curvaturas. Os resultados obtidos se mostraram eficientes na detecção de discontinuidades em imagens de profundidade.

Em termos de classificação de superfícies, este método é semelhante ao método apresentado por Hoffman e Jain, 1987. A importância deste trabalho está na metodologia desenvolvida para análise do erro nos cálculos dos valores de K e H .

4.6.5. Krishnapuram&Munshi

Krishnapuram e Munshi, 1991, apresentam uma análise do desempenho dos seguintes algoritmos de agrupamento:

- (1) *K-Means*;
- (2) *K-Means Fuzzy*;
- (3) *UFP-ONC (Unsupervised Fuzzy Partition-Optimum Number of Clusters)*.

Os algoritmos são aplicados a um conjunto variável de combinações de características extraídas para cada *pixel* da imagem de profundidade, usando conceitos de geometria diferencial, com o intuito de verificar qual combinação de características fornece melhores resultados quando utilizada pelos algoritmos de agrupamento apresentados.

As características utilizadas são:

- (1) Ângulo de orientação da normal;
- (2) Ângulo de inclinação;
- (3) Curvatura Gaussiana;
- (4) Curvatura Média;
- (5) Posição (coordenadas 3-D).

Para a segmentação de objetos poliédricos foram obtidos bons resultados aplicando o ângulo de orientação como única característica. Para objetos curvos é necessário utilizar pelo menos duas características no algoritmo de agrupamento.

Como não se trabalha com nenhuma estimativa sobre o número de regiões presentes na imagem, define-se o número máximo de grupos para a imagem, e utiliza-se uma medida de desempenho para comparação entre os diferentes conjuntos de partições obtidas pelos algoritmos de agrupamento. O critério que define esta medida de desempenho é baseado nos seguintes requisitos:

- (1) Separação clara entre grupos;
- (2) Máximo número possível de pontos próximos aos centróide dos grupos.

Segundo os autores, a combinação de características que forneceu melhores resultados, independente do algoritmo aplicado, foi aquela em que se utilizou o ângulo de orientação (constante em regiões planares) e a curvatura média (constante ou lentamente variável em regiões curvas).

No algoritmo *K-Means*, a determinação do melhor número de grupos entre o conjunto de partições obtidas não foi satisfatório, bem como a escolha dos centros de grupos corretos.

O algoritmo com que se obteve o melhor desempenho na discriminação das superfícies foi o *UFP-ONC*. No entanto, o *UFP-ONC* tende a formar grupos separados nos *pixels* pertencentes às regiões de borda, prejudicando uma boa

definição das mesmas. Assim, a não utilização de informação sobre bordas no cálculo das características prejudicou o desempenho deste método.

Segundo os autores, o tempo total de processamento do algoritmo *UFP-ONC* é muito alto, se comparado ao *K-Means*. Uma redução poderia ser realizada aumentando a amostragem no processo de agrupamento e/ou fornecendo uma boa estimativa do número de agrupamentos possíveis.

A maior limitação deste trabalho é não fornecer informações claras sobre os algoritmos usados.

4.6.6. Sabata et al.

Sabata et al., 1993, utiliza estruturas piramidais para a segmentação de imagens de profundidade num sistema de visão.

O trabalho consiste em duas etapas principais. A primeira etapa, que irá agrupar os *pixels* da imagem de acordo com a semelhança de suas propriedades locais (características), é ditada pelos dados de entrada e seu objetivo é criar uma imagem super-segmentada preliminar. A próxima etapa é dirigida pelo objetivo final do sistema de visão e deve reunir os resultados (pedaços de superfícies) da etapa anterior em regiões significativas homogêneas.

O critério de homogeneidade é função da representação de superfície adotada, que pode ser variável. Assim, a parte de mais alto nível do sistema é quem

decide qual a definição de região homogênea. Esta decisão irá ditar a estratégia de união sobre os resultados da etapa preliminar de segmentação.

Na segmentação preliminar, inicialmente, é realizada uma suavização da imagem. Em seguida, sete imagens são geradas: uma com os coeficientes (a,b,c) do vetor normal a cada ponto e seis imagens de intensidade luminosa, geradas a partir das informações 3-D e um algoritmo de *shading* para diferentes posições da fonte de luz (eixos X, -X, Y, -Y, Z, -Z). A geração de diversas imagens de intensidade luminosa tem como objetivo realçar as características que se apresentam pouco significativas em consequência da posição da fonte de luz. Na verdade, as informações criadas com esta metodologia são redundantes em sua maioria, e esta redundância é explorada pelos autores.

O próximo passo é o agrupamento piramidal para cada imagem gerada, que é composto por três estágios:

- (1) Inicialização, onde os nós da estrutura de dados piramidal são inicializados com os valores dos pixels da imagem;
- (2) Ligação de nós, um estágio iterativo onde o processo de agrupamento é realizado;
- (3) Geração da árvore, onde os *pixels* são atribuídos aos agrupamentos finais (rótulos).

Após todas as imagens serem particionadas, todas as partições são integradas numa sobreposição das mesmas, de forma que cada interseção entre duas regiões irá formar uma nova região, com o objetivo de se chegar a uma imagem final super-segmentada. Finalmente, todas as regiões assim obtidas são re-rotuladas.

A segunda etapa do processo realiza uma união dos segmentos extraídos na etapa anterior, baseado em um critério de homogeneidade que é definido de acordo com a representação final desejada para as superfícies. Duas estratégias são analisadas para a união dos segmentos adjacentes. Na primeira, para cada segmento é calculada uma função de aproximação, os segmentos adjacentes serão unidos se as suas funções atenderem ao critério de proximidade. No segundo caso, escolhe-se uma região (p.ex., a de maior área) e calcula-se sua função de aproximação que deve ser extrapolada para as regiões adjacentes; a união será realizada com aqueles segmentos cujo erro de adequação em relação à função extrapolada seja menor do que o limiar definido.

Este método de segmentação é independente da representação de superfície utilizada pelo sistema, o que é uma vantagem. No entanto, algoritmos piramidais são custosos computacionalmente em relação à memória utilizada e ao tempo de processamento.

Uma boa descrição sobre o funcionamento básico de algoritmos piramidais aplicado a segmentação de superfícies é apresentada por Grosky e Jain, 1986.

4.6.7. Bhandarkar&Siebert

Bhandarkar e Siebert, 1992, combinam técnicas de agrupamentos e detecção de bordas para segmentar imagens de profundidade contendo objetos poliédricos.

Inicialmente, é aplicado um algoritmo de agrupamento para gerar uma imagem segmentada, onde o número de segmentos desejado, fornecido pelo usuário, é adotado como o maior número possível de regiões visíveis na cena observada. Com isso, é possível ocorrer uma super-segmentação da imagem, sendo necessário realizar uma etapa posterior de união de segmentos. O critério de homogeneidade usado pelo algoritmo para unir grupos (regiões) está relacionado com a similaridade entre os parâmetros do plano associado a cada centro de grupo, a saber, dois ângulos (v, p) e uma distância d .

Uma vez obtidos os segmentos iniciais que compõem a imagem, suas bordas são identificadas não por um processo de detecção de bordas propriamente dito, mas através de um rastreamento nas direções x e y da imagem, considerando-se como borda os pontos onde a linha de rastreamento encontra um limite entre segmentos diferentes. Em seguida, realiza-se o cálculo do gradiente para cada *pixel* de borda, sendo a classificação entre borda de orientação ou borda de profundidade dada de acordo com a magnitude do gradiente.

As coordenadas dos vértices dos objetos e dos *pixels* componentes das bordas são refinadas utilizando-se informações sobre a interseção de suas superfícies adjacentes. Desta forma, é obtida uma descrição de grande precisão sobre a cena.

O método foge da classificação de superfícies por análise de curvatura, o que o torna menos suscetível a efeitos causados por ruídos, e os resultados obtidos têm grande fidelidade. No entanto, o método é aplicado apenas a objetos poliédricos, e não existe a garantia de que superfícies vizinhas com mesma orientação, mas com descontinuidade de profundidade, sejam corretamente separadas. Sua maior importância está no processo de refinamento das arestas dos poliedros resultantes.

4.6.8. Yokoya&Levine

Yokoya e Levine, 1989, apresentam uma abordagem híbrida para segmentação de imagens de profundidade, onde o resultado final é uma combinação de três mapas (imagens): um com as bordas de profundidade, outro com as bordas de orientação, e o último com pedaços de superfícies (não contendo descontinuidades de profundidade ou orientação) segmentados segundo análise das curvaturas Gaussiana (K) e Média (H).

Inicialmente, é obtido o mapa com as primitivas K e H de superfícies. O cálculo destas primitivas é realizado similarmente à metodologia apresentada na Seção 2.5, a diferença básica é que a vizinhança utilizada para realizar os cálculos é escolhida como a de menor erro quadrático, considerando-se todas as janelas que contém o *pixel* em questão. Este procedimento funciona como um filtro que irá suavizar a imagem. O agrupamento de *pixels* conectados com mesmo sinal no mapa $K-H$ irá gerar as primitivas de superfícies desejadas.

Os mapas com as bordas de profundidade e de orientação são obtidos analisando-se as diferenças entre os valores de profundidade e ângulo entre normais, respectivamente, na vizinhança de cada *pixel*. A segmentação final é obtida sobrepondo-se os três mapas e realizando-se o crescimento de cada região segmentada paralelamente, de modo que as regiões expandidas não ultrapassem os limites de borda anteriormente definidos.

A integração das abordagens permite uma boa descrição da imagem, devido à identificação de bordas e de superfícies. Além disso, a informação de borda soluciona o problema de regiões vizinhas com mesmo rótulo, observado no trabalho de Besl e Jain, 1988. A maior contribuição deste trabalho está na metodologia desenvolvida para seleção da melhor janela a ser utilizada, o que permite uma redução do erro nos cálculos das curvaturas K e H .

4.6.9. Gil et al

Gil et al., 1983, combina o mapa de bordas de duas imagens, uma de intensidade luminosa e outra de profundidade, de uma mesma cena para gerar uma imagem (mapa) de bordas final.

Inicialmente, é obtido o mapa de bordas para a imagem de intensidade luminosa pelo cálculo do gradiente segundo o operador de Kirsch (Gonzalez e Woods, 1992). Em seguida, é realizado um afinamento nas bordas deste mapa, considerando-se como ponto de borda aqueles pixels onde o valor do gradiente é um máximo local.

Para determinar o mapa de bordas da imagem de profundidade analisa-se os valores dos ângulos de curvatura, segundo uma determinada vizinhança, para cada *pixel* da imagem, devendo-se tomar cuidado nas regiões de borda de profundidade. A combinação dos dois mapas é dada por operações AND entre seus *pixels*, onde as informações de bordas do mapa da imagem de intensidade luminosa são confirmadas pelas informações de bordas do mapa de profundidade.

Algumas restrições sobre o tamanho mínimo dos segmentos têm como objetivo reduzir os efeitos causados por ruídos. Esta combinação de dados se mostra efetiva no sentido de que, uma vez que as duas imagens são afetadas de maneiras diferentes, a combinação de seus resultados irá permitir o preenchimento de falhas de informações. No entanto, não são definidas regiões mas sim segmentos de retas sem conexão, onde não se realiza nenhuma análise sobre as formas das superfícies e a conectividade dos segmentos.

4.6.10. Yu et al.

O trabalho de Yu et al., 1994 apresenta um método de segmentação e reconstrução aplicado a imagens de profundidade. O método realiza segmentação e adequação de superfícies de uma forma robusta, com uma alta tolerância a *outliers*, segundo os dados experimentais apresentados.

O ajuste de superfícies a um conjunto de pontos é realizada através da técnica denominada *Residual Consensus* (RESC), que é baseada no histograma dos resíduos dos ajustes obtidos.

A questão principal é a seguinte: Dado que os parâmetros de uma superfície podem ser determinados por p pontos (3 para um plano, 9 para uma superfície quádrlica), como escolher estes pontos, dentre o conjunto total de pontos que compõem a superfície, de modo a se obter os parâmetros da superfície mais adequada à região?

Para solucionar esta questão, Yu et al. usam técnicas de amostragem aleatória para tomar amostras de pontos. Em seguida, é feita a análise do resíduo para cada amostra, ou seja, é calculada a distância entre cada ponto, dentre todos da superfície, e a função da superfície adequada.

O algoritmo RESC consiste nos seguintes passos principais:

1. *Obter aleatoriamente K conjuntos de p pontos ($p = 3$ ou 9) do espaço de amostragem S .*

2. *Para cada conjunto de pontos, calcular o resíduo em relação aos dados originais usando as primitivas determinadas por estes pontos, e fazer o histograma dos resíduos:*

$$r_i = z_i - z$$

onde r_i é o resíduo no ponto i , z_i é o valor de profundidade no ponto i , e z é o valor de profundidade calculado a partir da equação da superfície.

3. *Para os K conjuntos, selecionar aquele cujo histograma (n^o de pontos versus resíduo) apresenta maior consenso residual, ou seja, exista um maior número de pontos com valores baixos de resíduo.*

4. *Determinar pelo histograma o desvio padrão σ dos resíduos, que é o nível de ruído da região cuja superfície foi adequada.*

5. *Rotular os pontos da superfície e removê-los de S , de modo que estes não sejam incluídos em processamentos posteriores.*

6. *Remover os outliers da região rotulada.*

7. *Repetir passos 1-6 até que $S = \emptyset$.*

A maior contribuição deste trabalho é permitir bons resultados de segmentação e adequação mesmo em imagens altamente ruidosas.

4.7. Comentários

As imagens de profundidade fornecem dados suficientes para a descrição precisa de uma cena. No entanto, estas informações devem ser corretamente segmentadas em elementos significativos da cena. Por este motivo, pode ser conveniente utilizar técnicas para segmentação que procurem super-segmentar a imagem garantindo, assim, que nenhuma região importante deixará de ser identificada e que as superfícies poderão ser reconstruídas com qualidade. Após esta super-segmentação, entretanto, deve-se realizar uma etapa de verificação da existência ou não de uma borda entre segmentos adjacentes para sua possível união. Esta etapa, no entanto, costuma ter um custo computacional elevado.

Nos processos de segmentação que utilizam análise de histograma os resultados são altamente dependentes de uma boa filtragem inicial das imagens. Além disso, não se obtém nenhuma descrição sobre a composição das regiões segmentadas. No entanto, para imagens onde existe uma boa distinção entre suas regiões componentes, esta técnica fornece bons subsídios para a definição do número de regiões no qual a imagem deve ser particionada. Esta característica pode ser de grande valia para a associação de outras técnicas, tais como crescimento de regiões e agrupamentos.

A abordagem de detecção de bordas permite criar uma boa representação aramada (*wire frame*) dos objetos contidos na imagem, desde que seja realizado um processamento suplementar (*edge closing*) para conectar os pedaços de segmentos extraídos, o que pode não ser trivial. A detecção de bordas isoladamente não oferece nenhuma informação sobre as superfícies dos objetos, dificultando a sua descrição em mais alto nível. Entretanto, este tipo de técnica apresenta bom desempenho para cenas contendo apenas objetos poliédricos.

A segmentação por crescimento de regiões permite uma boa descrição das superfícies dos objetos. Porém, esta abordagem é muito dependente de valores de limiar. Além disso, existe a dificuldade associada à extração de regiões-semente confiáveis.

A utilização de técnicas de agrupamento para segmentação de imagens de profundidade tem se mostrado muito interessante, pois é menos dependente em relação a valores de limiar. Esta característica é bastante conveniente quando se tem como objetivo desenvolver sistemas de visão mais genéricos. Dentre o conjunto de aplicações analisadas para esta abordagem observou-se que, em muitos casos, é desejável uma boa estimativa inicial do número de agrupamentos da imagem e da localização mais adequada para os centros de agrupamentos a serem escolhidos. Estas estimativas iniciais seriam utilizadas para reduzir o tempo de processamento necessário para determiná-las, bem como evitar que as mesmas sejam determinadas incorretamente. Assim, o desenvolvimento de um método adequado para determinar

estas estimativas, certamente, irá fornecer muito mais qualidade aos resultados obtidos pelas técnicas de segmentação por agrupamento.

Os trabalhos discutidos na Seção 4.6 mostram um grande elenco de técnicas usadas para solucionar o problema da segmentação. Embora seja difícil compará-los, uma classificação relacionada ao tipo de segmento obtido pode ser realizada. Basicamente, os segmentos resultantes podem ser dos seguintes tipos:

- (1) Baseados em sinais K e H ;
- (2) Planares;
- (3) Planares, côncavos, e convexos;
- (4) Superfícies de maior grau.

Na segmentação baseada em sinais K e H , pode-se fazer uma associação direta com 8 tipos de superfícies. Essa abordagem, entretanto, é muito sensível a ruídos na imagem, e existe grande dificuldade na determinação de valores de limiar para regiões planares.

A segmentação em regiões planares é de fácil computação, baseada nos valores das normais, além de que pedaços de regiões planares podem servir como base para um posterior processo de união em superfícies de maior grau.

Na segmentação em regiões planares, côncavas, e convexas, é necessário um maior nível de complexidade na obtenção do que para apenas regiões planares. Neste

caso, pode existir muito pouca compensação em relação ao aumento do custo computacional.

A segmentação em superfícies de maior grau produz resultados altamente discriminatórios. Entretanto, o custo computacional para sua obtenção mais elevado.

Algumas questões sobre segmentação de imagens de profundidade podem ainda ser levantadas:

- (1) Valores de limiar: A obtenção dos valores de limiar é uma etapa crítica dentro da segmentação de imagens, e muito ainda pode ser feito no sentido de tornar os métodos mais independentes deste problema, ou de se produzir meios automáticos mais robustos para sua solução.
- (2) Medidas de erro: É muito importante construir medidas de erro para o processo de segmentação que sejam realmente representativas da qualidade obtida. Muitas métodos não dão a esta questão a devida importância.
- (3) Robustez: A sensibilidade a ruídos sempre vai existir, de modo que é importante o desenvolvimento de metodologias robustas.
- (4) Integração segmentação/representação: Embora a representação do conteúdo das imagens esteja diretamente ligada à qualidade e ao tipo do processo de segmentação utilizado, esta integração tem sido pouco contemplada na literatura. Uma melhor associação dos dois métodos

pode gerar resultados adequados a tarefas tais como reconhecimento 3-D de objetos.

A Tabela 4.1 mostra um resumo sobre dos trabalhos descritos na Seção 4.6.

Tabela 4.1: Resumo dos métodos de segmentação

Fonte	Técnicas utilizadas	Tipos de segmentos resultantes	Principal contribuição	Principal limitação
Besl&Jain	mapa KH, crescimento de regiões, ajuste de quádricas	oito tipos baseados nos sinais K e H, quádricas	uso inovador de geometria diferencial e adequação de superfícies por ordem variável	exige valores de limiar rígidos
Taylor et al.	divisão-e-união	segmentos planares	rapidez	fragmentação nas bordas
Hoffman&Jain	agrupamento, testes estatísticos	segmentos planares, côncavos, e convexos	confiabilidade dos testes de planaridade e curvatura	agrupamento de dimensão-6
Abdelmalek	mapa KH	oito tipos baseados nos sinais K e H, segmentos planares e curvos	análise do erro nos cálculos das curvaturas K e H	alta sensibilidade a ruídos
Krishnapuram& Munshi	agrupamento, características de geometria diferencial	segmentos planares e curvos	estudo sobre combinação de características	ausência de detalhes na descrição dos algoritmos avaliados
Sabata et al.	agrupamento piramidal	variável	adequação de superfícies variável	alto custo computacional
Bhandarkar& Siebert	agrupamento, detecção de bordas	superfícies planares	refinamento de arestas	aplicado apenas a objetos poliédricos
Yokoya&Levine	mapa KH, detecção de bordas	oito tipos baseados nos sinais K e H	seleção de janelas com menor erro	alta sensibilidade a ruídos
Gil et al.	detecção de bordas	indefinido	integração de imagens de intensidade luminosa e profundidade	não fornece informações sobre as superfícies segmentadas
Yu et al.	estimação robusta, ajuste de quádricas	quádricas	boa resposta inclusive em imagens ruidosas	exige pré-segmentação da imagem

4.8. Sumário

Este capítulo descreveu o problema de segmentação em imagens de profundidade. No Capítulo 5, a seguir, é apresentado o método de segmentação desenvolvido, que associa as técnicas de agrupamento e detecção de bordas.

Capítulo 5: O método de segmentação

5.1. Introdução

Este capítulo apresenta o método de segmentação desenvolvido. O objetivo principal é, dada uma imagem de profundidade, segmentá-la em regiões bem aproximadas por planos, de modo que as suas superfícies componentes possam ser reconstruídas com pequeno erro. Para isso, aplica-se a técnica de agrupamento associada a técnicas de detecção de bordas, onde a informação utilizada como característica no algoritmo de agrupamento implementado é o vetor normal a cada ponto da imagem, que é constante em regiões planares. A utilização dos resultados da detecção de bordas permite solucionar as seguintes limitações da técnica de agrupamento aplicada:

- (1) Identificar uma estimativa inicial significativa para o número de regiões da imagem.
- (2) Separar regiões vizinhas de mesma orientação mas com descontinuidade de profundidade.

A informação sobre bordas também traz como benefício a possibilidade de eliminar efeitos espúrios encontrados nos *pixels* de bordas. Estes efeitos são devidos, principalmente, a problemas de aquisição e inconsistências nos cálculos realizados usando a vizinhança de tais *pixels*, pois os valores desta vizinhança são

discrepantes. Além disso, a informação geométrica sobre as bordas pode auxiliar na construção de modelos para reconhecimento 3-D de objetos.

As principais contribuições do método de segmentação como um todo são:

- (1) A determinação de uma estimativa inicial para o número K de regiões, a ser fornecido para o algoritmo de agrupamento, com base nos dados da própria imagem.
- (2) A utilização de uma nova abordagem para análise do resultado do agrupamento e determinação do número ótimo de grupos, com base no erro de reconstrução obtido para a imagem segmentada.

Inicialmente, na Seção 5.2, apresenta-se a definição de segmentação/reconstrução utilizada (Besl e Jain, 1988). Na Seção 5.3, apresenta-se um breve resumo sobre o método desenvolvido. As seções 5.4-5.11 descrevem as diversas etapas do método. Finalizando o capítulo, a Seção 5.12 faz uma breve comparação do método desenvolvido em relação aos discutidos no capítulo anterior, tomando como base a Tabela 4.1.

5.2. Definição de segmentação/reconstrução

A segmentação de imagens pode ser entendida como o processo de particionar uma imagem em segmentos significativos que correspondam a objetos,

partes de objetos, ou grupos de objetos da cena observada. Neste contexto, a reconstrução pode ser definida como o processo de obtenção de uma descrição da imagem com base em seus componentes identificados.

O problema da segmentação/reconstrução de imagens de profundidade é definido por Besl e Jain, 1988, como:

Definição 5.1: "Dada uma superfície digital, denominada $\tilde{g}(i, j)$ e especificada por seus valores z_{ij} , encontrar K funções de aproximação $\hat{f}_l(x, y)$ e K regiões \hat{R}_l da imagem sobre a qual estas funções são avaliadas, de modo que o erro total obtido

$$\varepsilon_{tot} = \|\tilde{g}(i, j) - \hat{g}(x(i), y(j))\|_I \quad (5.1)$$

entre a imagem reconstruída

$$\hat{g}(x, y) = \sum_{l=1}^K \hat{f}_l(x, y) \chi(x, y, \hat{R}_l), \quad (5.2)$$

avaliada nos pontos $(x(i), y(j))$ e os dados $\tilde{g}(i, j)$ é pequeno, e o número total de funções e regiões K também é pequeno, onde χ é uma função característica que assume valor igual a 1 se o ponto (x, y) pertence à região \hat{R}_l e igual a zero caso contrário".

De acordo com a Definição 5.1, pode-se observar que é desejável não apenas minimizar o erro de reconstrução das superfícies, mas também minimizar o

número de regiões resultantes da segmentação. Estes dois critérios são contraditórios e, por este motivo, é preciso encontrar uma solução de compromisso entre eles.

O método de segmentação desenvolvido baseia-se na definição apresentada por Besl e Jain, 1988, e utiliza os valores de reconstrução obtidos como medida de qualidade para o processo de segmentação.

5.3. Descrição da imagem

Neste trabalho, a imagem a ser segmentada está na forma r_{ij} , ou seja, consiste numa matriz onde cada *pixel* (i,j) contém um determinado valor r de profundidade. As coordenadas xyz de cada *pixel* da imagem são conhecidas, de acordo com a Seção 2.4.

5.4. Descrição sumária do método

Inicialmente, realiza-se na imagem de profundidade um processo de deteção de bordas de profundidade por um operador de borda padrão (p.ex., Sobel). Em seguida, calcula-se a normal a cada *pixel* p da imagem, usando as coordenadas (x,y,z) dos *pixels* vizinhos pertencentes a uma janela de dimensão 3x3 centrada no *pixel* p , e desprezando-se o efeito dos *pixels* pertencentes às bordas de profundidade já detectadas. De posse destas normais, identifica-se os *pixels* de

borda de orientação, que são aqueles onde existe uma variação significativa na orientação de sua normal em relação às normais de seus *pixels* vizinhos.

Com as informações de bordas de profundidade e bordas de orientação, gera-se o mapa de bordas da imagem. Este mapa sofre um processo de fechamento de bordas (*edge closing*) de modo que possa ser analisado com a finalidade de identificar o número de regiões fechadas que o compõem. Após desprezadas as regiões do mapa de bordas compostas por poucos *pixels*, ajusta-se um plano a cada região conexa. O erro de ajuste de planos é avaliado para cada região e, se for maior do que o limiar de planaridade estipulado, o número de regiões conexas obtido deve ser incrementado. Ao final da avaliação dos erros de ajuste planar, o número de regiões obtidas é fornecido ao algoritmo de agrupamento, como uma estimativa inicial do número de regiões K que compõem a imagem.

Uma vez gerado o mapa de bordas, calcula-se o erro de aproximação para as normais obtidas para cada *pixel* da imagem e o erro global de aproximação para toda a imagem, que representa a média dos erros de aproximação, desconsiderando-se as contribuições referentes tanto aos *pixels* de borda de profundidade quanto aos de bordas de orientação. Este erro de aproximação nos valores calculados para as normais é utilizado na determinação da medida de qualidade da segmentação obtida.

Neste ponto, o algoritmo de agrupamento é aplicado, utilizando-se como características os coeficientes (a,b,c) do vetor normal à superfície para cada *pixel*.

Depois da convergência do algoritmo, o resultado obtido é uma imagem composta por K regiões. As informações sobre o mapa de bordas são, então, utilizadas para separar regiões vizinhas de mesma orientação, que foram incorretamente agrupadas em uma mesma região. Em seguida, calcula-se os coeficientes da equação do plano mais adequado para cada região identificada. Com esta informação, pode-se gerar uma imagem reconstruída e avaliar o erro de reconstrução, com base no erro de aproximação de cada segmento. Assim, o número inicial de grupos pode ser aumentado ou diminuído, em função do erro de reconstrução obtido e do número de *pixels* que compõem cada região. As principais etapas do método desenvolvido são descritas detalhadamente a seguir.

5.5. Detecção de bordas e cálculo das normais

Dois diferentes tipos de bordas são identificadas: bordas de profundidade e bordas de orientação. Para identificar os *pixels* componentes das bordas de profundidade, pode-se utilizar um operador padrão para cálculo do gradiente em imagens de intensidade luminosa (Gonzalez e Woods, 1992; Pratt, 1991). No caso, o operador utilizado foi o DRF (*Direct Recursive Filter*), desenvolvido por Shen e Castan, 1986, por apresentar melhores resultados dentre os analisados (DRF, Sobel, Deriche). Após o cálculo do gradiente, o valor T_1 utilizado para o limiar que identifica se um *pixel* pertence ou não a uma borda de profundidade é dado com

base no valor médio T_m e no desvio padrão S dos valores encontrados para o gradiente da imagem, segundo a seguinte equação:

$$T_1 = T_m + 1.5 S \quad (5.3)$$

A Equação 5.3 foi utilizada com base na sugestão observada no trabalho de Haddon, 1988, que faz um estudo sobre determinação de valores de limiar, gerando resultados satisfatórios.

As bordas de orientação exigem um processo mais complexo para sua identificação. No caso, a abordagem usada foi calcular os coeficientes da normal a cada *pixel* e, a partir destes dados, identificar as bordas de orientação pela variação dos ângulos de orientação entre normais vizinhas.

A normal (a,b,c) a cada *pixel* (i,j) é calculada pela adequação de um plano, utilizando-se o Método dos Mínimos Quadrados (Noble e Daniel, 1988), em uma janela 3x3 centrada no *pixel* (i,j) , como mostra a Figura 5.1.

	$i-1,j$	
$i,j-1$	i,j	$i,j+1$
	$i+1,j$	

Figura 5.1: Janela 3X3 centrada no *pixel* (i,j) .

No cálculo da normal, despreza-se o efeito dos *pixels* de borda de profundidade incluídos na janela, de forma a reduzir efeitos espúrios. Os coeficientes da normal aos *pixels* de borda de profundidade não são calculados, mas sim recebem os valores dos últimos coeficientes anteriormente calculados. Com esta medida, evita-se obter coeficientes discrepantes para a normal à superfície nestes *pixels*.

O ângulo que mede a variação da normal entre um *pixel* e seus vizinhos é dado pelo maior valor de variação de sua normal em relação às de seus vizinhos. Estes valores são calculados através do produto escalar dos vetores normais, conforme é descrito na Seção 2.3. Quando este ângulo for maior do que um limiar T_2 , estipulado como sendo 10° empiricamente, considera-se que este *pixel* faz parte de uma borda de orientação.

5.6. Cálculo da estimativa do número de regiões

5.6.1. Construção do mapa de bordas

Uma vez que as bordas de profundidade e orientação foram detectadas, deve-se criar um mapa de bordas, a partir do qual será estimado o número inicial K de regiões que é fornecido ao programa de agrupamento. Para isso, o primeiro passo é reunir as bordas de profundidade e de orientação numa única imagem.

Devido a problemas inerentes aos métodos de detecção de bordas, tais como sensibilidade a ruídos e uso de valores de limiar incorretos na identificação dos *pixels* que compõem as bordas, é possível que o resultado da união das mesmas não seja um conjunto de regiões fechadas. Então, esta imagem resultante deve sofrer um processo de fechamento de bordas, de modo a gerar um conjunto de regiões fechadas.

Existem vários métodos para fechamento de bordas, encontrados na literatura. Os algoritmos utilizados são, geralmente, baseados na direção e magnitude das informações de gradiente usadas para construir as bordas. Os métodos mais comuns seguem a abordagem por processamento local ou processamento global (Gonzalez e Woods, 1992; Haralick e Shapiro, 1992; Pratt, 1991).

No método por processamento local, é realizada a análise da vizinhança dos *pixels* que estão em aberto na imagem de bordas, de modo a identificar aqueles vizinhos que têm valores de direção e magnitude de gradiente semelhantes aos dos *pixels* em aberto. Os *pixels* vizinhos que atenderem ao critério de similaridade definido são incorporados às bordas.

Os métodos globais são aqueles baseados, por exemplo, em informações obtidas a partir da aplicação da transformada de Hough (Illingworth e Kittler, 1988). Estas informações funcionam como uma máscara, que deve ser sobreposta aos *pixels* identificados originariamente como pertencentes a uma borda. Com esta

medida, é possível identificar quais *pixels* da imagem de gradiente devem ser utilizados para fechar as bordas. Esta abordagem funciona tanto para bordas retas quanto curvas, embora neste último caso o processo se torne muito mais complexo.

Independente do método utilizado, o maior problema associado ao fechamento de bordas é que este processo pode gerar bordas espúrias, o que acarreta a criação de pequenas regiões fechadas não representativas da realidade dos objetos observados na imagem. Na aplicação em questão, este problema é amenizado no processo de contagem do número de regiões.

5.6.2. Contagem do número de regiões

Uma vez que o mapa de bordas esteja construído adequadamente, o próximo passo é realizar a contagem do número de regiões fechadas que o compõem. Existem várias maneiras para realizar esta contagem. Jain et al, 1996, apresentam dois métodos, um recursivo e outro seqüencial, para fazer o rotulamento de regiões fechadas e, com isso, obter o número de regiões da imagem. O algoritmo utilizado neste trabalho é baseado nestes métodos, mas foi adaptado para o problema específico em questão, que é a identificação do número de regiões fechadas a partir de um mapa de bordas binário. A idéia básica do algoritmo é a seguinte:

1. Inicializar $K = 2$; (0 é fundo, 1 é borda)
2. Rastrear a imagem, até encontrar um pixel p com valor igual a zero;
3. Atribuir rótulo = K para o pixel p ;
4. Crescer a região, a partir do pixel p obtido no passo 2., agregando pixels vizinhos com valor igual a zero, e atribuindo a eles rótulo = K , até encontrar uma borda;
5. Incrementar K em uma unidade;
6. Repetir os passos 2. a 5., até não restarem mais pixels com valor igual a zero;
7. Avaliar a quantidade de pixels de cada região, se for menor do que um determinado limiar, definido em termos de uma porcentagem da imagem, decrementar em uma unidade o número de regiões;
8. Decrementar K em uma unidade.

O número de regiões final obtido pelo algoritmo é, então, repassado ao processo de agrupamento como sendo o número K inicial de regiões. Como essa estimativa pode sofrer ajustes, com base no erro de reconstrução encontrado, não é necessário um processo muito refinado para sua obtenção.

5.6.3. Objetos poliédricos x não-poliédricos

Para objetos poliédricos, o processo de contagem de regiões a partir do mapa de bordas fornece uma indicação significativa sobre o número de superfícies que compõem a imagem. Entretanto, para objetos não poliédricos, esta indicação por si só não é satisfatória, pois devemos dividir cada superfície em pedaços de regiões bem aproximadas por planos. Neste caso, não existe um número fixo

correto de regiões para a imagem, mas sim um número ótimo para o qual o erro de reconstrução está dentro do limite de tolerância desejado.

A necessidade de aumentar o número K de regiões, a ser fornecido ao algoritmo de agrupamento, é identificada pela simples análise do erro de ajuste de um plano a cada região delimitada pelo mapa de bordas. O valor de limiar a ser utilizado para decidir se uma superfície é planar ou não pode ser dado por um processo de calibração, utilizando-se a imagem de uma única superfície planar e obtendo-se o erro de ajuste para a mesma.

Os valores de erro encontrados indicam se o número K , obtido pela contagem de regiões, deve ser incrementado antes de ser utilizado pelo algoritmo de agrupamento. Não foi desenvolvida nenhuma metodologia para avaliar o quanto este número K deve ser aumentado, em função do erro obtido, visto ser este problema altamente dependente das características do conjunto de imagens utilizadas. Como alternativa, optou-se por incrementar o número K em duas unidades cada vez que uma região do mapa de bordas é identificada como não planar.

Apesar de todos os esforços para determinar uma estimativa K inicial mais próxima possível da realidade, o valor ideal final de regiões só é confirmado após a aplicação do algoritmo de agrupamento, com a avaliação do erro de reconstrução total, tomando como base a segmentação obtida. Se o erro de reconstrução está acima do desejado, aumenta-se o número K de regiões utilizado inicialmente e

reaplica-se o processo de agrupamento, até se obter um erro de reconstrução dentro do limite desejado.

5.7. Cálculo do erro de aproximação

O cálculo da normal à superfície para cada *pixel* p é realizado pela adequação de um plano aos *pixels* contidos numa janela 3x3 centrada no *pixel* p . Este cálculo pode conter um erro devido à aproximação da vizinhança do *pixel* por um plano. Este erro, denominado aqui erro de aproximação (ε_a), pode ser estimado para cada *pixel* p da imagem, considerando-se uma vizinhança v , pela seguinte equação:

$$\varepsilon_a^2(p) = \frac{1}{n_c} \sum_{i \in v} (d_p - a_p x_i - b_p y_i - c_p z_i)^2 \quad (5.4)$$

onde a_p, b_p, c_p, d_p são os coeficientes do plano que passa pelo *pixel* p , (x_i, y_i, z_i) são as coordenadas 3-D dos *pixels* que compõem a vizinhança do *pixel* p , e n_c é o número de *pixels* que contribuem efetivamente para o cálculo da normal no ponto p , desconsiderando-se os *pixels* de borda de profundidade já detectados.

O coeficiente d_p da equação do plano para cada região é calculado como sendo a média dos valores calculados para d_i para cada *pixel* i da região, segundo a seguinte equação:

$$d_p = \frac{1}{n} \sum_{i=1}^n d_i \quad (5.5)$$

onde n é o número de *pixels* da região.

O erro de aproximação global da imagem (ε_{ga}) é a média dos erros de aproximação de cada *pixel*, desconsiderando-se os *pixels* de borda, e é dado por:

$$\varepsilon_{ga} = \frac{1}{n - n_b} \sum \varepsilon_a^2(p) \quad (5.6)$$

onde n é o número total de *pixels* da imagem, e n_b é o número de *pixels* de borda.

As equações 5.4 e 5.6 são baseadas em equações similares apresentadas em Besl e Jain, 1988.

O erro de aproximação vai servir como parâmetro para a medida de qualidade dos processos seguintes, pois as componentes das normais, utilizadas como características de entrada, carregam este erro para a etapa de agrupamento. Como este erro é diretamente proporcional à quantidade de ruído na imagem, em imagens ruidosas o erro de aproximação deve ser maior do que o obtido para uma imagem sem ruído, e o erro de reconstrução obtido neste caso não pode ser atribuído somente ao processo de segmentação, mas também à qualidade da imagem.

5.8. Agrupamento

O algoritmo de agrupamento utilizado é baseado na minimização do erro quadrático das distâncias entre os elementos componentes de cada grupo (Dubes e Jain, 1976). O conceito básico para este tipo de algoritmo é descrito a seguir.

Seja o i -ésimo componente, $i=1, \dots, n$, de um conjunto de dados sob estudo

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iN})^T. \quad (5.7)$$

e $(x_{i1}, x_{i2}, \dots, x_{iN})^T$ seu vetor de características, onde N define o número de características utilizadas.

O número de componentes do conjunto, n , é assumido como sendo significativamente maior do que o número de características N .

Um agrupamento é uma partição $[C_1, C_2, \dots, C_K]$ que atribui a cada componente do conjunto um único rótulo de grupo.

Os componentes do grupo C_k formam o k -ésimo grupo, cujo centro é:

$$c_k = (c_{k1}, c_{k2}, \dots, c_{kN})^T \quad (5.8)$$

onde

$$c_{kj} = \frac{1}{M_k} \sum_{i \in C_k} x_{ij} \quad (5.9)$$

e M_k é o número de componentes no grupo k . Deste modo, o centro do grupo é o seu centróide.

O erro quadrático para o grupo k é:

$$e_k^2 = \sum_{i \in C_k} (x_i - c_k)^T (x_i - c_k) \quad (5.10)$$

e o erro quadrático para o agrupamento é:

$$E_K^2 = \sum_{k=1}^K e_k^2 \quad (5.11)$$

Assim, os objetivos do processo de agrupamento são: definir, para um dado K , um agrupamento que minimize E_K^2 , e encontrar um K satisfatório, muito menor do que n .

O algoritmo utilizado no método desenvolvido, cujos passos são descritos a seguir, tem como entradas as componentes da normal a cada *pixel*, ou seja, neste caso $x_i = (a_i, b_i, c_i)$, altamente discriminatória para superfícies planares, e o número K inicial de regiões desejadas, determinado a partir do mapa de bordas. Deseja-se que cada uma delas tenha um erro pequeno quando aproximada por um plano.

Algoritmo de agrupamento K-Means:

- 1. Escolher aleatoriamente K pixels na imagem, e tomar suas características como sendo os valores dos centros de grupos;*
- 2. Atribuir cada pixel da imagem ao grupo cujo centro é mais próximo;*
- 3. Atualizar os centros de grupos como sendo o valor médio de todos os seus pixels componentes;*
- 4. Repetir 2. e 3. até que o número máximo de iterações tenha sido atingido ou haja convergência.*

Após a aplicação do algoritmo de agrupamento, obtém-se uma imagem onde cada *pixel* possui um rótulo que o relaciona a uma região.

5.9. Refinamento da imagem segmentada

Após realizado o agrupamento, a imagem rotulada pode conter inconsistências. Estas inconsistências são devidas, principalmente, ao fato de que duas regiões vizinhas com mesma orientação, mas com profundidades diferentes, vão receber o mesmo rótulo no processo de agrupamento descrito. Existem duas principais metodologias para solucionar este problema:

- (1) Verificar a consistência de profundidade para cada região segmentada (Taylor et al, 1989);
- (2) Incluir mais características (p.ex., as coordenadas x,y,z , de cada *pixel*, conforme Hoffman e Jain, 1987) no vetor usado como entrada do algoritmo de agrupamento.

A duas opções são muito custosas. Por este motivo, decidiu-se utilizar uma nova abordagem para solucionar o problema, aproveitando as informações de bordas obtidas para a imagem. Esta abordagem consiste em sobrepor o mapa de bordas à imagem rotulada, onde o rótulo usado para estes *pixels* de bordas é igual a zero. Em seguida, uma adaptação do algoritmo de contagem de regiões (Seção 5.6.2) é aplicada, realizando uma nova rotulação das regiões. Desta forma, é possível solucionar o problema de regiões distintas vizinhas com mesma orientação.

Como última etapa de refinamento da imagem segmentada, deve-se garantir a absorção dos *pixels* de borda, que foram sobrepostos à imagem rotulada, às suas regiões vizinhas. Isto é realizado através da análise de similaridade entre as normais à superfície nos *pixels* de bordas e as de seus *pixels* vizinhos. Também foi utilizada uma opção para absorver os *pixels* das regiões pequenas por suas regiões vizinhas. O valor de limiar para executar esta absorção é dado em termos de uma porcentagem do tamanho total da imagem.

5.10. Cálculo das equações dos planos

A determinação do plano mais adequado segue o mesmo processo aplicado para o cálculo das normais, minimizando-se o erro quadrático médio para cada região, considerando-se todos os *pixels* que a compõem.

Como resultado, obtém-se os coeficientes da equação do plano que mais se adapta a cada região $k (a_k, b_k, c_k, d_k)$.

5.11. Cálculo do erro de reconstrução.

De posse das equações dos planos e dos pontos que compõem cada região, pode-se calcular o erro global de reconstrução (ε_{gr}) e o erro de reconstrução (ε_r) para cada *pixel*, baseado em equações definidas por Besl e Jain, 1988, e que para o caso de aproximações planares resultam nas equações 5.12 e 5.13.

$$\varepsilon_{gr}^2 = \frac{1}{n - n_b} \sum \varepsilon_r^2(p) \quad (5.12)$$

$$\varepsilon_r^2(p) = (d_k - a_k x_i - b_k y_i - c_k z_i)^2 \quad (5.13)$$

onde n_b é o número total de *pixels* de borda da imagem.

Neste caso, o erro de reconstrução é dado pela soma da distância de cada ponto ao seu plano equivalente, e fornece uma medida da qualidade da reconstrução.

Caso o erro de reconstrução esteja acima do desejado, tomando-se como base o erro de aproximação implícito da imagem (Equação 5.6), o número K de regiões deve ser aumentado e o processo deve retornar à etapa de agrupamento.

5.12. Comparação com outros métodos.

Sumarizando, o método de segmentação desenvolvido, quando inserido no contexto da Tabela 4.1, tem as seguintes características:

5.12.1. Técnicas utilizadas

Em relação às técnicas utilizadas, o método apresentado é semelhante ao de Bhandarkar e Siebert, 1992, discutido na Seção 4.6.8. Uma das diferenças básicas entre os dois métodos é que o aqui apresentado é diretamente aplicado a objetos não poliédricos. Bhandarkar e Siebert afirmam que o método por eles desenvolvido pode ser expandido para objetos não-poliédricos, mas não apresentam nenhuma discussão sobre como isto pode ser realizado.

Uma segunda diferença importante está na metodologia para determinar tanto a estimativa inicial do número de grupos quanto o número ótimo final de grupos. Bhandarkar e Siebert usam sempre o mesmo número inicial de grupos, igual a 20, para inicializar o algoritmo de agrupamento. Neste caso, não existe a opção para aumentar o número de grupos, e a sua redução é feita pela análise de similaridade entre normais a regiões vizinhas. Se duas regiões vizinhas possuem normais muito próximas, segundo um limiar definido, estas devem ser reunidas em uma só. No método apresentado neste capítulo, a estimativa inicial é dada em função do mapa de bordas, e a determinação do número ótimo é baseada no erro de reconstrução. Cabe ressaltar que no caso do trabalho de Bhandarkar e Siebert

sempre existe um número fixo correto de superfícies que compõem a imagem, pois os objetos são sempre poliédricos. No método aqui desenvolvido, este número pode não existir, pois o objeto observado na imagem não precisa ser poliédrico, mas sim pode ser apenas aproximado por pedaços de superfícies planares.

Como última diferença importante, está o fato de que Bhandarkar e Siebert não realizam detecção de bordas propriamente dita, mas sim identificam os limites entre as regiões, após realizada a segmentação, e utilizam estes limites como bordas. Com isso, muita informação valiosa sobre a topologia das superfícies pode ser perdida como, por exemplo, a separação de superfícies vizinhas de mesma orientação mas com descontinuidade de profundidade entre elas. Este problema é solucionado no método aqui desenvolvido pelo uso da informação sobre as bordas.

5.12.2. Tipos de segmentos resultantes

O trabalho de Taylor et al., 1989, discutido na Seção 4.6.2, também segmenta imagens de profundidade por agrupamento em pedaços de superfícies planares mas, como não utilizam informações de bordas, obtém muita fragmentação na posição das mesmas, mesmo para objetos poliédricos simples. O método de segmentação aqui descrito apresenta boa definição nas regiões de borda.

Outros trabalhos, a saber: Abdelmalek, 1990 (Seção 4.6.4), Besl e Jain, 1988 (Seção 4.6.1), Krishnapuram e Munshi, 1991 (Seção 4.6.5), e Yokoya e Levine, 1989 (Seção 4.6.8), segmentam objetos em superfícies planares e quádricas. A maior limitação destes é que, por utilizarem características de

geometria diferencial (valores de curvatura K e H), a identificação de planos têm grande sensibilidade a ruídos, quando comparados ao método aqui apresentado.

O trabalho de Hoffman e Jain, 1989, discutido na Seção 4.6.3, segmenta imagens de profundidade em superfícies planares, côncavas, e convexas por um algoritmo de agrupamento. Neste trabalho, o agrupamento é realizado com um vetor de característica de dimensão-6 (os três coeficientes da normal e as três coordenadas 3-D para cada *pixel*), de modo a tentar solucionar o problema de regiões vizinhas distintas mas com mesma orientação. Esta medida onera muito o custo de execução do algoritmo de agrupamento, conforme discutido na Seção 4.3.4, e não garante a solução do problema.

5.12.3. Principais contribuições

Como principais contribuições do método de segmentação aqui desenvolvido ressaltam-se:

- (1) Uma nova abordagem para determinação do número adequado de regiões para a imagem.
- (2) Uma nova abordagem para gerar uma estimativa inicial para o número de regiões da imagem.

Com o emprego de uma estimativa inicial para o número ótimo de regiões da imagem, aliada às informações de bordas, além da medida de qualidade do

agrupamento com base nos dados da reconstrução da imagem, o método de segmentação apresentado se torna eficiente pelos seguintes motivos:

- Usa apenas três elementos no vetor de características de entrada, que são os coeficientes do vetor normal a cada *pixel*.
- Preserva as bordas naturais do objeto, devido às informações de borda inseridas no processo de agrupamento.
- Separa diretamente regiões vizinhas diferentes mas com mesma orientação, também devido às informações de borda.
- Utiliza uma medida clara para atualização do número ótimo de regiões que compõem a imagem.

5.12.4. Principais limitações

Caso seja necessário segmentar a imagem em pedaços de superfícies de maior grau, a abordagem por agrupamento, conforme definida, pode se tornar ineficiente devido ao maior número de características de entrada do algoritmo. Além disso, o próprio processo de ajuste de quádricas é mais complexo do que por planos. Uma outra alternativa é contruir os pedaços de superfícies a partir de um *merging* entre os segmentos de planos obtidos. Estas alternativas compõem o conjunto de trabalhos futuros propostos na Seção 9.2.

Uma outra limitação do método de segmentação utilizado é a possibilidade de obter pouca definição na segmentação de regiões da imagem que possuam muitos detalhes. Esta limitação é inerente à técnica de agrupamento utilizada [Dubes e Jain, 1976; Duda e Hart, 1973; Jain e Flynn, 1996], pois o objetivo é minimizar o erro total do agrupamento, e a solução pode convergir para mínimos locais. Neste sentido, dependendo do tipo de algoritmo utilizado, ou mesmo do modo pelo qual o algoritmo foi inicializado, podem ser obtidas diversas soluções para o mesmo problema, como mostra a Figura 5.2.

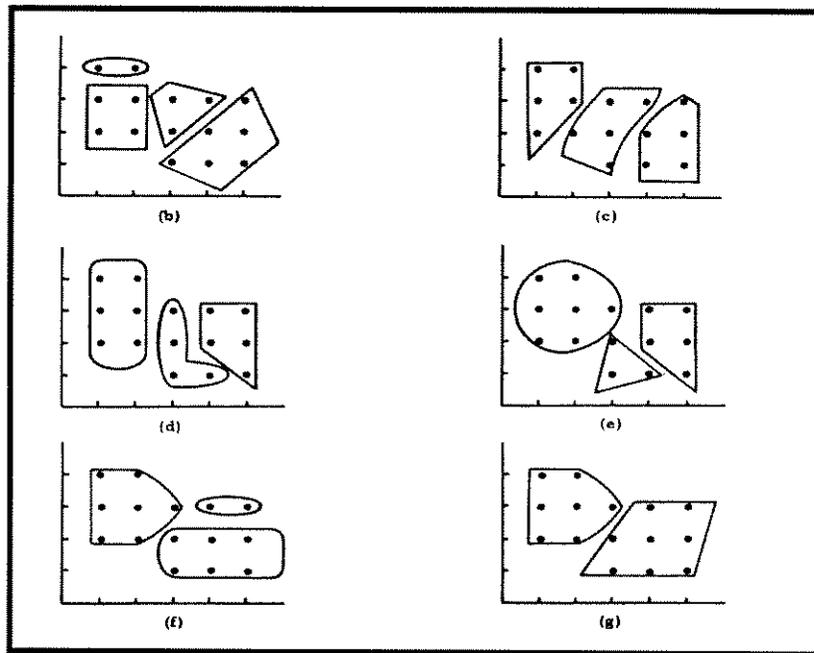


Figura 5.2: Várias soluções para o agrupamento de um mesmo conjunto de dados.

(Dubes e Jain, 1976)

A Figura 5.2 reflete um estudo sobre alguns algoritmos de agrupamento, realizado por Dubes e Jain, 1976. Portanto, se as regiões de interesse da imagem

não são adequadamente distintas, podem ocorrer erros de segmentação. Este problema se faz sentir mais gravemente em regiões com muitos detalhes, embora a detecção de bordas, inserida no processo de segmentação, possa auxiliar na obtenção da solução correta. Desta forma, esta pode ser considerada uma limitação significativa da técnica utilizada.

5.13. Sumário

Este capítulo apresentou o método de segmentação desenvolvido. Em seguida, no Capítulo 6, será apresentado o método de representação aplicado aos resultados da segmentação.

Capítulo 6: O método de representação

6.1. Introdução

Uma vez que se tenha realizado uma segmentação confiável na imagem de profundidade, o próximo passo é criar uma representação adequada para o conjunto de regiões segmentadas. Esta representação possibilitará a transformação da informação obtida para uma forma mais aplicável a problemas de visão computacional, tal como o reconhecimento 3-D de objetos.

O processo de representação é fortemente relacionado aos resultados da segmentação, pois uma separação incorreta dos elementos da imagem leva a uma representação errônea da mesma. Isto sugere que a solução dos dois problemas seja realizada de forma integrada, de modo que discrepâncias obtidas na representação atuem como uma realimentação ao processo de segmentação. No entanto, a despeito de sua importância, esta integração tem sido pouco contemplada na literatura, principalmente do ponto de vista topológico.

Este capítulo tem como objetivo apresentar o método de representação desenvolvido, cujos resultados podem ser utilizados como realimentação ao processo de segmentação. O método é baseado na representação por fronteira, e é dirigido à topologia dos objetos observados. A Seção 6.2 faz uma breve discussão sobre os trabalhos relacionados, encontrados na literatura. Em seguida, na Seção

6.3, apresenta-se como os dados são organizados na estrutura utilizada para a representação. A Seção 6.4 trata da obtenção dos vértices, arestas, e polígonos que vão compor a estrutura. Na Seção 6.5, apresenta-se alguns problemas associados à representação construída. Finalmente, na Seção 6.6, discute-se uma metodologia para o cálculo de medidas de qualidade para os resultados obtidos.

6.2. Trabalhos relacionados

Em alguns trabalhos, os processos de segmentação e descrição de imagens são implementados de forma integrada. Tanto Abdelmalek, 1990, quanto Yokoya e Levine, 1989, segmentam imagens de profundidade com base nos sinais de curvatura K e H , e a descrição final dos segmentos é dada em termos dos oito tipos básicos de regiões definidas por estes sinais. Besl e Jain, 1988, também segmentam imagens de profundidade com base nas curvaturas K e H mas, em seguida, geram uma descrição mais completa para os segmentos obtidos através do crescimento de regiões com ajuste variável de superfícies de primeiro até quarto grau.

Hoffman e Jain, 1987, usam segmentação por agrupamento com um vetor de característica de dimensão-6 (três coeficientes da normal mais as três coordenadas 3-D para cada *pixel* da imagem) onde a descrição final dos segmentos é limitada entre côncavo, convexo, ou planar. Newman et al., 1993, tomam os segmentos obtidos pelo método de Hoffman e Jain, e fazem um ajuste de quádricas (cone, cilindro, esfera, e plano) para cada um deles usando transformada de Hough.

Estes trabalhos já foram exaustivamente discutidos na Seção 4.6 e, apesar de gerarem uma descrição para as regiões segmentadas, estas descrições não são relacionadas à topologia da cena, ou seja, segmentos vizinhos podem não atender ao critério de continuidade de superfície, nem apresentam uma solução para a representação dos objetos observados na imagem.

Bhandarkar e Siebert, 1992, apresentam uma abordagem mais próxima ao problema que desejamos solucionar: a integração dos processos de segmentação e de representação como forma de melhorar o desempenho global do sistema desenvolvido. A idéia básica dos autores é associar informações sobre a topologia dos objetos observados em imagens de profundidade, de modo a auxiliar o processo de segmentação. Neste trabalho, após uma segmentação inicial por um algoritmo de agrupamento, utiliza-se dados sobre a interseção das superfícies planares e arestas identificadas para refinar as fronteiras entre regiões. Este processo de refinamento das coordenadas dos vértices e direção das arestas identificadas é uma contribuição importante do método. Como produto final, obtém-se informação para montar uma representação por fronteira para os objetos segmentados, no entanto, a construção desta representação, bem como seus problemas associados, não são discutidos. Como principais limitações do trabalho de Bhandarkar e Siebert podem ser citados:

- (1) É aplicado apenas a objetos poliédricos. Embora os autores afirmem que a sua representação pode ser estendida a objetos não-poliédricos, as questões básicas associadas não são discutidas como, por exemplo, o tratamento a ser aplicado a arestas curvas.
- (2) Superfícies vizinhas com mesma orientação são agrupadas em uma única região.
- (3) Não é discutida a estrutura utilizada para a representação.

Hoover et al., 1995, apresentam um trabalho, aplicado a visão computacional, no qual o objetivo principal é extrair um modelo por fronteira válido para um objeto. A entrada do sistema é o resultado da segmentação de uma única imagem de profundidade, que contém o objeto. Suas principais contribuições são:

- (1) Delimita um espaço (*space enveloped*) em torno do objeto, formado por suas partes visíveis mais uma extensão baseada nas suas bordas de profundidade (perfil). Esta característica é importante, por exemplo, para navegação autônoma, pois prevê o máximo volume possível ocupado pelo objeto, de modo a evitar colisões.
- (2) Tenta construir modelos válidos mesmo na ocorrência de informações de segmentação imprecisas e/ou erros de obtenção da topologia.

Os autores afirmam que seu trabalho é o primeiro a tratar o problema de modelagem de vértices formados por mais de três faces, bem como sugerem que várias vistas de um mesmo objeto podem ser utilizadas para construir um modelo 3-D completo. A principal limitação deste trabalho é, também, o fato de ser aplicado apenas a objetos poliédricos e a pouca discussão a respeito da estrutura utilizada para a representação.

Diversos trabalhos (Chen e Medioni, 1992; Faugueras e Herbert, 1987; Higuchi et al., 1995; Sapidis e Besl, 1995; Seales e Faugueras, 1995; Soucy e Laurendeau, 1996) utilizam triangulação para segmentar superfícies de objetos, sem restringí-los a poliédricos, em imagens de profundidade. Nenhum deles discute a construção de uma estrutura a ser utilizada na representação dos objetos.

É sabido que a representação por fronteira pode ser aplicada para um conjunto de triângulos, pois respeita a topologia do objeto, gerando superfícies contínuas, e tem baixo erro de reconstrução. Entretanto, apresenta alto custo computacional, principalmente em aplicações de reconhecimento 3-D, quando é necessário fazer o casamento de estruturas.

6.3. A estrutura utilizada

Neste trabalho, foi desenvolvido um método simples e eficiente para criar uma representação poliédrica por fronteira, dirigida à topologia, não apenas para imagens contendo objetos poliédricos mas também com objetos não-poliédricos. A

idéia básica é tomar as K regiões resultantes da segmentação (R_1, R_2, \dots, R_k) e construir uma representação poliédrica para este conjunto R de regiões, onde:

$$R = \cup_{i=1}^K R_i \quad (6.1)$$

O objetivo é que a representação seja robusta, no sentido de que não existam descontinuidades ou *cracks* nas superfícies do objeto, a menos daquelas que são inerentes ao próprio objeto. Como os objetos não são necessariamente poliédricos, a questão principal aqui é encaixar os segmentos de planos obtidos para este conjunto R de modo a formar uma superfície poliédrica, onde tanto as bordas de profundidade quanto as de orientação são preservadas.

O resultado da representação criada é uma estrutura para R contendo: um conjunto de regiões componentes; as coordenadas 3-D dos vértices de cada região, ordenados de modo a formarem um polígono; e a equação do plano associado a cada região.

Inicialmente, a informação que se tem é uma imagem segmentada em pedaços ou regiões que reunidas compõem um ou mais objetos. Nesta imagem, cada *pixel* de uma mesma região possui um mesmo rótulo, e todos os *pixels* têm suas coordenadas 3-D conhecidas (característica da imagem de profundidade). Além disso, a equação do plano mais adequado a cada região também é conhecida.

Para criar uma representação poliédrica robusta a partir da informação dada, é necessário que esta representação satisfaça as condições para ser uma

superfície poliédrica (Hoffman, 1989; Mäntyla, 1988; Preparata e Shamos, 1985), em resumo: a superfície deve ser dividida em faces planares, as quais se interceptam em arestas. Assim, as regiões serão aproximadas por polígonos e suas interseções serão segmentos de retas. Para atender a estas condições, é necessário considerar as interseções entre regiões adjacentes com base em suas equações de planos.

A solução proposta para criar a representação poliédrica é construir uma lista de vértices L_v e uma lista de arestas L_e , como mostram as Figuras 6.1 e 6.2.

Cada elemento da lista de vértices comporta a informação do rótulo das regiões a que pertence, suas coordenadas 3-D, e suas coordenadas de imagem. Cada elemento da lista de arestas contém a informação do rótulo das duas regiões que separa e dos rótulos de seus dois vértices componentes. Com estas informações, e utilizando o método proposto, pode-se chegar à representação final por regiões L_r desejada, cuja estrutura é mostrada na Figura 6.3.

O processo para determinar as coordenadas dos vértices e as arestas mais apropriadas para representar os objetos segmentados é apresentado na próxima seção.

6.4. Determinação dos vértices e arestas

Dentro do contexto deste trabalho, um vértice é definido por um *pixel* que contém duas ou três regiões diferentes da sua na sua vizinhança, considerando-se vizinhança-4 e uma janela 3x3 centrada no *pixel*. Baseado na definição de vértice apresentada, três situações podem ocorrer:

- (1) Um vértice pode ser formado pela interseção de três regiões adjacentes;
- (2) Um vértice pode ser formado pela interseção de quatro regiões adjacentes;
- (3) Um vértice pode pertencer a uma borda de profundidade (perfil).

Assim que um vértice dos casos (1) ou (2) é identificado, utiliza-se uma janela 5x5 para confirmar se o mesmo está numa região de borda de profundidade ou não. O aumento no tamanho da janela é para dar maior confiabilidade à identificação da descontinuidade de profundidade na região.

A restrição do *pixel* possuir no máximo quatro regiões em seu entorno está ligada ao fato de se utilizar vizinhança-4 na análise. Poder-se-ia ter utilizado vizinhança-8. Neste caso, a restrição do número máximo de regiões convergentes no vértice aumentaria para 8. Pode-se notar que, mesmo utilizando vizinhança-8, um vértice pode ser o ponto de convergência de um número de regiões maior do que 8. Neste caso, segundo o algoritmo implementado, serão identificados mais de um

vértice cujas coordenadas (tanto de imagem quanto 3-D) possuem grande proximidade entre si. Esta concentração de vértices é identificada no pós-processamento dos vértices, observando-se suas coordenadas, com base num valor de limiar definido. Assim, um conjunto de vértices, cujas distâncias sejam menores do que o limiar estipulado, são agrupados em apenas um, e as listas de arestas, vértices, e regiões são atualizadas.

Sabe-se que para determinar com precisão as coordenadas de um vértice do objeto é necessário utilizar as equações dos planos relativos a três regiões adjacentes no objeto. Então, o primeiro passo é identificar as regiões que formam cada vértice. Para isso, deve-se realizar uma varredura na imagem, verificando-se para cada *pixel* se este é vizinho de uma região diferente da sua, usando vizinhança-4. Cada *pixel* encontrado que tiver duas ou três regiões vizinhas, diferentes da sua, é identificado como vértice.

As Figuras 6.4(a)-(e) mostram exemplos de situações topológicas que podem ser encontradas, a partir da varredura na imagem, para uma vizinhança 3x3, centrada em um *pixel* p . A Figura 6.4(a) mostra a situação onde o *pixel* p está dentro de uma região homogênea (r_1) não sendo, portanto, identificado como vértice. Na Figura 6.4(b) tem-se que o *pixel* p encontra-se no limite entre duas regiões (r_1 e r_2). Também, neste caso, o *pixel* p não é identificado como vértice. Nas Figuras 6.4(c) e 6.4(d) o *pixel* p tem duas (r_2 e r_3) e três (r_2 , r_3 , e r_4) regiões diferentes da sua na vizinhança em questão, respectivamente, e deve ser inserido na

lista de vértices em ambos os casos. Na situação da Figura 6.4(e), de acordo com a definição de vértice, o *pixel* p não é identificado como vértice.

Quando um vértice é identificado, define-se um conjunto de arestas, uma para cada fronteira entre as regiões identificadas como vizinhas no vértice. Entretanto, como ainda podem não ser conhecidos os outros vértices que suportam as arestas, estas não devem ainda ser incluídas na lista final de arestas, e sim numa lista auxiliar de semi-arestas L_{he} . A informação do número do vértice que irá completar a semi-aresta é inicializado com o valor -1, até que este seja encontrado durante a identificação de vértices e montagem da Lista de semi-arestas. Cada uma das semi-arestas se tornará uma aresta completa quando o segundo vértice que a compõe for inserido na lista de vértices, e supõe-se que existem duas e apenas duas semi-arestas entre duas regiões vizinhas no vértice.

Devido à aproximação planar das regiões segmentadas, as coordenadas 3-D que melhor representam cada vértice em função de suas regiões vizinhas podem não ser exatamente as do ponto 3-D correspondente ao *pixel* identificado como vértice. Assim, deve-se calcular as novas coordenadas para cada vértice, considerando-se as interseções entre os planos vizinhos no vértice.

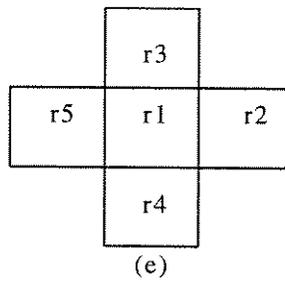
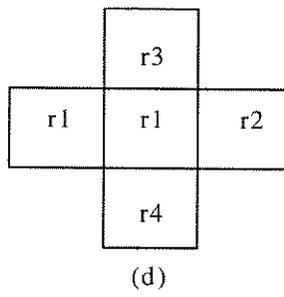
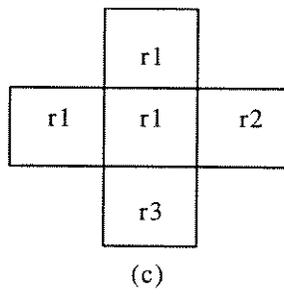
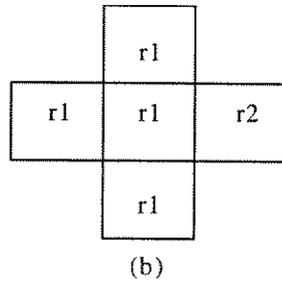
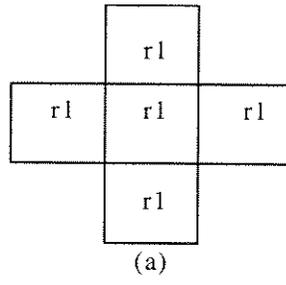


Figura 6.4: Exemplos de situações encontradas durante a varredura da imagem.

Uma outra consideração a ser observada, também devido ao fato de que os objetos a serem representados não precisam ser poliedricos, mas sim suas regiões segmentadas devem ser aproximadas por planos, é que tanto as bordas de profundidade quanto as bordas de orientação mantêm o seu formato original após a segmentação, ou seja, se estas bordas são curvas irão gerar limites curvos entre regiões. Por este motivo, estas bordas podem não ser adequadas como arestas na representação polidrica. A solução encontrada para este problema foi realizar uma poligonalização no mapa de bordas da imagem. Esta poligonalização tem como resultado um conjunto de vértices e suas arestas correspondentes, onde é possível existir mais do que uma aresta dividindo as mesmas regiões vizinhas.

Para o cálculo das coordenadas 3-D de cada vértice da representação foram criadas três alternativas, uma para cada situação possível, a saber:

- (1) O vértice é formado pela interseção de três regiões adjacentes.
- (2) O vértice é formado pela interseção de quatro regiões adjacentes.
- (3) Os vértices pertencem a uma borda de profundidade.

Na primeira situação, as coordenadas 3-D do vértice podem ser calculadas com exatidão por um sistema de equações lineares simples, usando as três equações dos planos relativos a cada região. Em seguida, as coordenadas 3-D calculadas para o vértice devem ser inseridas na lista de vértices.

Quando o vértice é formado por quatro regiões adjacentes, a aplicação do método de resolução por equações lineares pode levar a um sistema indeterminado, pois o número de equações de planos (igual a 4, uma para cada região da vizinhança) é maior do que o número de incógnitas (igual a 3, que são as coordenadas 3-D do vértice). Esta situação espacial pode ser melhor visualizada no exemplo da Figura 6.5.

A Figura 6.5(a) mostra a situação espacial para o caso de um vértice ser formado pela interseção de quatro regiões vizinhas. As Figuras 6.5(b) e 6.5(c) mostram como o estado da Figura 6.5(a) pode ser transformado espacialmente. Em resumo, um vértice formado por quatro regiões deve ser transformado em dois vértices ligados por uma aresta, de modo que cada um dos vértices fique vizinho a três regiões. O sentido da aresta criada será dado por convenção, considerando-se que a mesma irá dividir a região do *pixel* central com a região diretamente oposta àquela do *pixel* da vizinhança que tem o mesmo rótulo do *pixel* central. No exemplo da Figura 6.4(d) a situação vai ser transformada para a da Figura 6.5(c). Neste caso, deve-se inserir os dois vértices na lista de vértices, um para cada combinação possível de três regiões, após se calcular suas coordenadas 3-D. Segue-se o processo de inserção de semi-arestas.

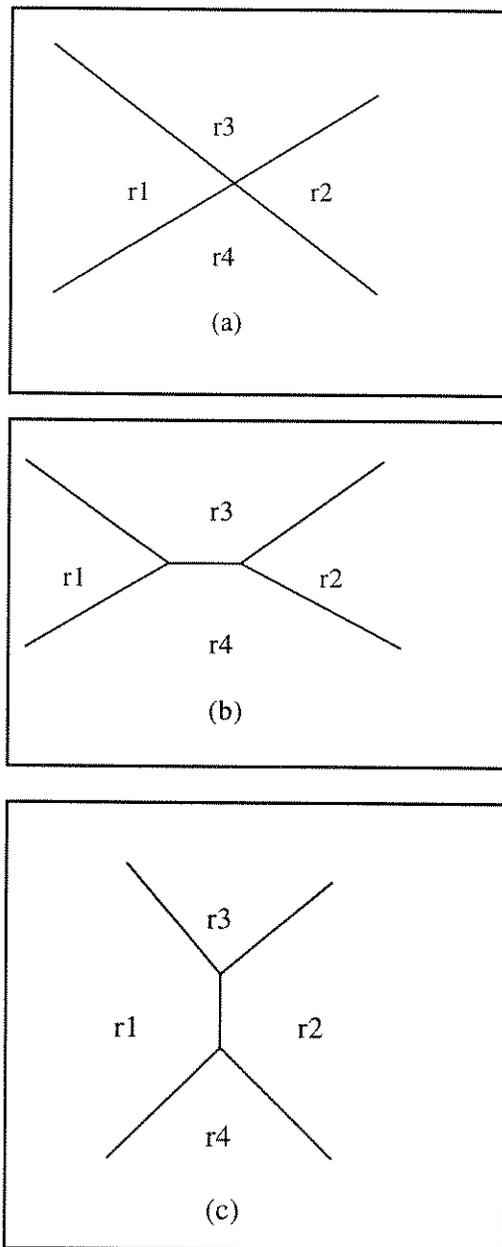


Figura 6.5: Soluções para a situação exemplificada na Figura 6.4(d).

No último caso, para vértices que pertencem a uma borda de profundidade, é necessário considerar a existência de uma descontinuidade entre suas regiões vizinhas. Por isso, deve-se projetar as coordenadas do vértice detetado em cada uma de suas regiões vizinhas. As coordenadas resultantes das projeções

são inseridas na lista de vértices, ou seja, um único vértice é transformado em vários outros, de acordo com o número de regiões vizinhas no vértice. Em seguida, deve-se inserir as semi-arestas correspondentes, a menos das semi-arestas relativas às bordas. As semi-arestas correspondentes às bordas, tanto de orientação quanto de profundidade, não irão compor a lista de semi-arestas, mas sim irão ser inseridas diretamente na lista de arestas por dois motivos:

- Devido ao processo de poligonalização realizado, estas não atendem ao critério de que podem existir duas e apenas duas semi-arestas separando duas regiões.
- Seus dois vértices componentes já são conhecidos diretamente pelo processo de poligonalização, sem que seja necessário realizar a varredura da imagem.

O algoritmo para construção da representação poliédrica dos objetos é mostrado na Figura 6.6.

Com os vértices, as arestas, e as equações do plano mais adequado a cada segmento, a representação desejada pode ser gerada, e a imagem pode ser reconstruída.

Algoritmo para representação poliédrica:

1. *Poligonalizar o mapa de bordas;*
2. *Para cada aresta do polígono:*
 - 2.1. *Para cada vértice da aresta:*
 - 2.1.1. *Caso o vértice não esteja inserido em L_v :*
 - 2.1.1.1. *Inserir o vértice em L_v ;*
 - 2.2. *Inserir a aresta em L_e ;*
3. *Realizar a varredura da imagem para detetar pixels candidatos a vértices:*
 - 3.1. *Para cada pixel detetado na varredura:*
 - 3.1.1. *Caso o ponto correspondente não esteja inserido em L_v :*
 - 3.1.1.1. *Se o pixel tem 4 regiões vizinhas diferentes:*
 - 3.1.1.1.1. *Separar o ponto em dois vértices, conforme Figura 6.5;*
 - 3.1.1.1.2. *Para cada ponto identificado como vértice:*
 - 3.1.1.2.1. *Se as regiões vizinhas ao ponto não contém descontinuidade de profundidade:*
 - 3.1.1.2.1.1. *Calcular as novas coordenadas do ponto pela interseção de suas regiões vizinhas;*
 - 3.1.1.2.1.2. *Inserir o ponto como um vértice em L_v e as semi-arestas correspondentes em L_{he} ;*
 - 3.1.1.2.2. *Caso contrário:*
 - 3.1.1.2.2.1. *Projetar o vértice em cada uma de suas regiões vizinhas;*
 - 3.1.1.2.2.2. *Inserir os novos vértices em L_v e as semi-arestas correspondentes em L_{he} ;*
 - 3.1.1.2.3. *Corrigir as listas de arestas, semi-arestas, e vértices;*
 - 3.1.2. *Caso o ponto esteja muito perto de uma das arestas do polígono:*
 - 3.1.2.1. *Dividir a aresta em duas e inserir o vértice no meio;*
 - 3.1.2.2. *Calcular as novas coordenadas do vértice;*
 - 3.1.2.3. *Corrigir as listas de arestas, semi-arestas, e vértices;*
 4. *Anexar a lista de semi-arestas L_{he} na lista de arestas L_e ;*
 5. *Avaliar lista de vértices:*
 - 5.1. *Se existem vértices muito próximos:*
 - 5.1.1. *Agrupar os vértices, calcular as coordenadas resultantes;*
 - 5.1.2. *Reorganizar as listas de vértices e arestas.*
 6. *Gerar L_r , com base em L_v e L_e .*

Figura 6.6: Algoritmo para representação poliédrica.

6.5. Principais problemas associados

Foram encontradas duas dificuldades principais na construção da representação desejada. A primeira delas refere-se ao valor de limiar utilizado para determinar a proximidade entre duas entidades. A segunda é relativa à inconsistência na determinação das coordenadas dos vértices que irão compor a estrutura. Estes problemas são discutidos nesta seção.

6.5.1. Limiar para proximidade entre duas entidades

Este problema é encontrado em três situações diferentes:

- (1) Quando existem dois ou mais vértices muito próximos, que devem ser ou não agrupados em um único vértice.
- (2) Quando existe um vértice muito próximo a uma aresta, e é necessário decidir se esta deve ou não ser dividida em duas com o vértice próximo no meio.
- (3) Quando existem três arestas se interceptando duas a duas, em pontos muito próximos.

Com base nos experimentos realizados, de acordo com o conjunto de imagens utilizadas, definiu-se um valor de limiar igual a 5 para a distância (Euclidiana) entre vértices, entre vértice e aresta, ou ainda entre interseções de

arestas, considerando-se coordenadas 3-D, para o qual estas entidades devem ser agrupadas.

No caso da verificação entre vértice e aresta, o procedimento é o seguinte:

- 1. Para cada vértice, calcular sua distância em relação a cada uma das arestas que compõem a representação.*
- 2. Para a menor distância calculada, verificar se esta é menor do que o limiar utilizado.*
- 3. Se for menor, quebrar a aresta ao meio e inserir o vértice, cujas novas coordenadas serão dadas pelas de sua projeção na aresta.*
- 4. Atualizar as listas de vértices, de semi-arestas, e de arestas.*

Para o caso da verificação entre vértices, os passos são os seguintes:

- 1. Avaliar a lista de vértices e calcular a distância entre cada dois deles.*
- 2. Agrupar os vértices que têm distância entre eles menor do que o valor de limiar estipulado. Atribuir suas coordenadas como o ponto médio dos vértices agrupados.*
- 3. Atualizar as listas de vértices e de arestas.*

No caso de interseção entre arestas, são identificados três vértices, um para cada interseção entre duas arestas. Então, o problema recai na mesma situação de análise de proximidade entre vértices.

Existe uma série de problemas associados com a questão de proximidade entre entidades: vértice-vértice, vértice-aresta, e aresta-aresta. Em algumas situações, esta proximidade não implica que as entidades devam ser agrupadas em uma só. Para solucionar estes problemas pode-se, por exemplo, inserir informações

sobre os próprios modelos dos objetos que podem aparecer na cena. Uma discussão sobre o assunto pode ser encontrada em Shimshoni e Ponce, 1997, do ponto de vista de construção de grafos de aparência para objetos poliédricos.

6.5.2. Inconsistências geométricas

Um outro problema, muito discutido na literatura relativa a representação por fronteira, é a existência de inconsistências geométricas na estrutura final construída. Mais especificamente, é necessário garantir que os vértices definidos para cada polígono da estrutura estão contidos no plano da região definida pelo polígono. No caso em questão, ou seja, representação por fronteira para objetos a partir de sua imagem de profundidade usando informações sobre bordas, duas questões devem ser observadas:

- (1) Inconsistências decorrentes dos vértices de junção entre bordas de orientação e/ou “bordas falsas”. Definimos bordas falsas como aquelas geradas pela aplicação do algoritmo de agrupamento, que não são coincidentes com nenhuma outra borda de orientação ou profundidade.
- (2) Inconsistências relativas aos vértices nas bordas de profundidade.

O primeiro tipo de inconsistência é tratado pelo refinamento das coordenadas dos vértices desta situação, pela interseção das equações de plano das regiões vizinhas ao vértice. Com este procedimento, ameniza-se a inconsistência geométrica associada.

No segundo caso, utilizam-se as projeções do vértice em cada um dos planos das suas regiões vizinhas. Com isso, novas coordenadas 3-D são obtidas, e elimina-se o erro associado, que é a distância do vértice ao plano, a menos de problemas de precisão numérica. Entretanto, este procedimento gera um aumento no número de vértices e arestas da estrutura.

A Figura 6.7 permite visualizar o problema discutido. Nesta figura, que é a descrição aramada de um paralelepípedo sobre um plano, pode-se observar que o vértice 1 vai ser transformado em 3 vértices, um relativo à região A, outra à região B e, finalmente, um terceiro relativo à região D. Devido ao critério de proximidade entre vértices, aqueles associados às regiões A e B devem ser agrupados em apenas um.

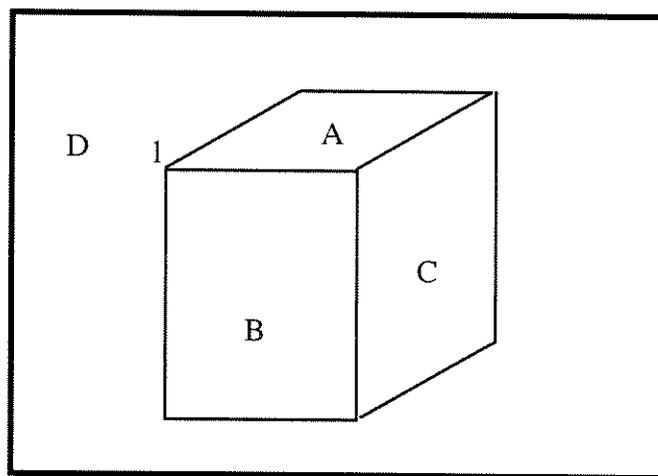


Figura 6.7: Exemplo de necessidade de duplicação de vértice.

Ainda considerando-se as inconsistências relativas aos vértices nas bordas de profundidade, em relação ao problema específico de construção de uma estrutura para conter a representação do objeto observado na imagem de profundidade, uma outra maneira de visualizar a necessidade da duplicação de vértices é mostrada na Figura 6.8.

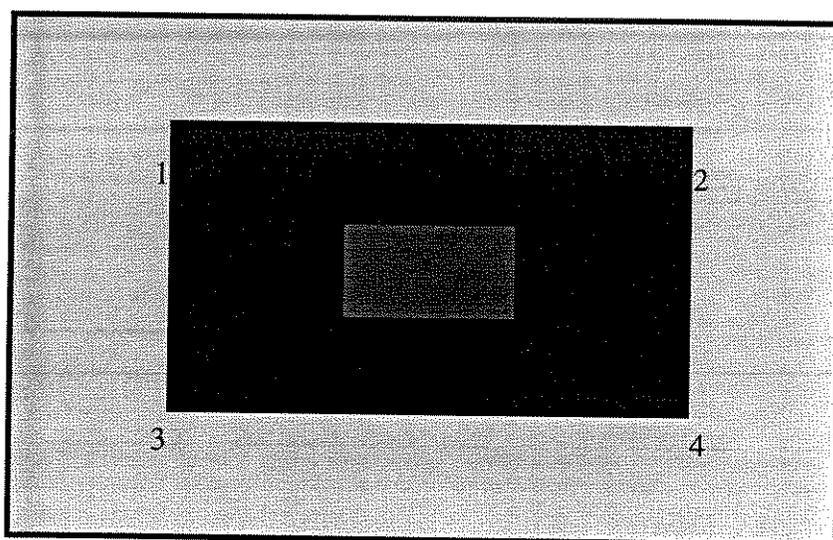


Figura 6.8: Outro exemplo de inconsistência relativa às bordas de profundidade.

A Figura 6.8 mostra a descrição aramada da face de um objeto que contém um buraco no seu interior. A descontinuidade de profundidade do plano que contém a face maior do objeto, em relação ao plano que contém o buraco, é grande o suficiente para que os vértices 5-6-7-8 sejam considerados como pertencentes a uma borda de profundidade. Neste caso, se não houver a duplicação de vértices, a região do buraco será composta pelos vértices 5-6-7-8, cujas coordenadas poderiam

não estar todas contidas no plano que define a região, mas sim no plano que contém a face maior do objeto.

6.6. Medida da qualidade da representação

Uma vez que a representação tenha sido criada, sua qualidade pode ser medida com base nos valores das coordenadas 3-D originais. Para calcular esta medida, a representação obtida deve ser colocada em correspondência com a imagem de profundidade original. Esta correspondência é realizada por um procedimento de calibração (Bellon, 1990), visto que a correspondência entre os pontos 3-D e os pontos da imagem é conhecida.

Depois da calibração, as coordenadas 3-D obtidas para os vértices são projetadas e suas arestas correspondentes são construídas, de modo a criar uma imagem 2D aramada. Esta imagem pode ser usada como uma ilustração visual da qualidade da representação.

Uma medida quantitativa da qualidade da representação é medida com base nas diferenças entre as coordenadas 3-D originais dos vértices e as correspondentes obtidas pela construção da representação, conforme a Equação 6.3 abaixo.

$$\varepsilon_R^2 = \sum_{i=1}^{nv} (p_{iR} - p_{io})(p_{iR} - p_{io})^T \quad (6.3)$$

onde ε_R é o erro da representação, nv é o número de vértices da representação, p_{iR} é o ponto da representação, e p_{io} é o ponto original.

6.7. Sumário

Este capítulo apresentou uma representação, baseada em B-Rep, que pode ser diretamente aplicada aos resultados obtidos para a segmentação descrita no Capítulo 5. Em seguida, no Capítulo 7, é realizada uma discussão sobre o problema de construção de modelos 3-D completos para objetos a partir de várias imagens (vistas) de profundidade.

Capítulo 7: Discussão sobre a construção de modelos

3-D completos

7.1. Introdução

Um dos objetivos deste trabalho é que seus resultados possam colaborar para o processo de reconhecimento 3-D de objetos. Dentro deste contexto, um próximo passo para o trabalho já realizado seria usar a representação criada para construir um modelo 3-D completo dos objetos visualizados. Este capítulo tem o intuito de discutir algumas das atuais abordagens para se construir este modelo, bem como indicar uma metodologia para se aplicar a representação desenvolvida na construção de modelos 3-D completos.

7.2. Casamento de vistas

A representação poliédrica descrita no Capítulo 6 é relativa apenas ao conjunto de superfícies visíveis da cena, que formam a imagem de profundidade. Para aplicações tais como reconhecimento 3-D de objetos muitas vezes é desejável que se construa um modelo 3-D completo dos objetos que compõem a cena. Este modelo deve ser gerado com a utilização de várias imagens. Estas imagens, no geral, são tomadas a partir de deslocamentos rotacionais (da câmera ou da plataforma que contém os objetos) de forma a cobrir um espaço de 360° em torno

dos objetos. Uma vez que as imagens tenham sido obtidas, é necessário realizar um casamento (*matching*) das vistas. A partir dos resultados deste casamento, pode-se construir o modelo 3-D desejado.

Então, basicamente, são necessárias três etapas para se construir um modelo 3-D completo para um objeto:

- (1) Aquisição de várias vistas;
- (2) Casamento entre as vistas;
- (3) Integração das vistas.

A próxima seção discute alguns trabalhos relacionados ao tema construção de um modelo 3-D completo.

7.3. Trabalhos relacionados

7.3.1. Seales & Faugueras

Seales e Faugueras, 1995, desenvolveram um sistema para reconstruir um modelo 3-D completo para objetos que contém tanto bordas suaves quanto acentuadas. A entrada do sistema é uma seqüência de imagens de intensidade luminosa do objeto. O número total de imagens adquiridas para cada objeto é 12, cobrindo um espaço de 360° em torno do objeto, com um espaçamento de 30° entre cada imagem tomada. A saída é uma malha de pontos que pode ser manipulada e

visualizada por ferramentas padrões de computação gráfica (Foley et al., 1996). O sistema usa visão trinocular (três câmeras: vista superior, vista inferior, vistas de perfil) e explora características de bordas nas imagens para extrair a informação 3-D e reconstruir o objeto, ou seja, a informação 3-D é baseada nas informações de bordas. Uma vez que esta informação seja conhecida, o conjunto de pontos 3-D obtidos é ajustado por triangulação, gerando uma malha que representa a superfície do objeto. Uma dificuldade observada neste sistema é a sua dependência da qualidade dos resultados da etapa de detecção de bordas, o que cria interpretações errôneas em muitos casos, como mostram os experimentos apresentados. Além disso, não é abordado como os resultados obtidos podem ser aplicados ao problema do reconhecimento 3-D de objetos.

7.3.2. Higuchi et al.

O trabalho de Higuchi et al., 1995, descreve um método para construir um modelo 3-D a partir de um conjunto de imagens de profundidade de um mesmo objeto, tomadas de diferentes pontos. A idéia básica neste trabalho também é construir malhas discretas para representar as superfícies observadas em cada cena. Cada malha deve ser mapeada como uma imagem esférica, e as transformações entre as diversas observações (vistas) são calculadas por meio de um casamento entre as respectivas imagens esféricas. As malhas são obtidas por meio de um processo inverso ao da triangulação, formando pequenos elementos hexagonais, onde os nodos são ajustados de modo a serem o mais próximos possíveis da

superfície original. Duas medidas de qualidade são obtidas: uma para o erro de transformação entre as vistas, e a outra para o erro de reconstrução da superfície, o que confere credibilidade ao método. Apesar de apresentar bons resultados, a qualidade dos resultados neste trabalho é altamente dependente de uma boa escolha para as vistas obtidas.

7.3.3. Masuda & Yokoya

Masuda e Yokoya, 1995, apresentam um método robusto para correspondência entre imagens de profundidade que admite até 50% de *pixels outliers* na imagem. Os *pixels outliers* são aqueles decorrentes de erros de aquisição ou ruído. As imagens são segmentadas de modo que seus *pixels* sejam classificados entre *inliers* e *outliers*. Apenas os *pixels* classificados como *inliers* são utilizados para determinar a correspondência entre as imagens. O método faz uso de complexas técnicas estatísticas para atingir a robustez desejada. O processo de registro associa uma adaptação do algoritmo ICP (Besl e McKay, 1992) com a representação de *quaternions* (Faugueras e Hebert, 1987). O método realiza apenas o registro das imagens, e não chega a desenvolver uma representação 3-D completa para os objetos reconstruídos.

7.3.4. Soucy & Laurendeau

Soucy e Laurendeau, 1996, apresentam uma estratégia para modelagem de superfícies baseada em triangulação hierárquica. As superfícies são obtidas através

do casamento de diversas imagens de profundidade relativas a diversas vistas de um mesmo objeto. A idéia principal é, a partir de uma triangulação inicial bem refinada, modificar a resolução do sistema de modo a adquirir triangulações mais grosseiras, com base em diferentes valores de limiar escolhidos para o erro de ajuste dos triângulos. O objetivo é obter representações mais comprimidas do objeto, de modo que a sua visualização possa ser mais rápida. Apesar de custoso, o método tem excelentes resultados no sentido de que garante a preservação da topologia do objeto como, por exemplo, descontinuidades de superfícies.

7.3.5. Chen & Medioni

O trabalho de Chen e Medioni, 1992, aborda o problema da obtenção de um modelo para objetos através da correspondência de múltiplas imagens de profundidade. Para determinar os parâmetros de transformação (três ângulos de rotação e um vetor de translação) entre duas imagens é utilizado um método iterativo, onde o número de pontos de controle cresce de acordo com o número de iterações realizadas, de modo a reduzir a complexidade de cada iteração. O método tem por objetivo minimizar as distâncias entre pontos nas duas imagens, sendo que os pontos de controle utilizados não necessitam representar características significativas das superfícies dos objetos. Uma vez que os parâmetros da correspondência entre as imagens são obtidos, estas são integradas para formar um modelo completo do objeto. Esta etapa de integração das imagens é pouco discutida no trabalho de Chen e Medioni.

7.3.6. Sapidis & Besl

O trabalho de Sapidis e Besl, 1995, apresenta um método para construção de superfícies polinomiais a partir de imagens de profundidade que utiliza uma extensão do algoritmo de crescimento de regiões apresentado em Besl e Jain, 1988. Neste trabalho, a imagem de profundidade é considerada como um conjunto de quadriláteros, cada um formado por quatro dos pontos visualizados. Cada quadrilátero é visto como a conjunção de dois triângulos. O algoritmo de crescimento de regiões, a partir de uma região-semente escolhida interativamente, vai incorporando triângulos vizinhos enquanto a distância entre os pontos incorporados e a função da superfície ajustada permaneça abaixo do limiar estipulado. Desta maneira, o problema da continuidade das superfícies obtidas, que ocorre em Besl e Jain, 1988, é solucionado. Com isso, é possível criar uma representação para os objetos visualizados adequada à integração com sistemas de CAD/CAM, que é o objetivo do trabalho. Como limitações deste trabalho está ainda a grande dependência de diversos valores rígidos de limiar e a escolha interativa das regiões-semente.

7.4. Usando a representação desenvolvida

Para criar um modelo 3-D completo usando o método de representação descrito no Capítulo 6, também é necessário o casamento de diversas imagens de profundidade. Como solução para esta questão, propõe-se aplicar para cada

imagem o processo de representação desenvolvido. A correspondência, então, deve ser realizada com base no conjunto de representações obtidas.

Como pode ser observado na Seção 7.2, o processo tradicional de correspondência entre duas vistas é baseado na determinação da transformação espacial entre as mesmas. Neste contexto, a existência de bordas de profundidade e/ou de orientação, obtidas no processo de segmentação, auxilia na determinação desta transformação. Estas bordas podem ser identificadas em vistas subsequentes, atuando como características de controle. Entretanto, o processo de correspondência de vistas não deve ser totalmente dependente da identificação de bordas. Por este motivo, deve-se aplicar métodos padrões para estimar a transformação, e utilizar a informação de bordas apenas para refinar o resultado obtido.

Além de conter informações de borda, outra vantagem do método desenvolvido é o seu menor custo computacional na geração das regiões, em relação à abordagem por triangulação, que é amplamente utilizada. Neste sentido, uma outra característica positiva da representação desenvolvida é que regiões poligonais planares são mais significativas do que triângulos, do ponto de vista de custo computacional necessário ao processo de correspondência. Entretanto, o cálculo das equações de planos para regiões em forma de triângulos são mais estáveis, o que reduz o problema de obtenção de superfícies contínuas.

Uma outra observação importante é que com triangulações é possível obter melhores resultados na reconstrução de objetos compostos por muitos detalhes. Entretanto, se o objetivo do método está voltado para aplicações de visão computacional, pode não ser necessário uma representação detalhada das superfícies dos objetos, mas apenas a identificação de suas características mais importantes.

Uma vez que a correspondência entre as diversas vistas do objeto esteja determinada, as mesmas devem ser integradas para formar o modelo 3-D completo. Na integração das representações obtidas, é importante criar uma representação única que inclua toda informação relevante sobre o modelo construído. A representação desenvolvida no Capítulo 6 pode ser diretamente utilizada para o modelo 3-D, sendo esta mais uma de suas vantagens.

Tomando como base a representação utilizada, propõe-se para a integração das mesmas um ajuste ou refinamento dos polígonos sobrepostos na etapa de correspondência entre as vistas. Mais explicitamente, espera-se que após a correspondência entre duas vistas os polígonos equivalentes estejam relativamente próximos. Deste modo, pretende-se analisar esta proximidade em termos de:

- (1) Coordenadas 3-D dos pontos pertencentes aos polígonos;
- (2) Equação do plano ajustado ao polígono.

Uma vez que dois polígonos P_1 e P_2 tenham sido identificados como correspondentes em duas representações R_1 e R_2 , relativas a duas vistas em

seqüência, deve-se criar um ou mais polígonos como resultado da associação dos mesmos. O número de polígonos resultante da integração de P_1 e P_2 depende do objetivo do sistema de visão, que pode ser minimizar o erro do ajuste de planos ou minimizar o número de regiões que irão compor o modelo. Deste modo, pode-se chegar a uma representação final completa similar à utilizada para cada vista.

Visando o reconhecimento 3-D, a representação obtida para o modelo completo pode ser transformada, por exemplo, para uma representação na forma de um grafo (Herman, 1986), cujos nodos representariam as primitivas (objeto, face, aresta, vértice) utilizadas. Neste caso, o reconhecimento de um objeto se dá através da correspondência entre o grafo relativo ao seu modelo construído e uma base de modelos em forma de grafos. Esta abordagem se mostra interessante para o tipo de representação desenvolvida. Entretanto, diversas outras abordagens podem ser utilizadas (Suetens et al., 1992; Kuan e Drazovich, 1986), devendo-se realizar um estudo profundo das mesmas quando do interesse em desenvolver um sistema para reconhecimento 3-D.

7.5. Sumário

Este capítulo discutiu o problema da construção de modelos 3-d completos para objetos a partir de várias vistas. No Capítulo 8, a seguir, são apresentados os resultados experimentais para os processos de segmentação e reconstrução.

Capítulo 8: Resultados experimentais

8.1. Introdução

Neste capítulo, são apresentados diversos resultados experimentais que ilustram a aplicabilidade do método de segmentação/representação desenvolvido. Inicialmente, na Seção 8.2, são apresentados alguns detalhes de implementação e obtenção das imagens. A seguir, na Seção 8.3, são apresentados resultados visuais experimentais decorrentes da aplicação do processo de segmentação a um conjunto de imagens de profundidade sintéticas, seguidos dos resultados visuais da representação para os exemplos de segmentação mostrados. A Seção 8.4 contém uma tabela que relaciona as medidas de erro desenvolvidas para quantificar os resultados obtidos. Na Seção 8.5 são mostrados dois exemplos numéricos relativos à estrutura obtida para duas imagens de profundidade após o processo de representação. Na Seção 8.6, apresentam-se os resultados visuais da aplicação dos métodos de segmentação e representação a uma imagem de profundidade de um objeto um pouco mais complexo. Na Seção 8.7, é mostrado um exemplo de aplicação em uma imagem real. A Seção 8.8 mostra um exemplo dos resultados visuais obtidos para a reconstrução da imagem de profundidade (Figura 8.2), plotados em 3-D. Finalmente, na Seção 8.9, são mostrados os resultados visuais de segmentação e reconstrução para uma imagem real.

8.2. Detalhes de implementação

As duas primeiras imagens de profundidade sintéticas utilizadas, mostradas nas Figuras 8.1 e 8.2, foram obtidas utilizando-se um software desenvolvido pelo Grupo ProSIm (DCA-FEEC-UNICAMP), baseado em *raytrace*. Estas imagens estão na forma r_{ij} , e foram gerados arquivos correspondentes com as coordenadas 3-D de cada *pixel*.

Para ilustrar a aplicabilidade do método de segmentação/representação em imagens ruidosas, aplicou-se um ruído Gaussiano (média = 0 e variância = 0.01) nas imagens geradas sinteticamente. Os resultados obtidos são mostrados no decorrer deste capítulo. O valor utilizado para a variância é baseado na amplitude dos valores das coordenadas (x,y,z) das imagens geradas. Este ruído é aplicado a cada uma das componentes (x,y,z) da imagem. Deste modo, o ruído final aplicado às coordenadas 3-D de cada *pixel* da imagem de profundidade têm variância superior a 0.01.

Os objetos utilizados (uma esfera e um paralelepípedo) foram escolhidos de modo a mostrar a aplicabilidade da segmentação e da representação, tanto em objetos poliédricos quanto em não poliédricos.

Para a realização dos testes, foram utilizadas diversas ferramentas do pacote KHOROS para processamento de imagens. Este pacote também foi utilizado como suporte para todo o software desenvolvido.

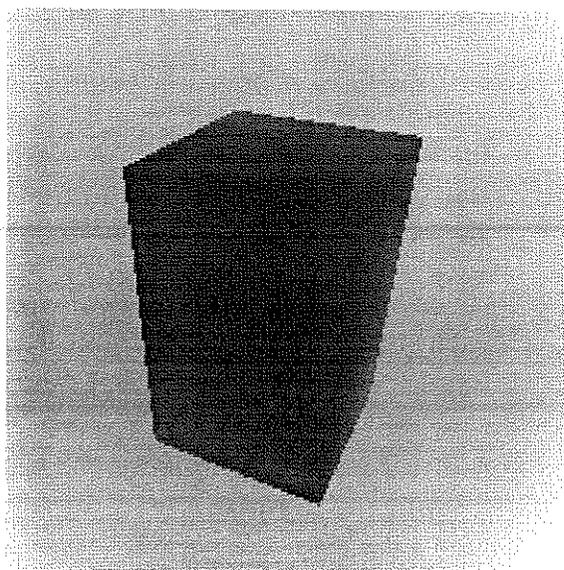


Figura 8.1: Imagem de profundidade, um paralelepípedo

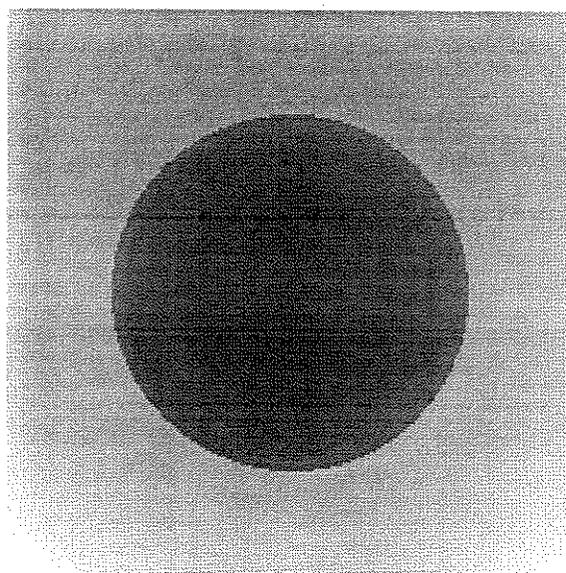


Figura 8.2: Imagem de profundidade, uma esfera.

8.3. Resultados visuais

As figuras 8.3-8.12 ilustram os resultados visuais obtidos para a segmentação das imagens de profundidade mostradas nas Figuras 8.1 e 8.2. Na segmentação das imagens adicionadas de ruído, observa-se algumas distorções na posição dos vértices, em comparação com as mesmas imagens sem ruído. Entretanto, devido ao refinamento das coordenadas 3-D dos vértices pela interseção das regiões vizinhas, esta distorção não chega a comprometer os resultados obtidos.

As figuras 8.13-8.22 ilustram os resultados visuais obtidos para a representação dos resultados de segmentação mostrados nas Figuras 8.3-8.12. Com base nestas figuras, pode-se notar que a representação permanece robusta mesmo quando o resultado da segmentação é degradado por ruído. A Figura 8.19 mostra uma pequena incoerência na visualização do plano superior esquerdo. Pode-se afirmar que este resultado é devido não apenas ao ruído incorporado, o que degrada o processo de ajuste de planos utilizado, mas também ao problema da discretização quando da obtenção das coordenadas de imagem para o ponto.

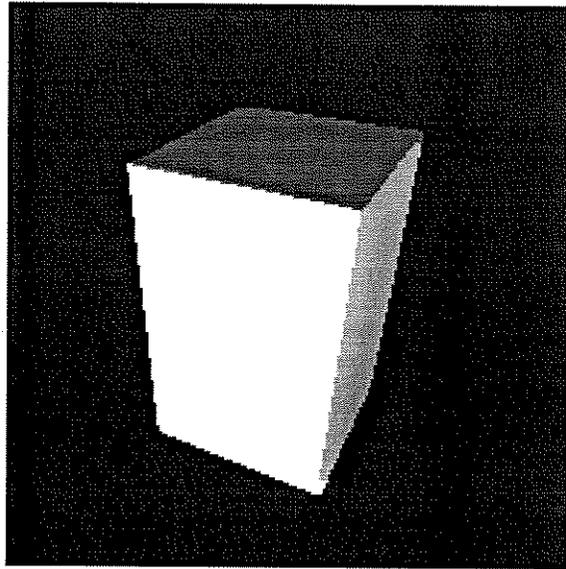


Figura 8.3: Resultado da segmentação para o paralelepípedo sem ruído.

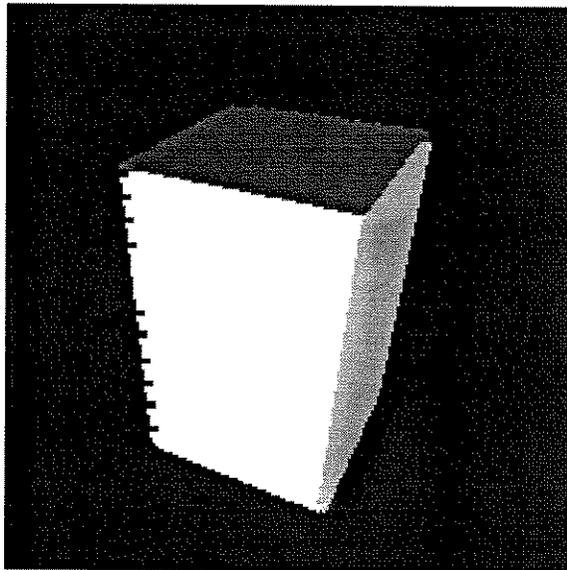


Figura 8.4: Resultado da segmentação para o paralelepípedo com ruído.

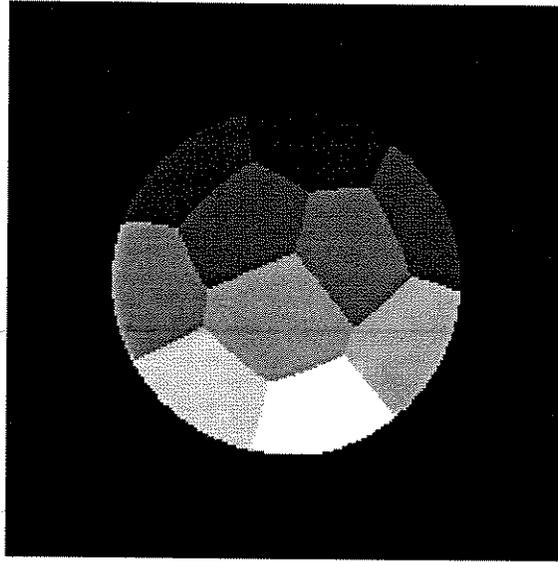


Figura 8.5: Resultado da segmentação para esfera sem ruído, $K = 10$.

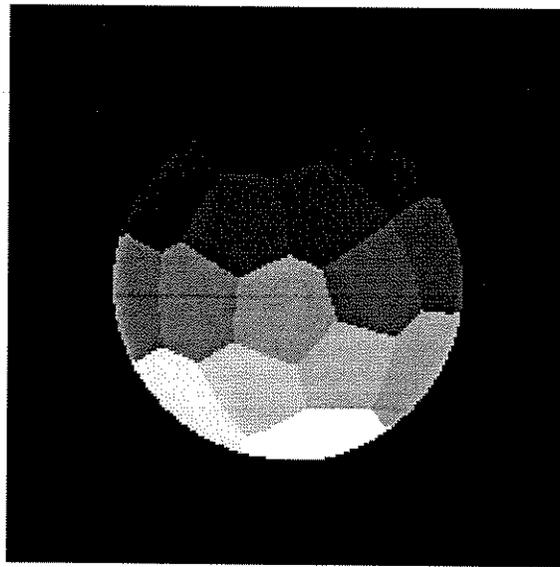


Figura 8.6: Resultado da segmentação para esfera sem ruído, $K = 15$.

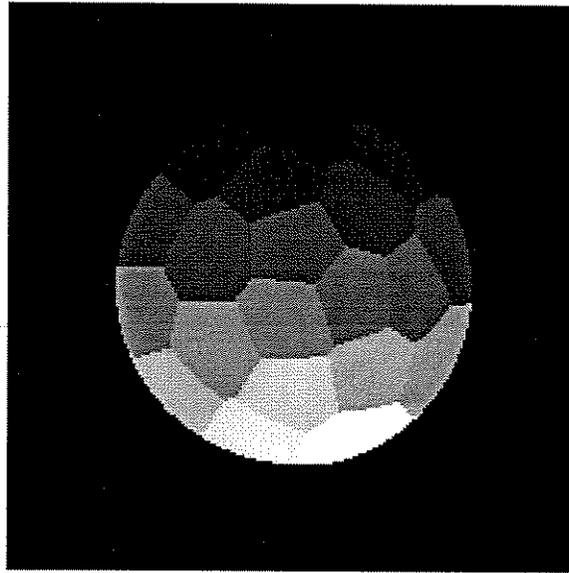


Figura 8.7: Resultado da segmentação para esfera sem ruído, $K = 20$.

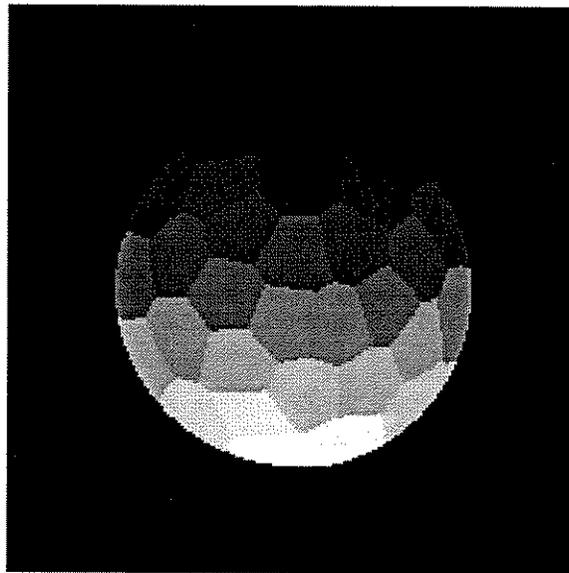


Figura 8.8: Resultado da segmentação para esfera sem ruído, $K = 30$.

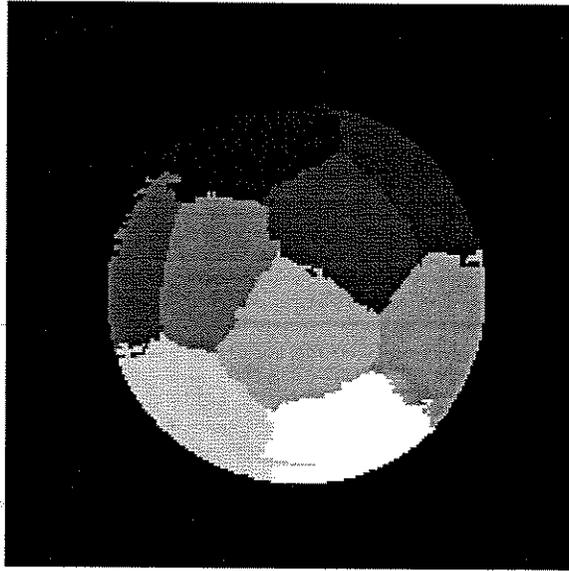


Figura 8.9: Resultado da segmentação para esfera com ruído, $K = 10$.

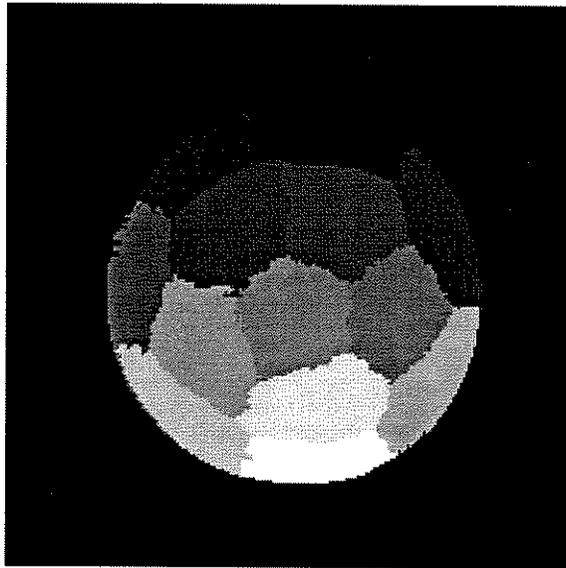


Figura 8.10: Resultado da segmentação para esfera com ruído, $K = 15$.

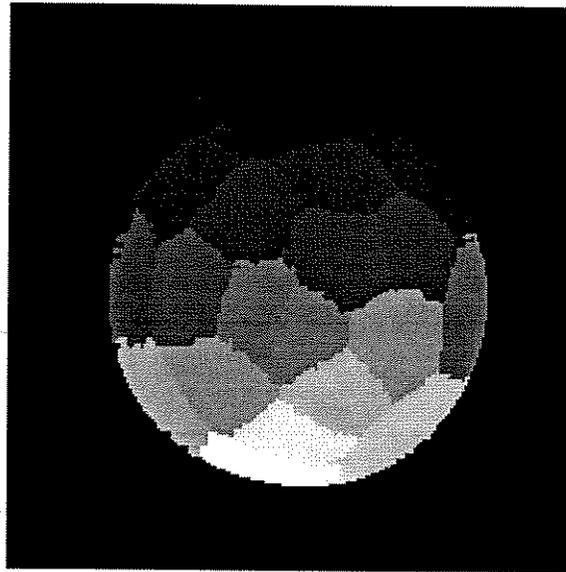


Figura 8.11: Resultado da segmentação para esfera com ruído, $K = 20$.

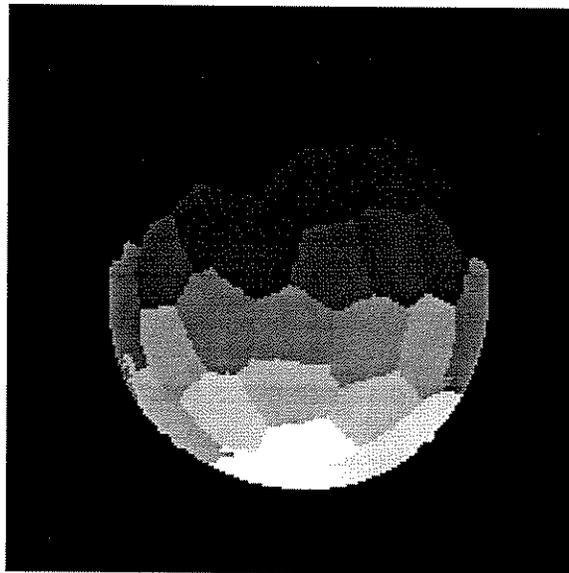


Figura 8.12: Resultado da segmentação para esfera com ruído, $K = 30$.

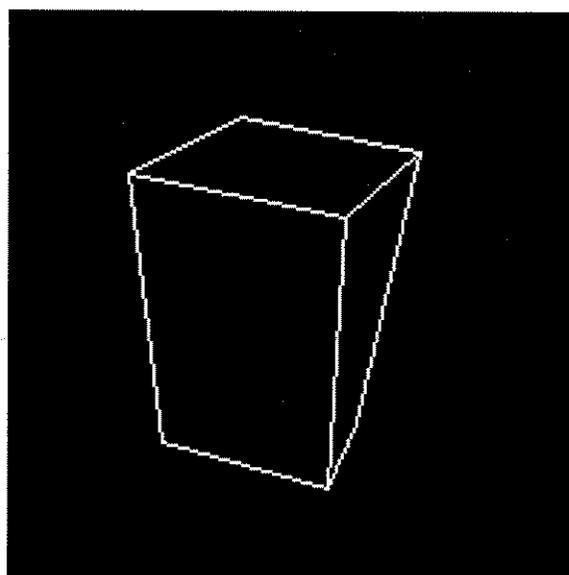


Figura 8.13: Resultado da representação para retângulo sem ruído.

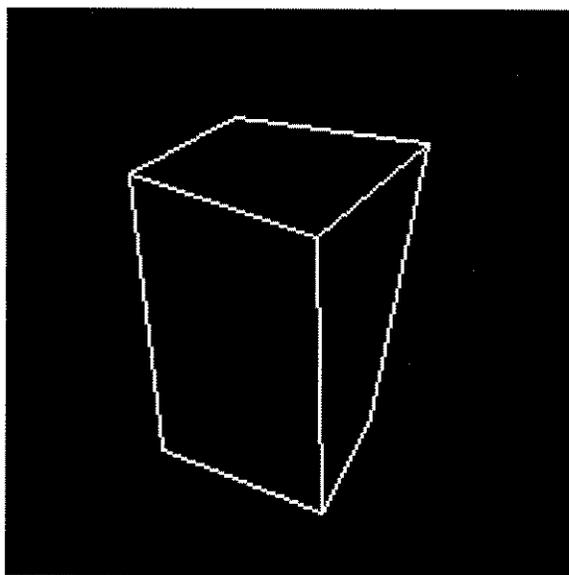


Figura 8.14: Resultado da representação para retângulo com ruído.

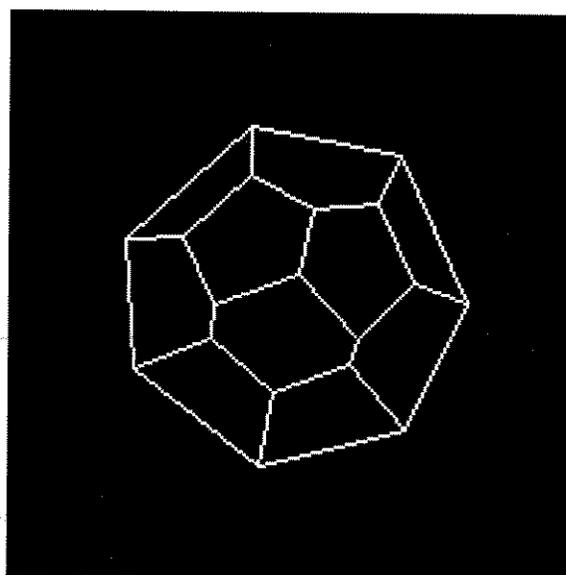


Figura 8.15: Resultado da representação para esfera sem ruído, $K = 10$.

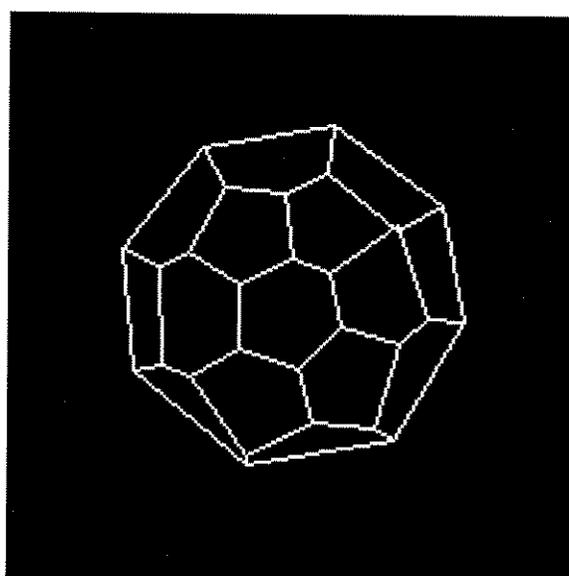


Figura 8.16: Resultado da representação para esfera sem ruído, $K = 15$.

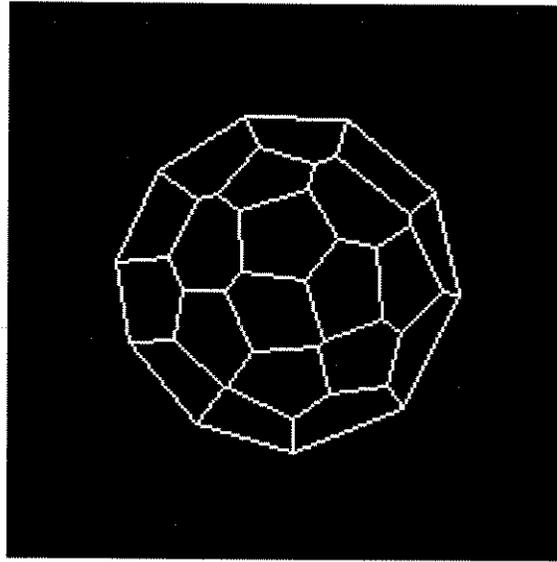


Figura 8.17: Resultado da representação para esfera sem ruído, $K = 20$.

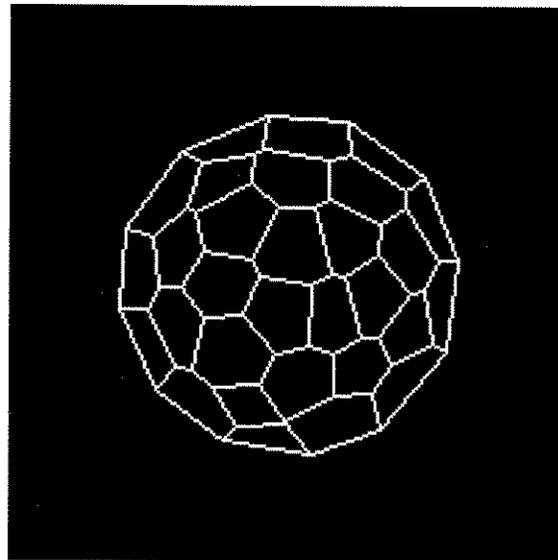


Figura 8.18: Resultado da representação para esfera sem ruído, $K = 30$.

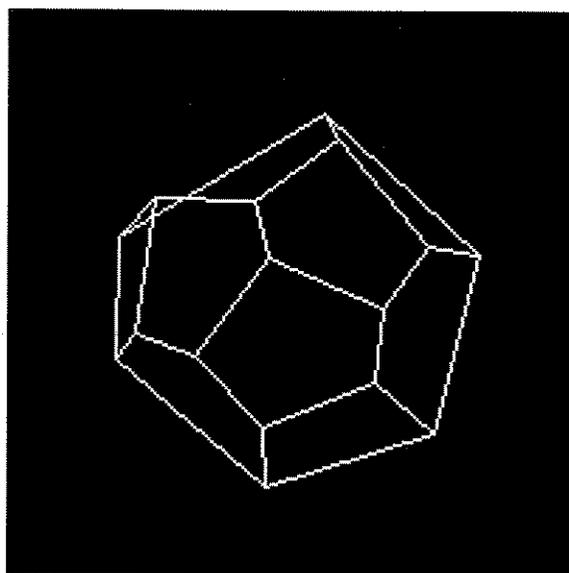


Figura 8.19: Resultado da representação para esfera com ruído, $K = 10$.

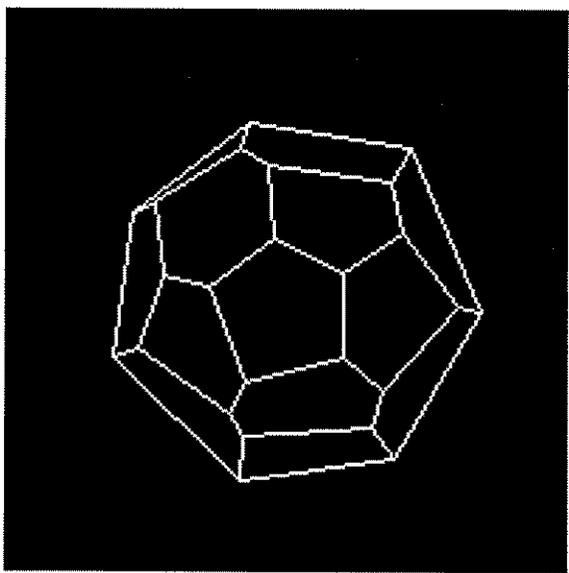


Figura 8.20: Resultado da representação para esfera com ruído, $K = 15$.

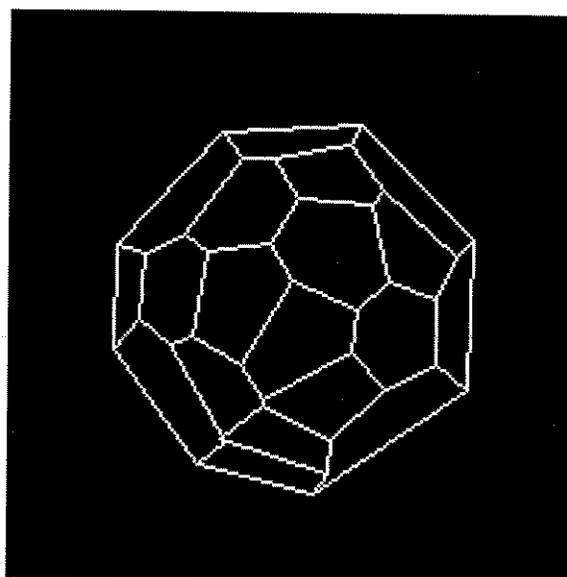


Figura 8.21: Resultado da representação para esfera com ruído, $K = 20$.

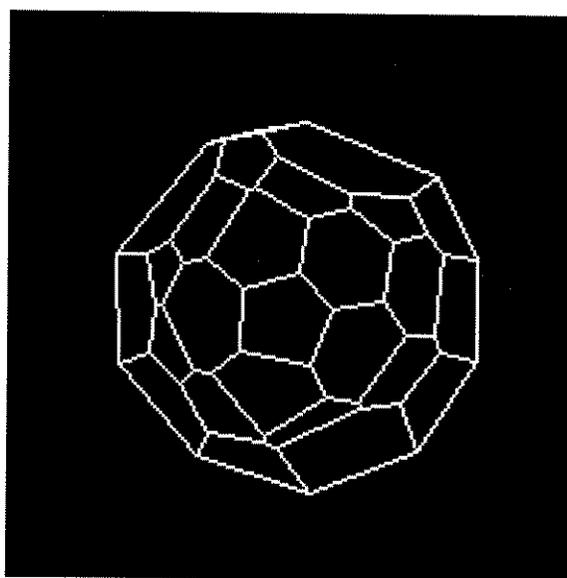


Figura 8.22: Resultado da representação para esfera com ruído, $K = 30$.

8.4. Tabela comparativa

Esta seção mostra uma tabela comparativa para as medidas de erros obtidas a partir da aplicação do método de segmentação/representação desenvolvido nos casos mostrados nas seções anteriores.

Tabela 8.1: Comparação das medidas de erro.

Imagem	ϵ_{ga}	ϵ_{gr}	ϵ_v	ϵ_v/nv
Paralelepípedo sem ruído	0.000072	0.119165	0.070175	0.01
Paralelepípedo com ruído	0.014901	1.564282	9.496048	1.36
Esfera sem ruído $K=10$	0.001574	0.610002	134.05811	6.67
Esfera sem ruído $K=15$	0.001574	0.455996	120.63270	4.79
Esfera sem ruído $K=20$	0.001574	0.324371	97.755966	2.57
Esfera sem ruído $K=30$	0.001574	0.225245	76.374603	1.32
Esfera com ruído $K=10$	0.006114	2.000572	550.27541	28.34
Esfera com ruído $K=15$	0.006114	1.516631	400.29523	16.68
Esfera com ruído $K=20$	0.006114	1.208808	262.17148	7.28
Esfera com ruído $K=30$	0.006114	0.978072	272.94876	5.45

Notação:

ε_{ga} = erro global de aproximação

ε_{gr} = erro global de reconstrução

ε_v = erro nas coordenadas 3D dos vértices

nv = número de vértices

Analisando-se os resultados da Tabela 8.1 observa-se que, surpreendentemente, a somatória dos erros nas coordenadas 3-D dos vértices é maior no caso da esfera dividida em 30 segmentos do que na esfera com 20 segmentos. Isto se deve ao problema da estabilidade no cálculo das equações dos planos, pois pequenas regiões (nas imagens em questão) apresentam as valores muito próximos para as coordenadas 3D de seus pontos componentes. Para solucionar este problema, métodos mais robustos vem sendo estudados (Yu et al., 1994).

Outra observação experimental é que o erro global de aproximação, no caso de imagens sem ruído de objetos poliédricos, é tão pequeno (próximo de zero) que não é adequado para uso como medida de qualidade da segmentação, conforme definido na Seção 5.7. Neste caso, este erro pode ser entendido como devido apenas a problemas numéricos.

8.5. Exemplo de estrutura obtida

Nesta seção, apresenta-se como exemplo a estrutura obtida para o paralelepípedo da Figura 8.1.

Lista de Vértices:

(Rótulo vértice, Coordenadas 3D, Rótulo regiões vizinhas, Coordenadas 2D)

0	-34.50	-20.15	-45.00	1	2	4	42	57
1	-10.00	-59.05	-47.91	1	3	4	113	168
2	-11.72	-20.41	-45.01	2	3	4	118	72
3	-10.00	-57.09	-66.71	0	1	3	123	144
4	-34.01	-58.54	-45.00	0	1	4	54	151
5	-33.42	-20.00	-68.06	0	1	2	81	37
6	-10.48	-20.00	-67.86	1	2	3	145	49

Lista de Arestas:

(Rótulo aresta, Rótulo regiões que divide, Rótulo vértices componentes)

0	0	2	2	4
1	1	2	3	4
2	2	3	2	3
3	3	4	1	3
4	0	5	1	4
5	4	1	1	3
6	0	6	1	2
7	5	1	1	4
8	6	3	1	2

Lista de Regiões:

(Rótulo região, Coeficientes plano adequado, Rótulo vértices)

1	0.366	0.500	0.785	107.752	3	4	1	5	0	6
2	0.000	1.000	0.000	20.408	0	2	3	6		
3	0.999	0.000	0.002	11.796	1	2	3	4		
4	0.000	0.000	1.000	45.007	0	2	1	5		

8.6. Um exemplo mais complexo

Nesta seção, apresenta-se um conjunto de figuras (8.23-8.26) que são os resultados visuais obtidos com a aplicação dos métodos de segmentação e representação desenvolvidos numa imagem de um objeto sintético mais complexo, obtida da Michigan State University.

O objeto mostrado foi escolhido de modo a ilustrar como a informação sobre as bordas soluciona o problema de regiões vizinhas de mesma orientação, mas com profundidades diferentes. Para este exemplo, o processo de poligonalização é uma etapa crucial na solução do problema, pois existe um grande número de bordas curvas no objeto.

A Figura 8.23 é a imagem de profundidade de uma vista do objeto. A Figura 8.24 é a imagem resultante após a sobreposição do mapa de bordas à imagem obtida pelo processo de agrupamento, conforme descrito na Seção 5.9. Após o refinamento da imagem anterior, obtém-se uma imagem segmentada final, que é mostrada na Figura 8.25. A Figura 8.26 é a representação obtida para a segmentação da Figura 8.25.

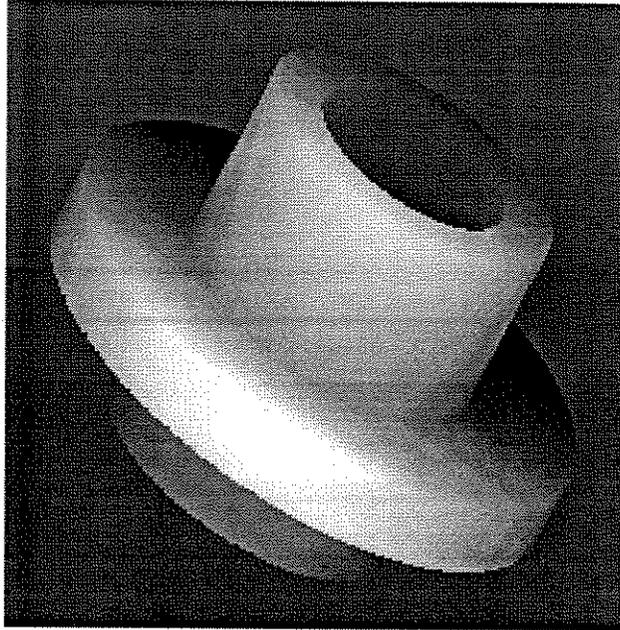


Figura 8.23: Imagem de profundidade.

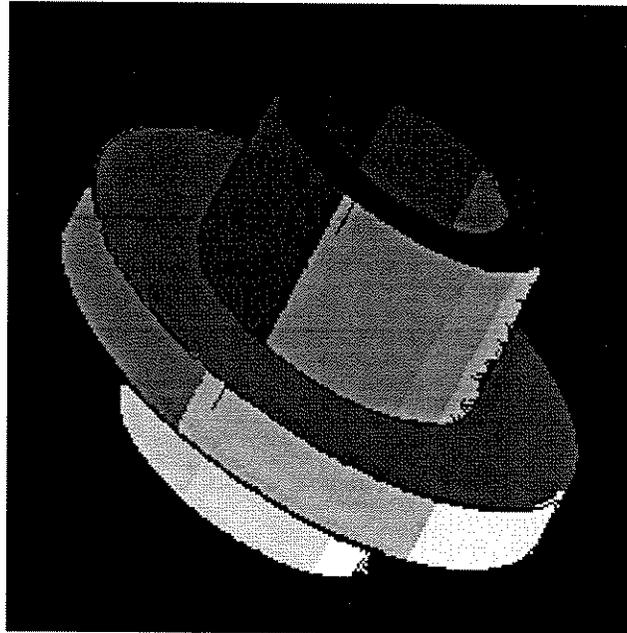


Figura 8.24: Resultado do agrupamento.

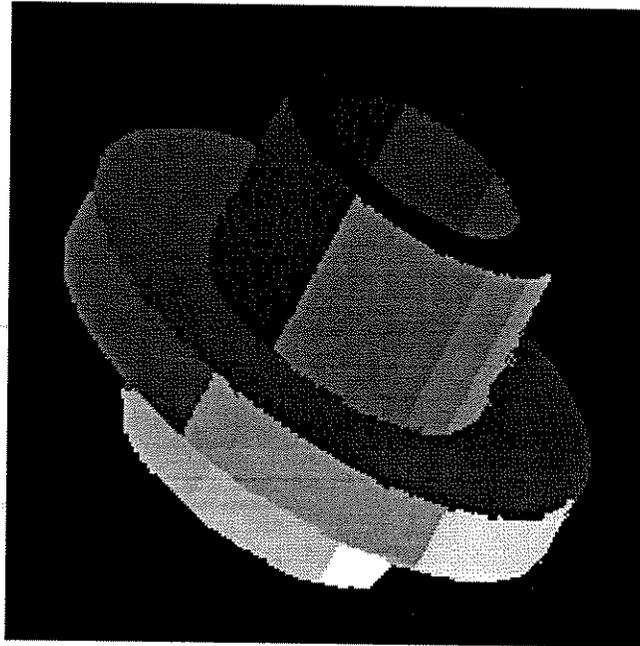


Figura 8.25: Imagem segmentada.

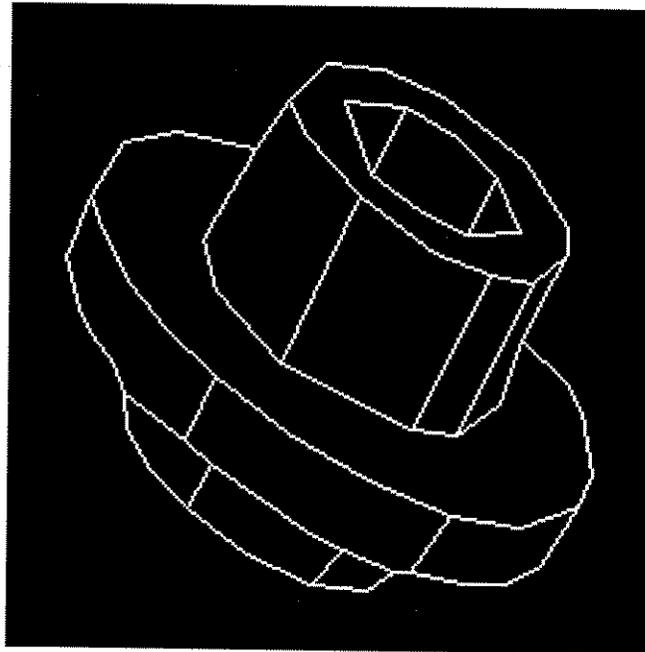


Figura 8.26: Representação do objeto.

8.7. Um exemplo real

Esta seção mostra os resultados visuais da aplicação dos métodos de segmentação e representação desenvolvidos em uma imagem real, obtida da University of South California. A Figura 8.27 é a imagem de profundidade de uma vista de um objeto poliédrico. O objeto foi escolhido de modo contemplar o problema de múltiplas faces convergindo para um único vértice. A Figura 8.28 é o resultado da segmentação para a imagem da Figura 8.27. Finalmente, na Figura 8.29, é mostrado o resultado obtido para a representação do objeto.

A maior dificuldade encontrada no processamento da imagem real foi a sua grande quantidade de ruído. Isto acarretou problemas no cálculo dos coeficientes da normal e, conseqüentemente, na obtenção da correta segmentação da imagem. Para solucionar estes problemas, as imagens com os dados sobre as coordenadas 3-D da cena observada sofreram um processo de suavização.

Na suavização, foram aplicados filtros que utilizam a Média e a Mediana, além de um filtro Gaussiano. Os resultados foram analisados para janelas de tamanho 3x3, 5x5, e 7x7. A suavização pelo filtro que utiliza a Mediana, com janela 5x5, forneceu o melhor resultado. Para obter resultados satisfatórios, foi necessário aplicar este filtro sequencialmente por pelo menos 5 vezes nas imagem xyz.

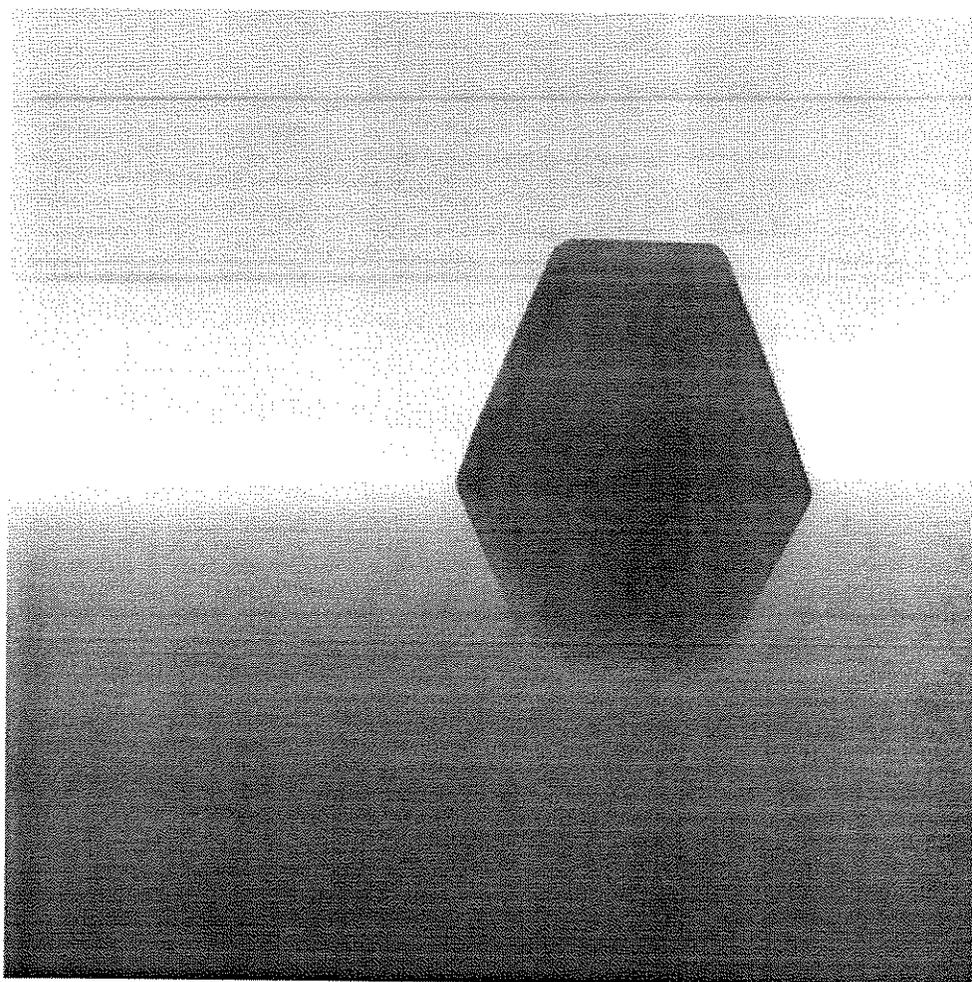


Figura 8.27: Imagem de profundidade real de um objeto poliédrico.

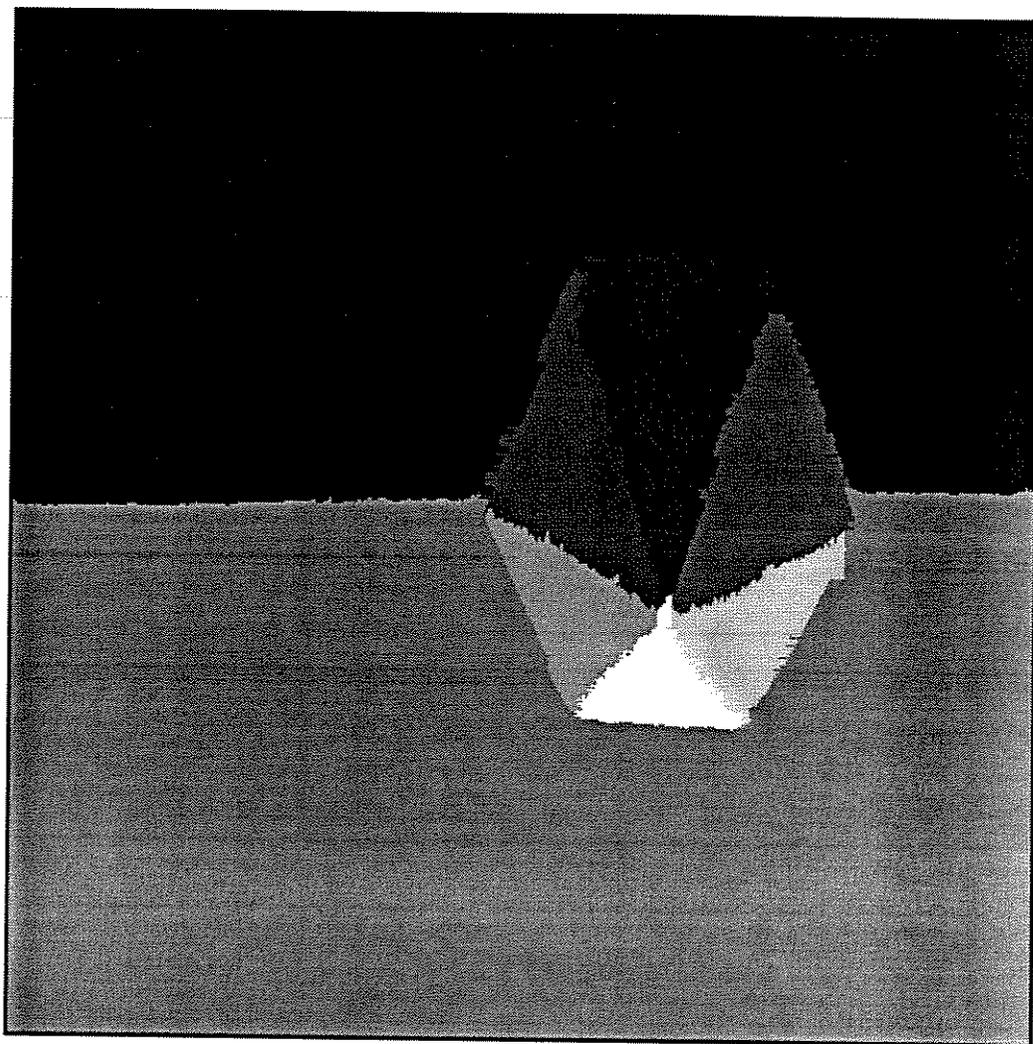


Figura 8.28: Resultado da segmentação.

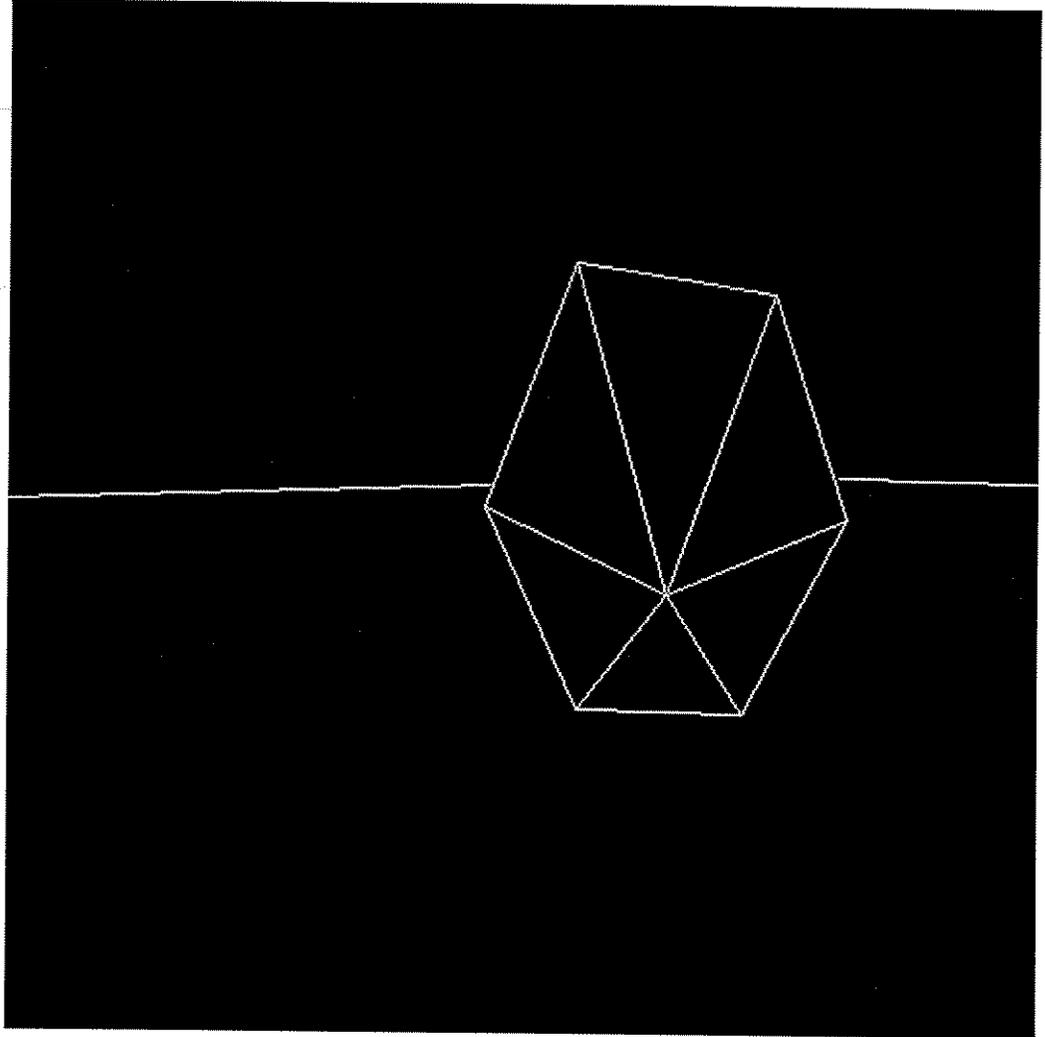


Figura 8.29: Resultado da representação.

8.8. Resultados visuais para a reconstrução 3-D

As Figuras 8.30 a 8.34 ilustram os resultados da reconstrução, obtidos para a imagem da esfera mostrada na Figura 8.2, plotados em 3-D. A Figura 8.30 é a imagem original, e as Figuras 8.31 a 8.34 mostram a reconstrução da imagem original resultantes da segmentação desta em 10, 15, 20 e 30 regiões, respectivamente. Neste conjunto de figuras, fica claro o aumento da qualidade da reconstrução com o aumento do número de regiões, para o caso de um objeto não-poliédrico.

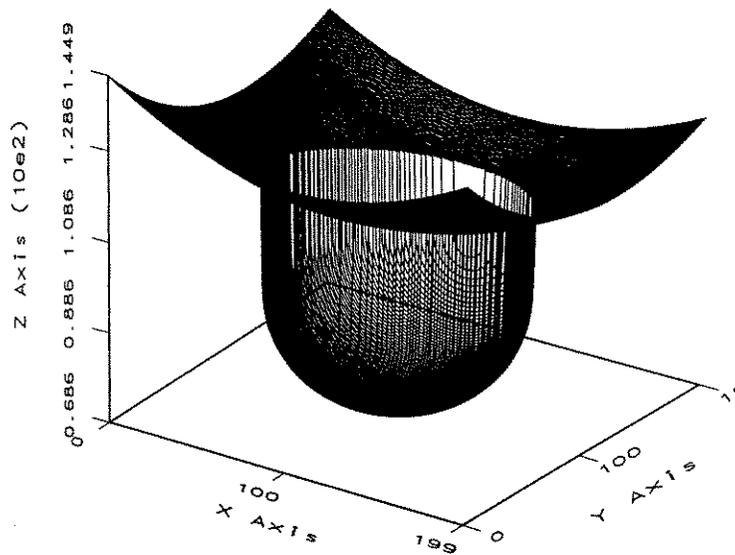


Figura 8.30: Esfera original plotada em 3-D.

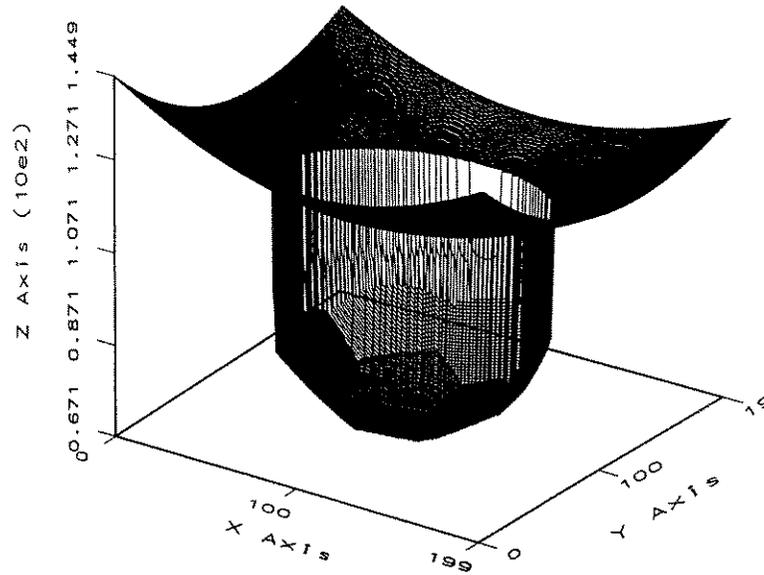


Figura 8.31: Esfera reconstruída a partir de 10 planos plotada em 3-D.

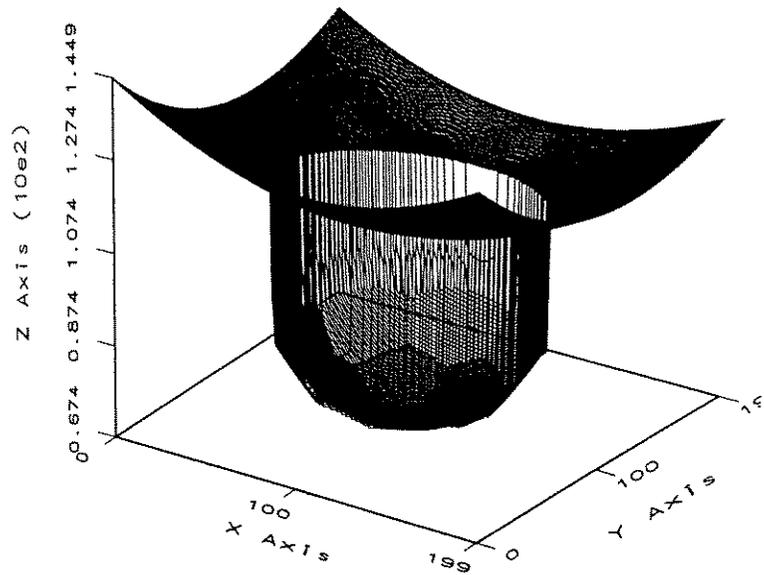


Figura 8.32: Esfera reconstruída a partir de 15 planos plotada em 3-D.

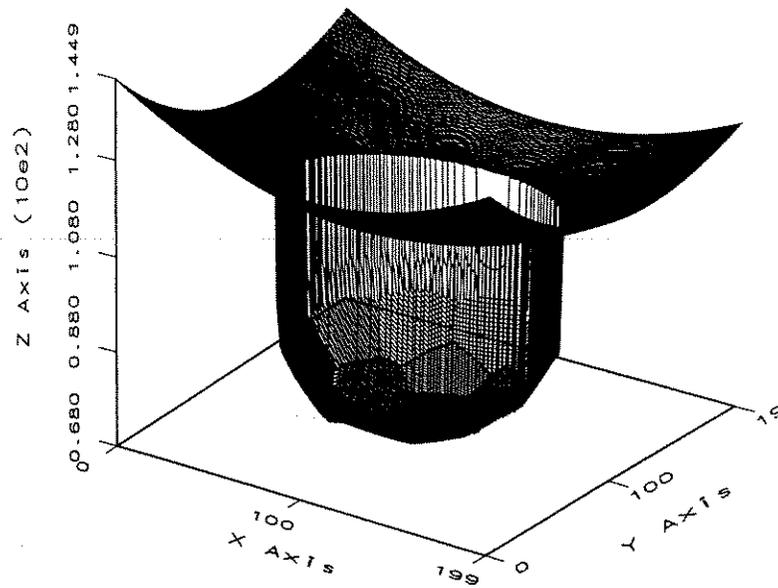


Figura 8.33: Esfera reconstruída a partir de 20 planos plotada em 3-D.

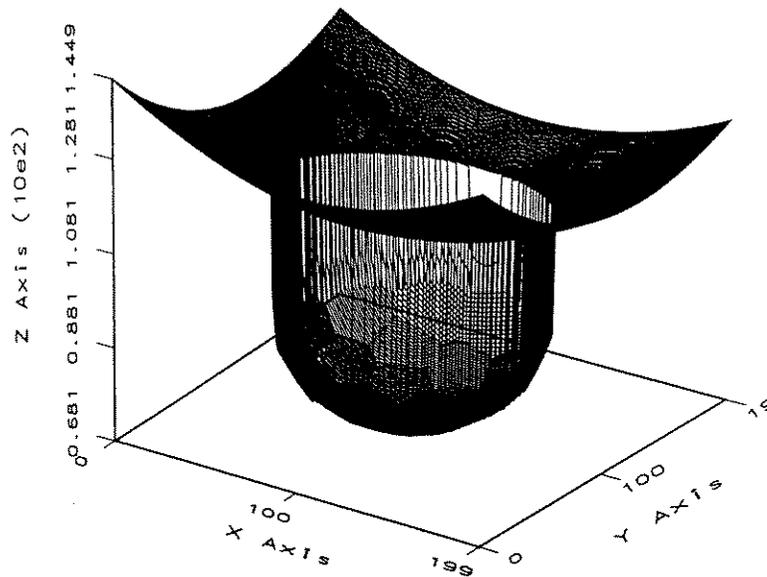


Figura 8.34: Esfera reconstruída a partir de 30 planos plotada em 3-D.

Capítulo 9: Conclusão e trabalhos futuros

9.1. Conclusão

Neste trabalho, foi apresentado um método integrado para segmentação e representação em imagens de profundidade. Este tipo de integração permite que resultados discrepantes na etapa de representação atuem como uma realimentação na etapa anterior de segmentação, melhorando o desempenho global do método. No entanto, a despeito de sua importância, esta integração tem sido pouco contemplada na literatura.

O processo de segmentação é baseado na associação da técnica de agrupamento *K-Means* com técnicas de detecção de bordas. A técnica de agrupamento é pouco dependente de valores rígidos de limiar, mas tem como limitação a dependência do conhecimento do número K de superfícies que compõem a imagem. A associação de técnicas de detecção de bordas com a técnica de agrupamento permite gerar uma estimativa inicial, não necessariamente precisa, para este número K . Uma vantagem do método desenvolvido é que esta estimativa pode ser atualizada com os resultados de reconstrução obtidos para as superfícies. Neste caso, o algoritmo de agrupamento é novamente aplicado, agora para um novo número K inicial de superfícies. A associação de técnicas de detecção de bordas ao método implementado, embora introduza sensibilidade a valores de limiar, tem ainda as seguintes vantagens:

- (1) Permite eliminar o efeito dos *pixels* de borda, que contém valores de profundidade discrepantes, gerando uma estimativa mais realista para as medidas de qualidade obtidas para a segmentação.
- (2) Contribui para a determinação da representação ou modelo dos objetos que compõem a imagem, pois bordas são características geométricas altamente significativas.

O algoritmo de agrupamento aplicado utiliza como característica de entrada um vetor com três componentes. Estas componentes são os coeficientes da normal a cada ponto da imagem, que são constantes em regiões planares. Assim, o resultado da segmentação é um conjunto de segmentos que podem ser bem ajustados por regiões planares.

Tendo em vista o conjunto de regiões planares obtidas na segmentação, optou-se por construir uma representação poliédrica para os objetos segmentados. Assim, as regiões segmentadas devem ser transformadas em polígonos, e suas interseções devem ser segmentos de retas. Após pequenos refinamentos necessários nas regiões segmentadas para que estas atendam às condições de formar uma superfície poliédrica, obtem-se uma representação simples e robusta, isto é, sem discontinuidades ou *cracks*. Esta representação, a princípio, está relacionada apenas às superfícies visíveis do objeto na imagem. No entanto, foi mostrado que a mesma representação pode ser aplicada também para a construção de modelos 3-D completos.

Em comparação com as representações baseadas em triangulação, que vêm sendo amplamente utilizadas, a representação por polígonos desenvolvida tem como principais vantagens sua maior simplicidade de implementação e seu menor custo computacional. Além disso, polígonos são entidades mais significativas do que triângulos, sendo por isso mais adequados para utilização em reconhecimento 3-D de objetos, do ponto de vista de custo computacional na etapa de casamento. Entretanto, na abordagem por triangulação não existe o problema de estabilidade na determinação das equações de planos, que é encontrado na abordagem por polígonos.

Além da vantagem de se utilizar uma abordagem integrada para solucionar os problemas de segmentação e representação, outras vantagens do método desenvolvido também podem ser ressaltadas. Destacam-se:

- (1) O desenvolvimento de uma nova abordagem, baseada no erro de reconstrução, para determinar o número adequado de regiões para a imagem.
- (2) A existência de medidas de qualidade significativas para o método permite que seus resultados sejam avaliados efetivamente.
- (3) Embora o método crie uma representação poliédrica para os objetos segmentados, sua aplicação não se restringe a objetos poliédricos e pode gerar boas aproximações também para objetos não-poliédricos. Além disso, caso necessário, pedaços de superfícies de maior grau

podem ser obtidas através da correta união de pedaços de superfícies planares.

(4) O método é adequado à interpretação de imagens pois, tomando-se diversas vistas de um mesmo objeto, pode-se gerar uma representação para cada vista e criar uma representação similar para um modelo 3-D completo pelo casamento das representações obtidas e posterior integração das informações.

As vantagens acima descritas vão de encontro a solucionar as dificuldades apontadas no Capítulo 1.

Em relação à escolha do método mais adequado para segmentação, o Capítulo 4 apresenta diversos indícios para a solução do problema, bem como indica vários trabalhos relacionados a serem consultados. Para representação, vários métodos são discutidos exaustivamente em Mäntyla, 1988, e Hoffman, 1989.

A definição do critério de homogeneidade e a determinação do número correto de regiões da imagem são questões estreitamente relacionadas. Como pode ser observado no Capítulo 4, muito esforço vêm sendo aplicado para dotar os sistemas de visão de métodos mais automáticos em relação a estes critérios. No entanto, a solução definitiva ainda está longe de ser alcançada.

A obtenção de técnicas robustas têm sido uma preocupação atual constante na maioria dos trabalhos sobre segmentação. A saída está na aplicação de ferramentas estatísticas, tais como eliminação de amostras incoerentes com o conjunto de dados a serem tratados.

Como indicação para a obtenção de estruturas de representação coerentes, mesmo na presença de resultados de segmentação imprecisos, uma saída é realimentar os resultados da representação na segmentação. Este trabalho procurou atuar neste sentido.

Como forma de reduzir o custo do processo de correspondência entre as representações obtidas e os modelos previamente armazenados, a solução óbvia é incluir características altamente discriminatórias na representação.

De uma forma geral, os problemas de uma determinada aplicação estão relacionados aos objetivos específicos do sistema de visão e às restrições apresentadas.

9.2. Trabalhos futuros

Como trabalhos futuros, os seguintes temas podem ser explorados:

- (1) A determinação de uma medida de qualidade mais precisa para a segmentação e a representação, baseada nas diferenças entre a superfície 3-D reconstruída e a original. Isto pode ser feito utilizando a

Lista de regiões obtida na representação como entrada em um processo de *ray-casting* (Foley et al., 1996), que irá criar uma imagem adequada para comparação com a original.

(2) Utilização de regiões que podem ser bem aproximadas por quádricas:

Seguindo o mesmo princípio utilizado para segmentação por planos, regiões quádricas podem ser obtidas de duas maneiras:

- Realizando-se a correta união de regiões planares: Neste caso, a maior dificuldade está em como agregar corretamente os segmentos planares.
- Realizando-se o agrupamento usando como característica os coeficientes da quádrica mais adequada a cada *pixel*, ao invés do vetor normal: O maior problema associado com esta abordagem está na obtenção confiável dos parâmetros quádricos, além do aumento do custo computacional por usar um vetor de características de maior grau.

(3) Representação por quádricas sem descontinuidades: Uma vez que se obtenha uma segmentação por quádricas, é necessário criar uma representação robusta para as regiões definidas. Neste caso, as interseções entre regiões não são mais retas e sim curvas, que devem ser bem delimitadas, e a representação final se torna mais complexa.

(4) Construção de um modelo 3-D completo: A construção de um modelo 3-D completo é etapa necessária para aplicação tanto em visualização quanto em reconhecimento 3-D de objetos. Neste caso, a dificuldade observada está na realização do casamento entre as diversas vistas necessárias ao processo, e na posterior integração destas vistas numa superfície contínua.

(5) Reconhecimento 3-D de objetos: De posse de um modelo 3-D completo, construído para um determinado objeto, é necessário realizar seu casamento com algum modelo previamente armazenado em uma base de dados. O modelo construído pode conter inconsistências, em decorrência de erros cometidos durante as diversas etapas necessárias para sua geração. Assim, o método de reconhecimento utilizado deve prover recursos suficientes para se chegar à alternativa mais confiável dentre as possibilidades existentes, e deve fornecer uma medida para esta confiabilidade. A maior dificuldade associada com o problema do reconhecimento 3-D de objetos está no desenvolvimento de um processo inteligente para realizar a busca na base de dados que contém os modelos dos objetos.

(6) A integração completa dos métodos de segmentação e de representação: Esta integração permite a realimentação dos resultados da representação no processo de segmentação, melhorando o desempenho global dos métodos.

Referências

- N. Abdelmalek, 1990. "Algebraic error analysis for surface curvatures and segmentation of 3D range images". **Pattern Recognition**, Vol.23, No.8, pp.807-817.
- M.A. Abidi e R.C. Gonzalez, 1992. **Data Fusion in Robotics and Machine Intelligence**. Academic Press.
- J.K. Aggarwal, 1990. "Segmentation and analysis of multi-sensor images". **Machine Vision for Three-Dimensional Scenes**. Academic Press, Inc. pp.267-299.
- J.K. Aggarwal e C.H. Chien, 1989. "3-D structures from 2-D images". **Advances in Machine Vision**. Springer Verlag, New York, pp.64-120.
- F. Arman e J.K. Aggarwal, 1993. "Model-based object orientation in dense-range images - a review". **ACM Computing Surveys**, Vol.25, No.1, pp.5-43.
- H. Asar, N. Nandhakumar e J.K. Aggarwal, 1990. "Pyramid-based image segmentation using multisensory data". **Pattern Recognition**, Vol.23, No.6, pp.583-593.
- D.H. Ballard e C.M. Brown, 1982. **Computer Vision**, Prentice Hall Englewood Cliffs, New Jersey.

- B.A. Barsky, 1984. "A description and evaluation of various 3-D models". **IEEE CG&A**, pp.38-52.
- B. Beckman, 1973. **Orthogonal polynomials for engineers and physicists**, The Golden Press Boulder, CO.
- O.R.P. Bellon e C.L. Tozzi, 1997. "An integrated approach for segmentation and representation of range images". Proc. of the **9th Int. Conf. on Image Analysis and Processing**, Vol. I, pp.279-286, Florença-Itália.
- O.R.P. Bellon, 1996. "Tutorial: Segmentação de imagens de profundidade". CD-ROM do **XIX Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens**, Caxambu-MG.
- O.R.P. Bellon, J.E.C. Castanho, e C.L. Tozzi, 1995. "A hybrid approach for solving the range image segmentation/reconstruction problem". Proc. of the **IEEE Int. Conf. on Systems, Man, and Cybernetics**, Vol.1, pp.252-257, Vancouver-Canada.
- O.R.P. Bellon, 1990. "Visão Computacional: Um Sistema para Localização de Objetos Poliédricos no Espaço 3-D", **Dissertação de Mestrado**, DCA-FEEC-UNICAMP.
- J. Beraldin, M. Rioux, F. Blais, L. Cournoyer e J. Domey, 1992. "Registered intensity and range imaging at 10 mega-samples per second". **Optical Engineering**, Vol.31, No.1, pp.88-94.

- P. Besl, 1989A. Active optimal range imaging sensors. **Advances in Machine Vision**. Springer Verlag, New York, pp.1-62.
- P. Besl, 1989B. **Surfaces in Range Image Understanding**. Springer Verlag, New York.
- P. Besl, 1986. "Invariant surface characteristics for 3D object recognition in range images". **Computer Vision, Graphics, and Image Processing**, Vol.33, pp.33-80.
- P. Besl e R. Jain, 1988. "Segmentation through variable-order surface fitting". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.10, No.2, pp.167-192.
- S. Bhandarkar e A. Siebert, 1992. "Integrating edge and surface information for range image segmentation". **Pattern Recognition**, Vol.25, No.9, pp.947-962.
- M. Brown, 1985. "Feature extraction techniques for recognizing solid objects with ultrasonic range sensor". **IEEE J. Robotics and Automation**, Vol.1, No.4, pp.191-205.
- Y. Chen e G. Medioni, 1992. "Object modelling by registration of multiple range images". **Image and Vision Computing**, Vol.10, No.3, pp.145-155.
- G.B. Coleman e H.C. Andrews, 1979. "Image segmentation by clustering". **Proc. of the IEEE**, Vol.67, No.5.

- R. Hoffman e A. Jain, 1987. "Segmentation and classification of range images".
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.9,
No.5, pp.608-620.
- A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K.
Bowyer, D.W. Eggert, A. Fitzgibbon, e R.B. Fisher, 1996. "An experimental
comparison of range image segmentation algorithms". **IEEE Transactions
on on Pattern Analysis and Machine Intelligence**, Vol.18, No.7, pp.673-
688.
- J. Illingworth e J. Kittler, 1988. "A survey of the Hough transform". **Computer
Vision, Graphics, and Image Processing**, Vol. 44, No.1, pp.87-116.
- A.K. Jain e P.J. Flynn, 1996. Image segmentation using clustering. **Advances in
Image Understanding**. IEEE Computer Society Press. pp.65-83.
- R. Jain, R. Kasturi, e B.G. Schunck, 1996. **Machine Vision**. McGraw-Hill.
- R. Jarvis, 1983A. "A perspective on range finding techniques for computer vision".
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.5,
No.2, pp.122-139.
- R.A. Jarvis, 1983B. "A laser time-of-flight range scanner for robot vision". **IEEE
Transactions on Pattern Analysis and Machine Intelligence**, Vol.5, No.5,
pp.505-512.

- W. Grosky e R. Jain, 1986. "A pyramid-based approach to segmentation applied to region matching". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.8, No.5, pp.639-650.
- J. Hackett e M. Shah, 1989. "Segmentation using intensity and range data". **Optical Engineering**, Vol.28, No.6, pp.667-674.
- J.F. Haddon, 1988. "Generalised threshold selection for edge detection". **Pattern Recognition**, Vol.21, No.3, pp.195-203.
- R.M. Haralick e L. G. Shapiro, 1992. **Computer and Robot Vision: Volume I**. Addison-Wesley Publishing Company.
- M. Herman, 1986. "Representation and incremental construction of a three-dimensional scene model". **Techniques dor 3-D Machine Perception**, Elsevier Science Publishers B.V., North Holland, pp. 149-183.
- K. Higuchi, M. Herbert, e K. Ikeuchi, 1995. "Building 3-D models from unregistered range images". **Graphical Models and Image Processing**, Vol.57, No.4, pp.315-333.
- C.M. Hoffman, 1989. **Geometric&Solid Modeling: An Introduction**. Morgan Kauffman Publishers.

- R. Krishnapuram e A. Munshi, 1991. "Cluster-based segmentation of range images using differential-geometric features". **Optical Engineering**, Vol.30, No.10, pp.1468-1478.
- D.T. Kuan e R.J. Drazovich, 1986. "Model based interpretation of 3-D range data".
Techniques for 3-D Machine Perception, Elsevier Publishers B.V., North Holland, pp.219-230.
- I. Kweon e T. Kanade, 1992. "High-resolution terrain map from multiple sensor data".
IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.14, No.2, pp.278-292.
- P. Lancaster e K. Salkauskas, 1986. **Curve and Surface Fitting: An Introduction**. Academic Press.
- R.C. Luo e M.G. Kay, 1989. "Multisensor integration and fusion in intelligent systems". **IEEE Transactions on Systems, Man, and Cybernetics**, Vol.19, No.5, pp.901-931.
- M. Mäntylä, 1988. **An Introduction to Solid Modeling**. Computer Science Press.
- T. Masuda e N. Yokoya, 1995. "A robust method for registration and segmentation of multiple range images". **Computer Vision and Image Understanding**, Vol.61, No.3, pp.295-307.

- A. Mitiche e J.K. Aggarwal, 1983. "Detection of edges using range information". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.5, No.2, pp.174-178.
- A. Mitiche e J.K. Aggarwal, 1986. "Multiple sensor integration/fusion through image processing: a review". **Optical Engineering**, Vol.25, No.3, pp.380-386.
- T.S. Newman, P.J. Flynn e A.K. Jain, 1993. "Model-based classification of quadric surfaces". **CVGIP: Image Understanding**. Vol.58, No.2, pp.235-249.
- D. Nitzan, A. Brain e R. Duda, 1977. "The measurement and use of registered reflectance and range data in scene analysis". **Proceedings of the IEEE**, Vol.65, No.2, pp.206-220.
- D. Nitzan, 1988. "Three-dimensional vision structure for robot applications". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.10, No.3, pp.291-309.
- B. Noble e J.W. Daniel, 1988. **Applied Linear Algebra**. Prentice-Hall Int.
- N. Pal e S. Pal, 1993. "A review on image segmentation techniques". **Pattern Recognition**, Vol.26, No.9, pp.1277-1294.
- B. Parvin e G. Medioni, 1989. "Adaptive multiscale feature extraction from range images". **Computer Vision, Graphics, and Image Processing**, Vol.45, pp.346-356.

- W.K. Pratt, 1991. **Digital Image Processing**. Wiley Interscience.
- F.P Preparata e M.I. Shamos, 1985. **Computational Geometry: An Introduction**. Springer-Verlag, NY-USA.
- B. Sabata, F. Arman e J.K. Aggarwal, 1993. "Segmentation of 3D range images using pyramidal data structures". **CVGIP: Image Understanding**, Vol.57, No.3, pp.373-387.
- P.K. Sahoo, S. Soltani e A.K.C. Wong, 1988. "A survey of thresholding techniques". **Computer Vision, Graphics, and Image Processing**, Vol.41, pp.233-260.
- N. Sapidis e P.J. Besl, 1995. "Direct construction of polynomial surfaces from dense range images through region growing". **ACM Transactions on Graphics**, Vol.14, No.2, pp.171-200.
- W.B. Seales e O. Faugeras, 1995. "Building three-dimensional object models from image sequences". **Computer Vision and Image Understanding**, Vol.61, No.3, pp.308-324.
- S.Z. Selim e M.A. Ismail, 1984. "K-Means-type algorithms: a generalized convergence theorem and characterization of local optimality", **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.6, pp.81-86.
- J. Shen e S. Castan, 1986. "An optimal linear operator for edge detection". **Proc. of CVPR'86**, Miami, pp.109-114.

- M. Soucy e D. Laurendeau, 1996. "Multiresolution surface modeling based on hierarchical triangulation". **Computer Vision and Image Understanding**, Vol.63, No.1, pp.1-14.
- P. Suetens, P. Fua, e A.J. Hanson, 1992. "Computational strategies for object recognition". **ACM Computing Surveys**, Vol.24, No.1, pp.1-61.
- S. Tamura, E. Kim, R. Close, e Y. Sato, 1994. "Error correction in laser scanner three-dimensional measurement by two-axis model and coarse-fine parameter search". **Pattern Recognition**, Vol.27, No.3, pp331-338.
- R. Taylor, M. Savini e A. Reeves, 1989. "Fast segmentation of range imagery into planar regions". **Computer Vision, Graphics, and Image Processing**, Vol.45, pp.42-60.
- M.A. Wani e B.G. Batchelor, 1994. "Edge region-based segmentation of range images". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.16, No.3, pp.314-319.
- N. Yokoya e M. Levine, 1989. "Range image segmentation based on differential geometry: a hybrid approach". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.11, No.5, pp.643-649.
- X. Yu, T.D. Bui e A. Krzyak, 1994. "Robust estimation for range image segmentation and reconstruction". **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.16, No.5, pp.530-538.

G. Zhang e A. Wallace, 1993. "Physical modeling and combination of range and intensity edge data". **CVGIP: Image Understanding**, Vol.58, No.2, pp.191-220.