

Universidade Federal de Juiz de Fora
Programa de Pós-Graduação em Modelagem Computacional

**Uma Metodologia para Reconstrução de Superfícies a
Partir de Luz Estruturada**

Por
Anderson Rodrigues Vilela

Juiz de Fora, MG - BRASIL
MARÇO DE 2008

Vilela, Anderson Rodrigues

Uma Metodologia para Reconstrução de Superfícies a Partir de Luz Estruturada / Anderson Rodrigues Vilela; Orientador: Prof. Dr. Flávio de Souza Barbosa , Co-orientador: Prof. Dr. Fernando Marques de Almeida Nogueira. - 2008.

98 fl.: il.

Dissertação (Mestrado em Modelagem Computacional) – Faculdade de Engenharia Civil, Universidade Federal de Juiz de Fora, Juiz de Fora, 2008.

1. Computação gráfica. 2. Processamento de imagens. 3. Reconstrução tridimensional. I. Barbosa, Flávio de Souza. II. Nogueira, Fernando Marques de Almeida. III. Título.

CDU 681.3.04:74

**UMA METODOLOGIA PARA RECONSTRUÇÃO DE SUPERFÍCIES
A PARTIR DE LUZ ESTRUTURADA**

Anderson Rodrigues Vilela

DISSERTAÇÃO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO EM
MODELAGEM COMPUTACIONAL DA UNIVERSIDADE FEDERAL DE JUIZ
DE FORA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (M.SC.) EM
MODELAGEM COMPUTACIONAL.

Aprovada por:

Prof. Flávio de Souza Barbosa, D.Sc.
(Orientador)

Prof. Fernando Marques de Almeida Nogueira, D.Sc.
(Co-orientador)

Prof. Marcelo Bernardes Vieira, D.Sc.

Prof. Ricardo Cordeiro de Farias, Ph.D.

Prof. Carlos Magluta, D.Sc.

JUIZ DE FORA, MG - BRASIL

MARÇO DE 2008

Uma imagem vale mais que mil palavras.

Sabedoria popular

*Dedico esta dissertação a meus pais,
que sempre investiram em minha formação
e ao meu filho, que tem contribuído
de forma fundamental para que
eu me esforce mais e possa apoiá-lo
como fui apoiado pelos meus pais*

Agradecimentos

Dedico meus sinceros agradecimentos:

- ao professor Flávio, pela orientação, incentivo e acima de tudo pelo companheirismo e amizade em todos esses longos 4 anos de Iniciação Científica e Mestrado;
- ao professor Fernando, por ter agregado o conhecimento em processamento de imagens ao trabalho e por ter cedido a “churrasqueira” para os ensaios experimentais.
- à FAPEMIG, Fundação de Amparo à Pesquisa de Minas Gerais, por ter investido no início de minha jornada no mundo da pesquisa com uma bolsa de iniciação científica;
- à CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo apoio financeiro durante o mestrado, sem o qual tudo seria mais complicado;
- à UFJF, Universidade Federal de Juiz de Fora, por oferecer formação superior de qualidade e gratuita;
- ao MMC, Mestrado em Modelagem Computacional, por ter me selecionado para a turma inaugural em seus primeiros passos;
- a todos amigos e familiares que contribuíram direta e indiretamente para que eu chegasse até aqui.

Resumo da Dissertação apresentada à UFJF como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UMA METODOLOGIA PARA RECONSTRUÇÃO DE SUPERFÍCIES A PARTIR DE LUZ ESTRUTURADA

Anderson Rodrigues Vilela

Março/2008

Orientador : Flávio de Souza Barbosa

Co-orientador : Fernando Marques de Almeida Nogueira

A reconstrução de superfícies a partir de imagens bidimensionais é um problema presente em diversas áreas de conhecimento, tais como: reconstrução de órgãos ou partes do corpo humano a partir de tomografias; reconhecimento automático de caracteres; visão computacional para montagem e inspeção; processamento de imagens aéreas e de satélite para previsão do tempo, dentre outras. Em todos esses casos faz-se necessário um estudo detalhado sobre processamento de imagens e sobre a escolha correta da técnica de reconstrução a ser adotada, a fim de que se obtenha bons resultados.

Apresenta-se neste trabalho um modelo de reconstrução de superfícies baseado numa técnica de reconstrução tridimensional através de luz estruturada. Utiliza-se uma fonte de luz retilínea incidindo sobre os objetos da cena como luz estruturada e aplica-se movimento controlado aos objetos fazendo com que suas superfícies sejam varridas por essa luz. Através das imagens captadas por uma câmera da projeção da luz estruturada sobre os objetos, faz-se a reconstrução de suas superfícies

O modelo é validado através de uma metodologia proposta utilizando imagens sintéticas obtidas via simulação computacional e imagens obtidas via ensaio experimental. Resultados da identificação de diversas superfícies tridimensionais são apresentados em exemplos de aplicação.

Abstract of Dissertation presented to UFJF as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A Methodology for Surface Reconstruction from Structured Light

Anderson Rodrigues Vilela

March/2008

Advisor: Flávio de Souza Barbosa

Co-advisor: Fernando Marques de Almeida Nogueira

The surfaces reconstruction from two-dimensional images is a problem present in various areas of knowledge, such as reconstruction of organs or parts of the human body from tomography; automatic recognition of characters; computer vision for assembly and inspection, processing of aerial and satellite images for weather forecast, among others. In all these cases it is necessary detailed studies on image processing and the right choice of the technique of reconstruction to be adopted in order to achieve good results.

It is presented in this work a model for reconstruction of surfaces based on a technique of three-dimensional reconstruction through structured light. The technique employs a source of straight light focused on the objects of the scene as structured light and applies controlled movement to the objects allowing their surfaces to be swept by that light. Through images captured by a camera containing the projection of the structured light on the objects, the surfaces are reconstructed.

The model is validated through a methodology proposed using synthetic images obtained by computer simulation and images obtained by experimental test. Results of the identification of various three-dimensional surfaces are presented as examples of application.

Sumário

1	Introdução	1
1.1	Objetivo do Trabalho	4
1.2	Escopo do Trabalho	5
2	Aspectos Gerais Sobre Processamento de Imagens	6
2.1	Representação de Imagens Digitais	6
2.2	Digitalização da Imagem	7
2.3	Passos Fundamentais do Processamento de Imagens	9
2.4	Aquisição de Imagens	14
2.5	Técnicas de Segmentação	15
2.5.1	Limiarização	15
2.5.2	Rotulação	20
3	Visão Computacional	26
3.1	Introdução	26
3.2	Formação da Imagem	27
3.2.1	Equações de Colinearidade Diretas	29
3.2.2	Equações de Colinearidade Inversas	33

3.3	Calibração da Câmera	36
3.4	Técnicas de Reconstrução	39
3.4.1	<i>Shape From Shading</i>	39
3.4.2	<i>Shape From Motion</i>	40
3.4.3	<i>Shape From Texture</i>	41
3.4.4	<i>Shape From Focus</i>	41
3.4.5	<i>Shape From Stereo</i>	42
3.4.6	Luz Estruturada	43
4	Um Modelo Para Reconstrução de Superfícies	45
5	Validação da Metodologia	53
5.1	Validação Através de Vídeos Gerados Artificialmente	53
5.1.1	Geração Artificial de Imagens	53
5.1.2	Calibração da Câmera	55
5.1.3	Processamento das Imagens	58
5.1.4	Apresentação dos Resultados	63
5.2	Validação Através de Ensaio Experimental	65
5.2.1	Aquisição Experimental de Imagens	65
5.2.2	Calibração da Câmera	68
5.2.3	Processamento das Imagens	68
5.2.4	Apresentação dos Resultados	69
6	Resultados	71

6.1	Cubo 2	71
6.2	Cilindro 1	72
6.3	Cilindro 2	73
6.4	Metade de um Cilindro Vazado 1	74
6.5	Metade de um Cilindro Vazado 2	75
6.6	Metade de um Toróide Reto	76
7	Conclusões	78

Lista de Figuras

2.1	Convenção de eixos.	7
2.2	Exemplos de discretização espacial de uma imagem.	10
2.3	Exemplos de discretização em amplitude de uma imagem.	11
2.4	Passos fundamentais do processamento de imagens [15].	12
2.5	Detalhes dos sensores para aquisição de imagens.	14
2.6	Exemplo de aplicação do limiar global.	17
2.7	Aplicação do <i>semithresholding</i> à imagem da figura 2.6(a).	18
2.8	Imagem com 3 modos dominantes.	19
2.9	Resultado da aplicação da limiarização multinível.	21
2.10	Vizinhança entre <i>pixels</i>	22
2.11	Exemplo do crescimento de regiões usando pontos iniciais conhecidos.	24
2.12	Exemplo de rotulação.	25
3.1	Referencial para o espaço-imagem e referencial para o espaço-objeto.	28
3.2	Formação da imagem	29
4.1	Esquema do modelo utilizado.	46
4.2	Vistas e imagem obtida pela câmera.	47

4.3	Determinação das coordenadas no referencial de tela e do espaço- imagem.	48
4.4	Determinação das coordenadas no referencial do espaço-objeto.	48
4.5	Detalhes da vista lateral.	49
5.1	Código do Pov-Ray® para geração da imagem da figura 5.2	54
5.2	Imagem gerada artificialmente	55
5.3	Seqüência de imagens obtida artificialmente	56
5.4	Imagem de controle com as coordenadas conhecidas	58
5.5	Seleção das regiões de interesse	59
5.6	Resultados do processo de limiarização (cores invertidas)	60
5.7	Resultados do processo de rotulação (cores invertidas)	60
5.8	Cálculo da linha média (cores invertidas)	61
5.9	Resultado do processamento do cubo	64
5.10	Detalhamento do sistema experimental para reconstrução tridimen- sional	66
5.11	Detalhe da fonte de luz - conexão entre o apontador <i>laser</i> e o resfriador de processador	66
5.12	Detalhe da esteira, cubo utilizado no ensaio experimental e triângulo de controle	67
5.13	Câmara escura utilizada para controlar a iluminação do ensaio	67
5.14	Seqüência de imagens obtida experimentalmente	68
5.15	Imagem de controle com as coordenadas conhecidas	69
5.16	Resultado experimental do processamento do cubo	70

6.1	Vista superior e imagem obtida pela câmera	72
6.2	Resultados do processamento do cubo 2	72
6.3	Vista superior e imagem obtida pela câmera	73
6.4	Resultados do processamento do cilindro 1	73
6.5	Vista superior e imagem obtida pela câmera	74
6.6	Resultados do processamento do cilindro 2	74
6.7	Vista superior e imagem obtida pela câmera	75
6.8	Resultados do cilindro vazado 1	75
6.9	Vista superior e imagem obtida pela câmera	76
6.10	Resultados do cilindro vazado 2	76
6.11	Vista superior e imagem obtida pela câmera	77
6.12	Resultados do toróide	77

Lista de Tabelas

2.1	Número de bits utilizados no armazenamento para alguns valores de N e k	8
5.1	Parâmetros obtidos pela calibração e impostos na geração das imagens	58
5.2	Parâmetros obtidos pela calibração e impostos na geração das imagens	65
5.3	Parâmetros obtidos pela calibração e impostos na geração das imagens	70

Capítulo 1

Introdução

Os métodos de processamento de imagens digitais se situam em duas áreas de aplicação principais: na melhoria de informações específicas para a interpretação humana e no processamento de dados de uma cena para interpretação automática feita por um computador. O presente trabalho foca a segunda área citada, que aborda problemas genericamente chamados de problemas de percepção de máquina, onde o interesse é estabelecer procedimentos para extrair informações de imagens através do processamento por computadores. Exemplos do tipo de informações que podem ser extraídas são dados estatísticos, como histogramas, e medições de distâncias multidimensionais.

Problemas típicos em percepção de máquina, que rotineiramente utilizam técnicas de processamento de imagens, são o reconhecimento automático de caracteres, visão computacional para montagem e inspeção, processamento de impressões digitais, processamento de imagens aéreas e de satélite para previsão do tempo, entre outros.

Um exemplo é a técnica proposta por Lee *et al* [1] para medição remota em tempo real de deslocamentos em pontes utilizando processamento de imagens. Uma câmera com lente telescópica acoplada obtém uma seqüência de imagens de um alvo instalado no ponto de medição. Um algoritmo para o reconhecimento do alvo é utilizado, calculando os deslocamentos através do deslocamento e geometria do alvo

em função do tempo. A metodologia é verificada através de testes de laboratório utilizando uma mesa vibratória e de um teste de campo em uma ponte. Os resultados são comparados com os obtidos por sensores de contato mostrando que a técnica pode ser aplicada com boa exatidão.

Como outro exemplo, uma técnica para determinação tridimensional de frequências naturais de estruturas a partir de imagens tomadas por apenas uma câmera é apresentada por Nogueira *et al* [2]. As coordenadas tridimensionais do centro de um alvo circular preto e fundo branco fixado sobre a estrutura são determinadas com precisão sub-*pixel* através de um processo de calibração realizado para cada imagem da seqüência de imagens (vídeo) tomada pela câmera, formando a resposta temporal do alvo analisado. Posteriormente, determinam-se as frequências naturais da estrutura por meio da Transformada Discreta de Fourier. Resultados para imagens sintéticas são apresentados, ratificando a eficiência da metodologia proposta.

O efeito do borramento na análise de frequências naturais de estruturas é abordado por Nogueira *et al* [3], mostrando que para um tempo de exposição constante, este fenômeno interfere nas análises, mas é minimizado através de uma estratégia proposta pelo trabalho.

Técnicas de processamento de imagens também são aplicadas à biologia, e o trabalho de Figueiredo *et al* [4] é um exemplo. Nele, é proposto um método para contagem automática de hemácias em imagens médicas.

Uma outra aplicação do processamento de imagens presente em diversas áreas do conhecimento é a reconstrução tridimensional a partir de imagens bidimensionais. Por exemplo, imagens obtidas a partir de tomografias podem ser utilizadas para reconstruir as formas tridimensionais de órgãos ou partes de órgãos, assim como imagens aerofotogramétricas são utilizadas em levantamentos topográficos. De forma semelhante, é feita a identificação automática de volumes edificados partindo-se de vistas de um projeto arquitetônico.

A reconstrução tridimensional de uma cena pode ser realizada através de várias técnicas, entre as quais pode-se citar: reconstrução baseada na informação ra-

diométrica contida em uma imagem - *Shape from Shading* [5]; reconstrução através de parâmetros geométricos e da variável tempo para a determinação da estrutura tridimensional da cena - *Shape from Motion* [6]; a partir da modelagem dos efeitos que a distância focal e a abertura do diafragma da câmera exercem sobre as imagens capturadas - *Shape from Focus* [7]; reconstrução através da projeção de um padrão de textura sobre uma cena de interesse - *Shape from Texture* [8]; e determinação de forma utilizando duas ou mais câmeras em posições distintas - *Shape from Stereo* [9].

Cada uma das técnicas citadas possui vantagens e desvantagens, no entanto, a denominada *Shape from Stereo*, é a mais difundida. A maior dificuldade desta técnica consiste em estabelecer a correspondência de *pixels* homólogos entre as imagens automaticamente. Para cenas dinâmicas, uma outra dificuldade em empregar esta técnica é sincronizar os instantes das tomadas das câmeras. Esta dificuldade pode inviabilizar a sua utilização, uma vez que pequenos intervalos de tempo de defasagem entre os instantes das tomadas das câmeras influenciam significativamente nos resultados obtidos.

Um trabalho que merece destaque no que se refere à reconstrução tridimensional é apresentada por Dalmia *et al* [10] onde uma abordagem computacional para extrair dados tridimensionais de resolução controlável é apresentada. É utilizada uma câmera em movimento e o processamento é feito em tempo real. Vários experimentos são apresentados para demonstrar a exatidão e a sua funcionalidade em uma variedade de cenas.

Uma forma de se evitar o uso de 2 câmeras da técnica *Shape from Stereo* é a substituição de uma delas por uma ou mais fontes de luz com estrutura conhecida. Desta forma não existe o problema de correspondência entre imagens, já que apenas uma imagem da cena é utilizada, uma vez que os padrões apresentados pela estrutura da luz incidindo sobre a cena, juntamente com os dados bidimensionais da imagem, permitem obter as coordenadas tridimensionais do objeto em estudo.

Um método para a medição de superfícies tridimensionais através de câmeras e luz estruturada é proposta por Izquierdo *et al* [11]. O método inclui algoritmos para

calibração sub-*pixel*, detecção de uma faixa de *laser* e reconstrução tridimensional. Diferentes estratégias de calibração e algoritmos para reconstrução são testados de acordo com sua exatidão e velocidade.

Wang *et al* [12] apresenta um sistema baseado em luz estruturada para modelagem de objetos. O sistema consiste em um projetor digital, uma câmera de vídeo e uma mesa giratória. A calibração do sistema é implementada de forma automática através de uma imagem de uma amostra de características conhecidas sob a luz. A luz estruturada é composta de faixas de três cores emitidas por um projetor. A validade e a exatidão do método são demonstrados por vários experimentos.

O presente trabalho utiliza uma técnica de luz estruturada onde se projeta um feixe de *laser* que percorre a superfície em estudo. Uma câmera é utilizada para armazenar as imagens e um algoritmo de processamento de imagens reconstrói as coordenadas da superfície dos objetos da cena.

O algoritmo para o processamento das imagens faz inicialmente a seleção de regiões de interesse para diminuir o tempo de processamento, fazendo a segmentação e rotulação destas regiões. As coordenadas de pontos de interesse são extraídas e é feita a reconstrução tridimensional, que utiliza dados de uma calibração previamente realizada.

A validação da metodologia é feita primeiramente através de imagens geradas artificialmente, eliminando qualquer forma de ruído. Em seguida, um ensaio experimental é proposto de forma a analisar o método em condições reais. Resultados obtidos pelo método são também analisados atestando a viabilidade da metodologia proposta.

1.1 Objetivo do Trabalho

Propor uma metodologia para reconstrução de superfícies a partir de luz estruturada composta por um feixe retilíneo de luz incidindo sobre os objetos da cena.

1.2 Escopo do Trabalho

Este trabalho está dividido em 7 capítulos e, visando dar ao leitor uma melhor compreensão do seu conteúdo, faz-se uma descrição resumida do teor de cada capítulo e seus principais objetivos.

O capítulo 1 situa este trabalho entre as áreas de atuação do processamento de imagens. Uma revisão de trabalhos relacionados que inspiraram a proposição da metodologia também é apresentada, juntamente com um breve resumo do conteúdo dos capítulos

São apresentados no capítulo 2 aspectos básicos sobre imagens digitais, mostrando como é feita a representação dessas imagens, sua digitalização e a convenção de eixos adotada no presente trabalho. Alguns aspectos sobre processamento de imagens utilizados no desenvolvimento da metodologia, como os passos fundamentais para o processamento de imagens e técnicas de segmentação também são mostrados.

Em seguida, no capítulo 3, é feita uma revisão sobre aspectos geométricos da visão computacional, onde será mostrado como ocorre a formação da imagem bem como serão desenvolvidas as equações que fazem a correspondência entre um ponto da imagem e um ponto no espaço. Ao final do capítulo, faz-se uma revisão das técnicas de reconstrução tridimensional mais aplicadas.

Um modelo para a reconstrução de superfícies é proposto no capítulo 4 utilizando um feixe de luz retilíneo como luz estruturada.

Com o sistema desenvolvido, no capítulo 5 as etapas do processo de reconstrução utilizando a metodologia proposta são detalhadas. A metodologia é verificada em aplicações de reconstrução tridimensional de objetos utilizando imagens obtidas artificialmente e experimentalmente.

Outros resultados são mostrados no capítulo 6, confirmando a confiabilidade do sistema e fazendo algumas considerações quanto a problemas encontrados.

Finalmente, as últimas considerações e propostas de trabalhos futuros são feitas no capítulo 7.

Capítulo 2

Aspectos Gerais Sobre Processamento de Imagens

Neste capítulo serão apresentados a representação de imagens adotada neste trabalho, aspectos básicos do processo de digitalização de imagens e fundamentos de processamento de imagens.

2.1 Representação de Imagens Digitais

O termo imagem corresponde a uma função bidimensional de intensidade de luz $f(i, j)$, onde o valor f em um ponto de coordenadas (i, j) é proporcional à intensidade de luz, ou brilho, da imagem naquele ponto. Quando se considera que a imagem é formada por um processo físico, os valores de $f(i, j)$ são proporcionais à energia irradiada pela fonte e pelo albedo da superfície iluminada. Então os valores de intensidade devem ser não nulos e finitos: $0 < f(i, j) < \infty$. Regiões onde o brilho da imagem é maior, o valor da função $f(i, j)$ também o é.

Para que seja possível a obtenção de uma imagem digitalizada de forma única, faz-se necessário definir um sistema de coordenadas. A figura 2.1 mostra a convenção de eixos usada neste trabalho, chamada de sistema referencial de tela.



Figura 2.1: Convenção de eixos.

Em imagens monocromáticas, a intensidade f no ponto (i, j) é o nível de cinza l da imagem naquele ponto e pertence ao intervalo $[L_{MIN}, L_{MAX}]$, onde L_{MIN} é o valor mínimo e L_{MAX} é o valor máximo que l pode assumir. Normalmente faz-se uma translação de forma que $L_{MIN} = 0$ e o intervalo fique representado no intervalo $[0, L]$ onde $l = 0$ é o valor mínimo e corresponde ao preto e $l = L$ é o valor máximo e corresponde ao branco. Todos os valores intermediários são níveis de cinza variando do preto até o branco.

2.2 Digitalização da Imagem

Para uma imagem ser tratada por um computador, a função de imagem $f(i, j)$ deve ser discretizada espacialmente e em amplitude. Então esta função contínua $f(i, j)$ é aproximada por elementos igualmente espaçados arranjados na forma de uma matriz $N \times M$, onde $i, j \in \mathbb{N}^*$, com $1 \leq i \leq N$ e $1 \leq j \leq M$ como mostra a equação 2.1, e cada elemento da matriz é uma quantidade discreta.

$$f(i, j) = \begin{bmatrix} f(1, 1) & f(1, 2) & \dots & f(1, M) \\ f(2, 1) & f(2, 2) & \dots & f(2, M) \\ \dots & \dots & \dots & \dots \\ f(N, 1) & f(N, 2) & \dots & f(N, M) \end{bmatrix}. \quad (2.1)$$

Uma imagem digital monocromática pode ser considerada uma matriz cujos índices de linha e coluna identificam um ponto na imagem e o correspondente valor do elemento na matriz identifica o valor da escala de cinza naquele ponto. Estes elementos são chamados em inglês de *picture elements*, e são usualmente chamados pela abreviação *pixels*.

O tamanho de uma imagem digital varia de acordo com a aplicação e deve ser escolhido para se obter uma boa discretização. Normalmente escolhem-se valores para as dimensões da imagem e do número de níveis de cinza que são potências de 2. Desta forma pode-se referir à discretização pelo número de bits usados para o armazenamento.

A discretização em amplitude de $f(i, j)$ é feita dividindo-se o intervalo contínuo $[0, L]$, apresentado na seção 2.1, em 2^k partes. O valor de k indica o número de *bits* usados para representar a intensidade.

A tabela 2.1 mostra exemplos de número de bits usados para armazenar a imagem para diversos valores de N e k , considerando $N = M$.

Tabela 2.1: Número de bits utilizados no armazenamento para alguns valores de N e k .

k \ N	1	2	3	4	5	6	7	8
32	1.024	2.048	3.072	4.096	5.120	6.144	7.168	8.192
64	4.096	8.192	12.288	16.384	20.480	24.576	28.672	32.768
128	16.384	32.768	49.152	65.536	81.920	98.304	114.688	131.072
256	65.536	131.072	196.608	262.144	327.680	393.216	458.752	524.288

A figura 2.2 mostra exemplos de discretização espacial para uma imagem digitalizada em 512×512 , 256×256 , 128×128 , 64×64 , 32×32 e 16×16 *pixels*. Em seguida, exemplos de discretização da amplitude da mesma imagem são mostrados na figura 2.3 utilizando 256, 64, 16, 8, 4 e 2 níveis de cinza, correspondendo respectivamente a 8, 6, 4, 3, 2 e 1 *bits* para representação dos níveis de cinza.

Uma imagem digital é dita de boa qualidade dependendo da finalidade para a qual ela é usada. A figura 2.3(f), por exemplo, é suficiente para que se identifique que há uma flor na imagem e com boa aproximação determinar a forma das pétalas, mas não é possível identificar folhas atrás da flor. Como outro exemplo, a figura 2.2(d) pode ser utilizada para identificar tanto a flor e as folhas em seu fundo, mas a determinação da forma das pétalas é muito grosseira. Então, o conceito de imagem de boa qualidade é subjetivo, podendo a imagem ser considerada de boa qualidade para algumas aplicações e de má qualidade para outras. Deve-se escolher uma discretização razoável para que se ocupe o mínimo de espaço de armazenamento com um nível de qualidade aceitável.

A formação de imagens digitais coloridas segue os mesmos princípios fundamentais apresentados para imagens em escala de cinza. Entretanto, necessita-se de mais de uma matriz para se ter as informações de intensidade de cada *pixel*. Assim sendo, por exemplo, no sistema de representação RGB (Red - vermelho, Green - verde e Blue - azul) tem-se 3 matrizes de intensidade, uma para cada cor. Como as imagens utilizadas neste trabalho são monocromáticas, convida-se o leitor para a leitura da referência [13] para se obter maiores informações sobre imagens coloridas.

2.3 Passos Fundamentais do Processamento de Imagens

Segundo Gonzalez *et al* [14] os métodos que compõem o processamento de imagens são divididos em duas grandes categorias: métodos nos quais as entradas e saídas são imagens e métodos em que as entradas são imagens e saídas são atri-



(a) 512x512



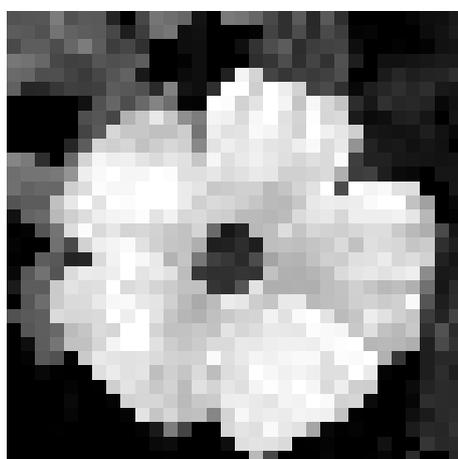
(b) 256x256



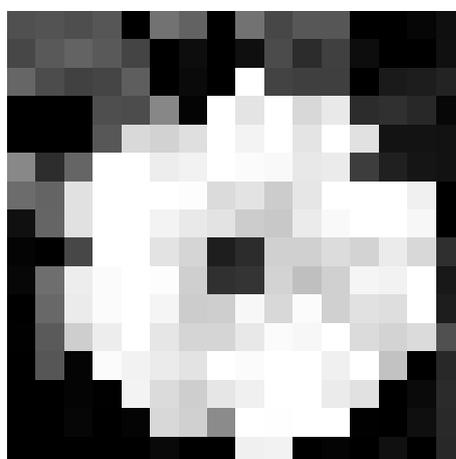
(c) 128x128



(d) 64x64



(e) 32x32



(f) 16x16

Figura 2.2: Exemplos de discretização espacial de uma imagem.



(a) 256 cores (8 bits)



(b) 64 cores (6 bits)



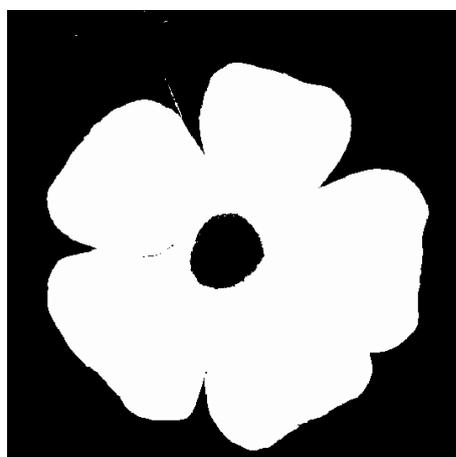
(c) 16 cores (4 bits)



(d) 8 cores (3 bits)



(e) 4 cores (2 bits)



(f) 2 cores (1 bits)

Figura 2.3: Exemplos de discretização em amplitude de uma imagem.

butos extraídos destas imagens. Estas categorias são divididas em outras menores mostradas no esquema da figura 2.4. Dependendo dos objetivos do processamento, não são aplicadas todas estas etapas.

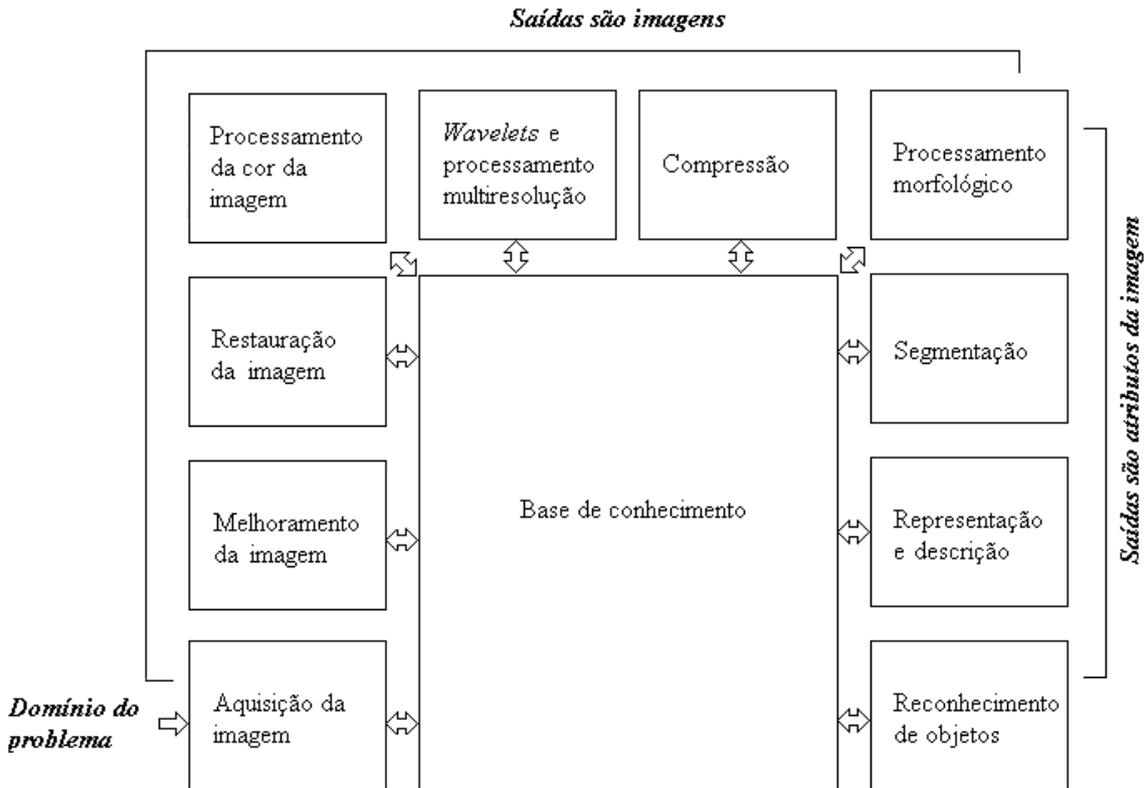


Figura 2.4: Passos fundamentais do processamento de imagens [15].

A seguir, cada etapa mostrada na figura 2.4 são comentadas.

Aquisição das imagens: é o primeiro processo mostrado na figura 2.4. Geralmente envolve a digitalização, armazenamento, pré-processamento e transformações de escala, entre outros. Esta etapa será comentada com mais detalhes na seção 2.4.

Melhoramento: Nesta etapa, detalhes da imagem de interesse ou que podem estar imperceptíveis são realçados. Exemplos comuns são alterações do contraste da imagem, filtros e alterações de cores. Usa preferências humanas subjetivas.

Restauração: é a etapa na qual se melhora a aparência da imagem, é uma etapa que não é subjetiva como o melhoramento de imagens, se baseia quase sempre em modelos matemáticos e probabilísticos de degradação de imagens.

Processamento de cores: O uso de cores em processamento automático de imagens é motivado por dois principais fatores: na análise automática de imagens - a cor é uma característica que pode simplificar a extração e identificação de objetos de uma cena; e na análise de imagens feitas por meios humanos - a motivação pelas cores advém do fato que o olho humano é capaz de identificar milhares de intensidade de cor, comparado com centenas de níveis de cinza.

Wavelets: é o fundamento para se representar uma imagem em vários níveis de resolução. *Wavelets* podem ser empregados de várias maneiras e objetivos diferentes, como compressão e representação piramidal.

Compressão: Nesta etapa são aplicadas técnicas para diminuir o “espaço” necessário para armazenamento de imagens, e por conseqüência, a largura de banda necessária para sua transmissão.

Processamento morfológico: São aplicadas ferramentas para extrair componentes da imagem que são importantes na representação e descrição de formas presentes.

Segmentação: Separa a imagem em suas partes ou objetos constituintes. A segmentação automática é uma das etapas mais desafiadores do processamento de imagens. O procedimento de segmentação para a solução de problemas de identificação individual de objetos é muito complexo. Se a segmentação não for feita corretamente, têm-se falhas que prejudicam todo o processamento. Em geral, quanto mais precisa é a segmentação, melhor é o processo de reconhecimento.

Representação e descrição: Normalmente segue a etapa de segmentação, processando a sua saída para um formato que pode ser melhor entendido pelo computador. Um exemplo é decidir se os dados representam um contorno ou uma região. Após a escolha de uma representação, um método para a descrição dos dados deve ser aplicado. Uma descrição seria uma seleção de atributos presentes nos dados que são importantes, representando informações quantitativas ou informações básicas para diferenciar uma classe de objetos de outra.

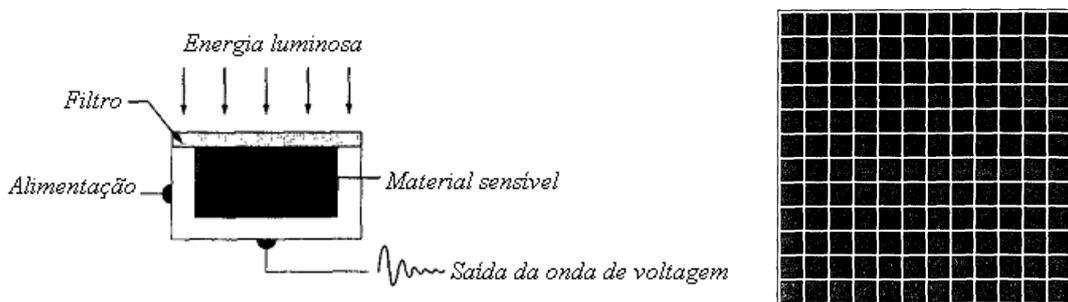
Reconhecimento: é a etapa que atribui um “nome” ao objeto, informando se

ele é um veículo, uma figura geométrica ou uma letra, por exemplo. Os dados da etapa de descrição são usados por métodos para o reconhecimento automático de objetos.

Para tornar o processamento de imagens mais rápido e eficiente, uma base de conhecimentos deve ser usada em cada etapa. Esses conhecimentos podem ser prévios para cada problema, por exemplo, regiões da imagem onde se é esperado encontrar os objetos de interesse. Neste caso, elimina-se o restante da imagem do processamento, trabalhando somente com estas regiões. Esse conhecimento pode ser mais complexo, como uma base de dados de objetos que podem ser encontrados na imagem. Após o processamento, técnicas de correlação são aplicadas para se escolher dentre estes objetos da base, qual é o que melhor se aplica ao caso.

2.4 Aquisição de Imagens

A aquisição de imagens é feita por sensores como mostrado na figura 2.5(a). O plano sensor de uma câmera é formado por uma matriz destes sensores (figura 2.5(b)). A resposta de cada sensor é proporcional à energia luminosa projetada sobre a superfície do sensor sendo que cada sensor é responsável por gerar um *pixel*.



(a) Detalhe de um sensor

(b) Matriz de sensores

Figura 2.5: Detalhes dos sensores para aquisição de imagens.

2.5 Técnicas de Segmentação

O primeiro passo da análise de imagens normalmente é a sua segmentação, que divide a imagem em partes ou objetos. Segmentar é separar partes de interesse em uma imagem. Separar objetos do fundo de uma imagem pode ser um exemplo, assim como detectar linhas, pontos, formas geométricas e contornos presentes na imagem. O nível e a técnica de segmentação variam dependendo do problema a ser resolvido. Uma simples imagem binária pode ser suficiente em alguns casos, já em outros, regiões rotuladas, linhas e contornos devem ser obtidos.

Nos casos de detecção de descontinuidades, cantos e contornos (bordas) a segmentação é feita baseada na descontinuidade de intensidade. Limiares baseados na distribuição das propriedades dos *pixels*, como a intensidade ou cor são definidos agregando *pixels* de propriedade semelhantes e definindo assim regiões na imagem. Na segmentação orientada a regiões, noções de vizinhança e conectividade de *pixels* são aplicadas.

A seguir são dadas algumas noções sobre as técnicas mais usadas de segmentação.

2.5.1 Limiarização

Nesta seção serão descritas algumas técnicas de limiarização (tradução do termo em inglês *thresholding*). A limiarização é uma das mais simples e a mais usada técnica de segmentação. O seu objetivo é dividir a imagem em regiões de interesse e remover todas as outras regiões que não são importantes.

O método mais simples de limiarização usa um único parâmetro limite para classificação dos *pixels* (limiar). Em muitos casos, com este método, nenhum limiar oferece um bom resultado de segmentação sobre toda a imagem. Nestes casos, técnicas usando limiar variável, em multiníveis e baseadas em várias medidas estatísticas são usadas.

Quando o limiar depende somente do valor do *pixel*, o limiar é chamado global. Se depender tanto do valor do *pixel* quanto de alguma outra propriedade, é chamado

local. Se também depender das coordenadas espaciais i e j , é chamado dinâmico. Quando a limiarização resultar em uma imagem binária, ou seja, seus *pixels* são pretos (0) ou brancos (1), o processo pode ser denominado também de *binarização*.

A seguir são mostradas diferentes estratégias para a seleção do limiar adequado.

a) Limiar Global

A técnica de limiarização global é usada para isolar objetos de interesse tendo valores diferentes do fundo da imagem. Cada *pixel* é classificado como sendo parte de um objeto de interesse ou como fundo.

Supondo-se que a imagem em escala de cinza mostrada na figura 2.6(a) possua o histograma da figura 2.6(b) de distribuição de ocorrências de níveis de cinza. Nesta imagem $f(i, j)$ compõem objeto claros em um fundo escuro, de forma que os *pixels* estejam agrupados em dois modos dominantes. Uma maneira óbvia de extrair os objetos do fundo é selecionar um limiar T que separa estes modos. Então, para qualquer ponto (i, j) em que $f(i, j) > T$ é chamado de um ponto do objeto, senão, é chamado de ponto do fundo.

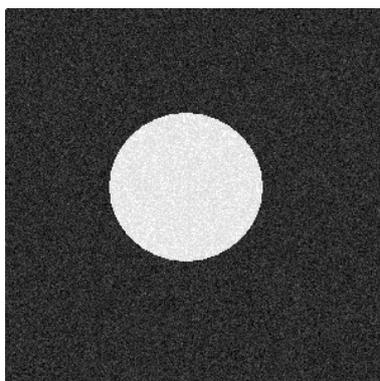
A imagem *limiarizada* $g(i, j)$ é definida como

$$g(i, j) = \begin{cases} 1, & \text{se } f(i, j) > T, \\ 0, & \text{se } f(i, j) \leq T, \end{cases} \quad (2.2)$$

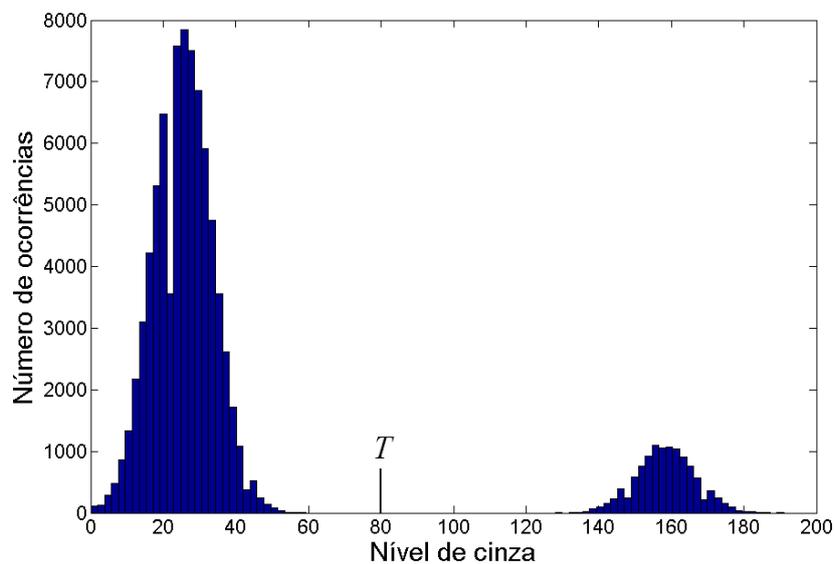
onde os *pixels* rotulados com o valor 1 ou qualquer outro valor diferente de 0 correspondem aos objetos e os rotulados com o valor 0 correspondem ao fundo. A figura 2.6(c) mostra o resultado desta técnica aplicada à imagem da figura 2.6(a).

b) *Semithresholding*

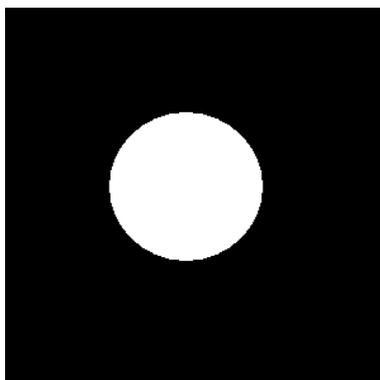
Semithresholding é uma variação da limiarização global. *Pixels* com valores dentro de um dado intervalo mantêm seus valores originais. Os outros *pixels* assumirão valor 0.



(a) Imagem com um objeto claro em um fundo escuro



(b) Histograma de distribuição de ocorrências de níveis de cinza da imagem da figura 2.6(a)



(c) Imagem da figura 2.6(a) após a aplicação do limiar global

As regiões de interesse são isoladas usando:

$$g(i, j) = \begin{cases} f(i, j), & \text{se } f(i, j) > T, \\ 0, & \text{se } f(i, j) \leq T. \end{cases} \quad (2.3)$$

A figura 2.7 mostra a limiarização da imagem da figura 2.6(a) utilizando esta técnica.

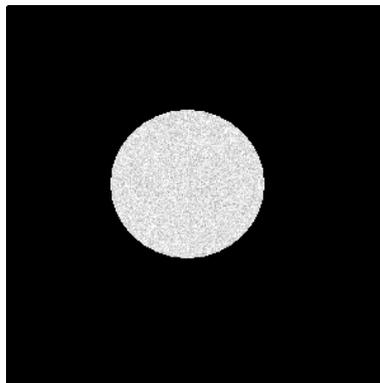


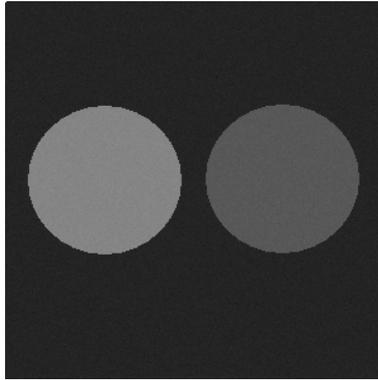
Figura 2.7: Aplicação do *semithresholding* à imagem da figura 2.6(a).

c) Limiarização Multinível

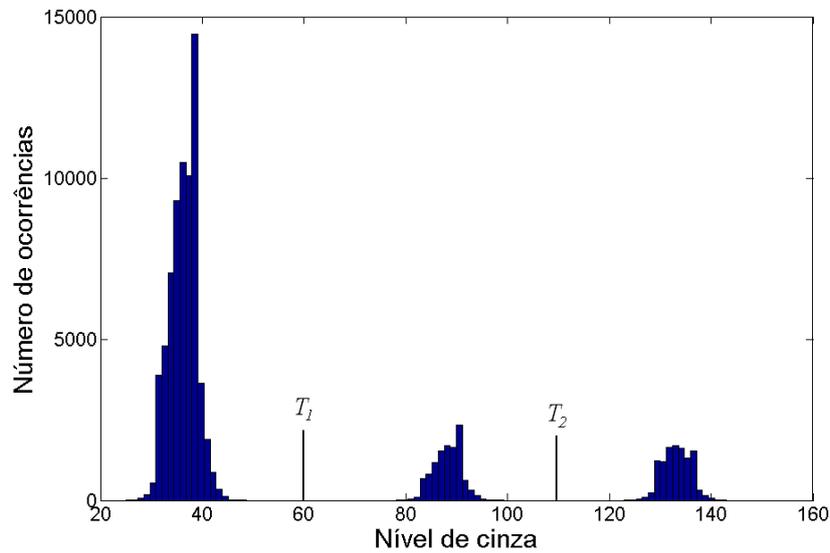
As técnicas limiarização global e *semithresholding* segmentam a imagem assumindo que ela tem somente dois tipos de regiões, uma mais clara e outra mais escura. Certamente a imagem pode conter mais do que dois tipos de regiões, e a limiarização multinível é a técnica a ser aplicada nestes casos.

Esta técnica é uma extensão das duas técnicas anteriores que permite a segmentação dos *pixels* em múltiplas classes.

A figura 2.8(a) mostra um exemplo. Três modos dominantes caracterizam o histograma de distribuição de ocorrências de níveis de cinza da imagem (figura 2.8(b)), que poderiam ser de dois objetos claros de cores diferentes sobre um fundo escuro. Esta abordagem classifica um ponto (i, j) pertencendo a uma classe de objetos se $T_1 < f(i, j) \leq T_2$, ou a outra classe de objetos se $f(i, j) > T_2$. Como o fundo tem-se $f(i, j) \leq T_1$.



(a) Imagem com dois objetos em um fundo escuro



(b) Histograma de distribuição de níveis de cinza da imagem da figura 2.8(a)

Figura 2.8: Imagem com 3 modos dominantes.

Para uma imagem $f(i, j)$ e um intervalo limiar $[T_1, T_2]$ dados, as imagens *limiarizadas* $g_1(i, j)$ e $g_2(i, j)$ são dadas por:

$$g_1(i, j) = \begin{cases} f(i, j), & \text{se } T_1 \leq f(i, j) \leq T_2 & \text{objeto mais escuro,} \\ 0, & \text{se } f(i, j) < T_1 & \text{fundo.} \end{cases} \quad (2.4)$$

$$g_2(i, j) = \begin{cases} f(i, j), & \text{se } f(i, j) \geq T_2 \quad \text{objeto mais claro,} \\ 0, & \text{se } f(i, j) < T_2 \quad \text{fundo.} \end{cases} \quad (2.5)$$

Para obter as imagens $g_1(i, j)$ e $g_2(i, j)$ binárias:

$$g_1(i, j) = \begin{cases} 1, & \text{se } T_1 \leq f(i, j) \leq T_2 \quad \text{objeto mais escuro,} \\ 0, & \text{se } f(i, j) < T_1 \quad \text{fundo.} \end{cases} \quad (2.6)$$

$$g_2(i, j) = \begin{cases} 1, & \text{se } f(i, j) \geq T_2 \quad \text{objeto mais claro,} \\ 0, & \text{se } f(i, j) < T_2 \quad \text{fundo.} \end{cases} \quad (2.7)$$

A figura 2.9 mostra o resultado da aplicação desta técnica à imagem da figura 2.8(a).

Este tipo de limiarização é geralmente menos confiável do que o limiar simples, porque é difícil escolher limiares que efetivamente isolam regiões de interesse, especialmente quando o número de modos no histograma é grande. Nestes casos o estudo de técnicas mais complexas de limiarização deve ser feito. As referências [14] e [13] contêm informações relevantes sobre outras técnicas.

2.5.2 Rotulação

Primeiramente, seguem dois conceitos preliminares: vizinhança e conectividade entre *pixels*:

- **Vizinhança:** É o conjunto de *pixels* localizados relativamente próximos a um *pixel* de referência. Normalmente se trabalha com dois tipos de vizinhança, mostrados na figura 2.10: a vizinhança de 4 *pixels* e a vizinhança de 8 *pixels*.
- **Conectividade:** é um conceito importante no estabelecimento de regiões e contornos. Para determinar se os *pixels* estão conectados se seus vizinhos

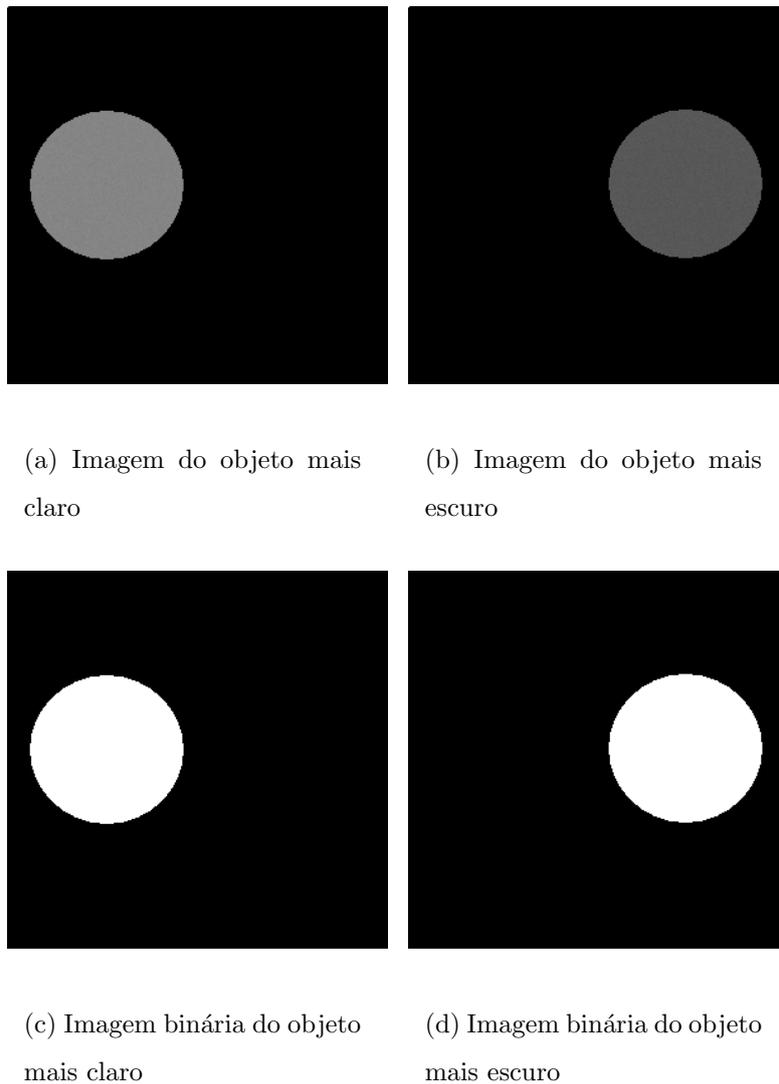


Figura 2.9: Resultado da aplicação da limiarização multinível.

obedecem a algum critério quanto ao seu nível de cinza. Em imagens binárias, verifica-se se os vizinhos têm o valor 1 ou 0.

Seja F a representação de toda a imagem. Pode-se dizer que a segmentação é um processo que particiona F em n sub-regiões, F_1, F_2, \dots, F_n da seguinte maneira:

1. $\cup_{i=1}^n F_i = F$;
2. F_i é uma região conectada, $i = 1, 2, \dots, n$;
3. $F_i \cap F_j = \emptyset$ para todo i e j , $i \neq j$;

	1	2	3	4	5
1	6	8	8	7	9
2	7	6	7	9	7
3	1	0	1	1	0
4	2	0	1	0	0
5	1	0	2	1	1

(a) Vizinhança de 4 *pixels*

	1	2	3	4	5
1	6	8	8	7	9
2	7	6	7	9	7
3	1	0	1	1	0
4	2	0	1	0	0
5	1	0	2	1	1

(b) Vizinhança de 8 *pixels*Figura 2.10: Vizinhança entre *pixels*.

4. $P(F_i) = VERDADEIRO$ par $i = 1, 2, \dots, n$;

5. $P(F_i \cup F_j) = FALSO$ para $i \neq j$

onde $P(F_i)$ é um predicado lógico os pontos do conjunto F_i e \emptyset é o conjunto vazio, o operador \cup é o operador de união e \cap o operador de interseção.

A primeira condição indica que segmentação deve ser completa, isto é, todo *pixel* deve estar em uma região. A condição (2) exige que os pontos em uma região devem estar conectados. A terceira condição indica que as regiões devem ser disjuntas. A condição (4) lida com as propriedades que devem ser satisfeitas pelos *pixels* em uma região segmentada, por exemplo $P(F_i) = VERDADEIRO$ se todos os *pixels* em F_i tem a mesma intensidade. Finalmente a condição (5) indica que a regiões F_i e F_j são diferentes no mesmo sentido do predicado P .

Como o próprio nome diz a rotulação, é um procedimento que agrupa *pixels* ou sub-regiões em regiões maiores. A abordagem mais simples é a agregação de *pixels*, que começa com um conjunto de pontos “semente” e a partir deles regiões são formadas acrescentando um destes pontos seus *pixels* vizinhos que possuem propriedades similares, como o nível de cinza, a textura ou a cor.

Para ilustrar este procedimento, aplica-se uma segmentação sobre a figura 2.11(a), onde os valores das células representam os níveis de cinza. Consideram-se dois pontos de coordenadas (3, 2) e (3, 4) como “sementes”. Usar dois pontos iniciais resulta em uma segmentação consistindo em, no máximo, duas regiões: F_1 associada à semente (3, 2) e F_2 associada à semente (3, 4). A propriedade P a ser usada é incluir um *pixel* se a diferença de intensidade for maior que 4.

A figura 2.12 mostra um exemplo que utiliza uma imagem binária para o processo de rotulação. Neste exemplo, *pixels* com valor 1 são agregados em regiões, e o valor de suas intensidades passa a conter o número da região. Os algoritmos mais comuns de rotulação, por padrão, percorrem as colunas de *pixels* da imagem procurando regiões. A primeira região identificada recebe o valor 1, a segunda 2, e assim em diante.

A abordagem utilizando pontos “semente” também pode ser adotada na rotulação do exemplo da figura 2.12 de forma semelhante à adotada no exemplo da figura 2.11

	1	2	3	4	5
1	6	8	8	7	9
2	7	6	<u>7</u>	9	7
3	1	0	1	1	0
4	2	0	<u>1</u>	0	0
5	1	0	2	1	1

(a) Matriz original

	1	2	3	4	5
1	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
2	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
3	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
4	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
5	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>

(b) Resultado da segmentação usando uma diferença absoluta de menos de 4 entre os níveis de intensidade

	1	2	3	4	5
1	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
2	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
3	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
4	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
5	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>

(c) Resultado da segmentação usando uma diferença absoluta de menos de 9

Figura 2.11: Exemplo do crescimento de regiões usando pontos iniciais conhecidos.

	1	2	3	4	5
1	1	1	0	0	0
2	1	0	1	1	1
3	0	0	1	1	1
4	1	0	1	1	1
5	1	0	0	0	0

(a) Matriz da imagem original

	1	2	3	4	5
1	1	1	0	0	0
2	1	0	3	3	3
3	0	0	3	3	3
4	2	0	3	3	3
5	2	0	0	0	0

(b) Resultado da rotulação com vizinhança de 4 *pixels*

	1	2	3	4	5
1	1	1	0	0	0
2	1	0	1	1	1
3	0	0	1	1	1
4	2	0	1	1	1
5	2	0	0	0	0

(c) Resultado da rotulação com vizinhança de 8 *pixels*

Figura 2.12: Exemplo de rotulação.

Capítulo 3

Visão Computacional

3.1 Introdução

A visão é o sentido que um grande número de espécies animais dispõe e que os capacita a adquirir, processar, extrair e representar informações do mundo que os cerca a partir de imagens bidimensionais.

O desenvolvimento computacional das câmeras de vídeo analógicas, e mais recentemente das câmeras de vídeo digitais, tornaram possível desenvolver sistemas de visão artificial que realizam certas tarefas similares as dos sistemas de visão biológicos. Tais sistemas estão ainda muito distantes de alcançarem a robustez e capacidade de generalização dos sistemas biológicos, haja visto a complexidade desse sentido, mas em algumas aplicações já é possível realizar as tarefas de aquisição, processamento, extração e representação da informação automaticamente com relativo sucesso [1, 10, 16].

Existem várias denominações similares para o termo visão computacional. Entre as mais comuns e utilizadas estão os termos: visão artificial, visão robótica, visão de máquina e visão cibernética.

Em certas aplicações não é necessária a informação tridimensional dos objetos que estão sendo analisados, como por exemplo, a inspeção de placas de circuitos

impresso, contagem de células e várias outras aplicações que envolvem processos de análise qualitativo e quantitativo [4, 16]. Porém quando a informação tridimensional da cena em questão é necessária para a obtenção de outras informações de mais alto nível, a complexidade dos sistemas aumenta significativamente. Como exemplos de aplicação de sistemas de visão artificiais que necessitam da informação tridimensional da cena pode-se citar a navegação autônoma, geração de mapas topográficos, inspeção tridimensional de produtos industrializados e robótica em geral.

A reconstrução tridimensional de uma cena é uma tarefa extremamente rápida, precisa e eficiente em sistemas de visão biológicos e infelizmente ainda obscura do ponto de vista do conhecimento de seu funcionamento. Várias teorias buscam explicar esse funcionamento, mas poucas soluções já foram concretizadas e consagradas pela comunidade científica.

A seguir será desenvolvida a formulação que governa as transformações entre o espaço bidimensional e o tridimensional e algumas técnicas de reconstrução que se encontram na literatura.

3.2 Formação da Imagem

Inicialmente, define-se como cena o conjunto de objetos, pessoas, paisagens e tudo o que compõe o cenário observado através de uma câmera. A luz refletida pelos objetos presentes em uma cena é capturada pelo sistema sensor da câmera proporcionalmente a sua intensidade. A posição que cada raio luminoso irá ocupar neste plano é modelado matematicamente por equações chamadas de equações de colinearidade.

As equações de colinearidade diretas modelam um mapeamento de pontos contidos no espaço tridimensional da cena, chamado de espaço-objeto, para um espaço chamado de espaço-imagem. Já as equações de colinearidade inversas modelam um mapeamento de pontos contidos no espaço-imagem para o espaço-objeto.

Antes de formalizar estas equações, faz-se necessário que os sistemas referenciais

sejam arbitrados. Na figura 3.1, X , Y e Z representam os eixos cartesianos no espaço-objeto; x , y e z representam os eixos cartesianos no espaço-imagem; i e j , representam os eixos cartesianos no sistema bidimensional da imagem, chamado de referencial de tela; as letras do alfabeto grego representam os ângulos de rotação entre os eixos dos dois primeiros sistemas citados: κ representa a rotação em torno do eixo Z , ϕ representa a rotação em torno do eixo Y e θ representa a rotação em X ; CP representa o centro de perspectiva da câmera e f é a distância entre o plano de formação da imagem e o CP . A reta perpendicular ao plano sensor da câmera que passa pelo CP é chamada de eixo óptico.

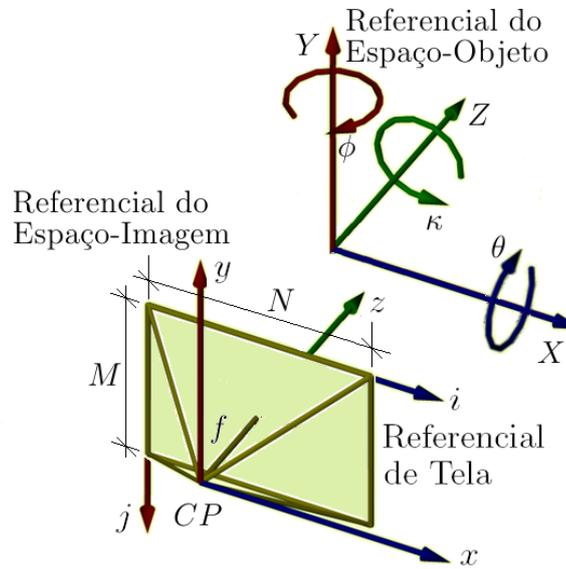


Figura 3.1: Referencial para o espaço-imagem e referencial para o espaço-objeto.

As coordenadas no referencial do espaço-imagem são obtidas a partir das coordenadas no referencial de tela definido nas figuras 2.1 e 3.1 através das seguintes operações:

$$\begin{cases} x = (i - N/2)E_x, \\ y = (M/2 - j)E_y, \\ z = f. \end{cases} \quad (3.1)$$

onde:

- (x, y, z) são as coordenadas no referencial do espaço-imagem;

- (i, j) são as coordenadas no referencial de tela;
- N e M são as dimensões horizontal e vertical da imagem em *pixels*;
- E_x e E_y são fatores de escala nas direções x e y .
- f é a distância focal.

Os sistemas adotados obedecem a regra da mão esquerda. Nestes sistemas, as rotações positivas são contadas no sentido horário quando o sistema é visto a partir de um eixo positivo e em direção à origem. Interpreta-se este sistema como sendo superposto à tela do computador, com valores positivos do eixo Z aparecendo mais distantes do observador.

3.2.1 Equações de Colinearidade Diretas

As equações que modelam um mapeamento de pontos contidos no espaço-objeto, para o espaço-imagem ou equações de colinearidade diretas, podem ser expressas por uma simples semelhança de triângulos, como mostra a figura 3.2.

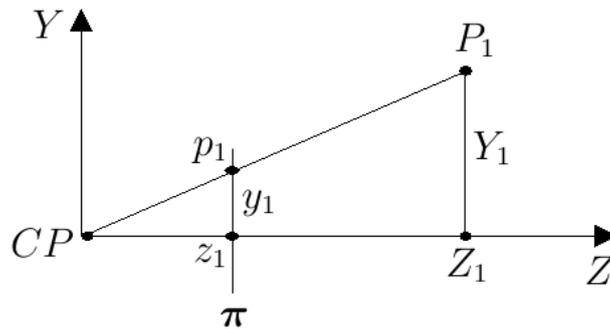


Figura 3.2: Formação da imagem

Na figura 3.2, P_1 é um ponto no referencial do espaço-objeto de coordenadas $(0, Y_1, Z_1)$ e p_1 é a imagem deste ponto no plano sensor π de coordenadas $(0, y_1, z_1)$. CP é o centro perspectivo e, portanto, o ponto z_1 é o lugar geométrico onde o eixo óptico intercepta o plano sensor (denominado ponto principal). Define-se então a distância focal da câmera f como sendo a distância $\overline{CPz_1}$. Por semelhança dos triângulos $\overline{P_1Z_1CP}$ e $\overline{p_1z_1CP}$ obtém-se que:

$$\frac{Y_1}{y_1} = \frac{Z_1}{z_1}, \quad (3.2)$$

e portanto

$$y_1 = \frac{z_1 Y_1}{Z_1}. \quad (3.3)$$

Por analogia à expressão 3.3, tem-se:

$$x_1 = \frac{z_1 X_1}{Z_1}, \quad (3.4)$$

onde:

- (x_1, y_1, z_1) são as coordenadas do ponto P_1 no referencial do espaço-imagem;
- (X_1, Y_1, Z_1) são as coordenadas do ponto P_1 no referencial do espaço-objeto.

Admitindo que o CP está na origem do sistema referencial do espaço-imagem, z_1 é igual a distância focal f , assim as coordenadas no referencial do espaço-imagem serão representadas por (x, y, f) , onde f é constante para qualquer ponto neste espaço. Portanto as equações 3.3 e 3.4 ficam:

$$y = \frac{fY_i}{Z_i} \quad \text{e} \quad (3.5)$$

$$x = \frac{fX_i}{Z_i}, \quad (3.6)$$

onde:

- X_i, Y_i e Z_i são coordenadas de um ponto P no sistema referencial do espaço-objeto;
- x, y e $z = f$ são coordenadas de um ponto p no sistema referencial do espaço-imagem;

As equações 3.5 e 3.6 são denominadas equações de colinearidade diretas e permitem dizer que um ponto P_1 no referencial do espaço-objeto, a imagem deste ponto no referencial do espaço-imagem (p_1) e o centro perspectivo CP são colineares, possibilitando determinar as coordenadas de um ponto no referencial do espaço-imagem a partir das coordenadas deste ponto no referencial do espaço-objeto.

Entretanto as equações 3.5 e 3.6 foram definidas considerando que a origem do sistema global de coordenadas coincide com a origem do sistema imagem de coordenadas (CP) e também não se considerou rotações entre esses dois sistemas. Assim sendo, para o caso geral onde o plano sensor não é paralelo ao plano XY , faz-se necessário primeiro, transformar as coordenadas do sistema referencial do espaço-objeto para o sistema referencial do espaço-imagem e a partir destas, utilizar as equações 3.5 e 3.6. Esta transformação entre os dois sistemas de coordenadas é modelada por uma transformação isogonal no espaço [17] dada pela equação 3.7:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \lambda \cdot \mathbf{R} \cdot \begin{bmatrix} X - X_{CP} \\ Y - Y_{CP} \\ Z - Z_{CP} \end{bmatrix}, \quad (3.7)$$

onde:

- (X, Y, Z) são coordenadas de um ponto P no sistema referencial do espaço-objeto;
- (X_{CP}, Y_{CP}, Z_{CP}) são coordenadas do centro perspectivo;
- λ é um fator de escala;
- \mathbf{R} é a matriz de rotação, definida na equação 3.8.

$$\mathbf{R} = \begin{bmatrix} \cos \phi \cos \kappa & \cos \theta \sin \kappa + \sin \theta \sin \phi \cos \kappa & \sin \theta \sin \kappa - \cos \theta \sin \phi \cos \kappa \\ -\cos \phi \sin \kappa & \cos \theta \cos \kappa - \sin \theta \sin \phi \sin \kappa & \sin \theta \cos \kappa + \cos \theta \sin \phi \sin \kappa \\ \sin \phi & -\sin \theta \cos \phi & \cos \theta \cos \phi \end{bmatrix}. \quad (3.8)$$

Desenvolvendo o produto matricial da equação 3.7, resulta:

$$X_i = \lambda [r_{11}(X - X_{CP}) + r_{12}(Y - Y_{CP}) + r_{13}(Z - Z_{CP})], \quad (3.9)$$

$$Y_i = \lambda [r_{21}(X - X_{CP}) + r_{22}(Y - Y_{CP}) + r_{23}(Z - Z_{CP})] \quad e \quad (3.10)$$

$$Z_i = \lambda [r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]. \quad (3.11)$$

Portanto as coordenadas X_i , Y_i e Z_i são as coordenadas X , Y e Z transformadas para o sistema referencial do espaço-imagem, e agora estas coordenadas estão no mesmo sistema com que foram definidas as equações 3.5 e 3.6, podendo ser substituídas nestas, resultando:

$$x = f \frac{\lambda [r_{11}(X - X_{CP}) + r_{12}(Y - Y_{CP}) + r_{13}(Z - Z_{CP})]}{\lambda [r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]} \quad e \quad (3.12)$$

$$y = f \frac{\lambda [r_{21}(X - X_{CP}) + r_{22}(Y - Y_{CP}) + r_{23}(Z - Z_{CP})]}{\lambda [r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})]}. \quad (3.13)$$

Cancelando λ nas expressões 3.12 e 3.13, tem-se:

$$x = f \frac{r_{11}(X - X_{CP}) + r_{12}(Y - Y_{CP}) + r_{13}(Z - Z_{CP})}{r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})} \quad e \quad (3.14)$$

$$y = f \frac{r_{21}(X - X_{CP}) + r_{22}(Y - Y_{CP}) + r_{23}(Z - Z_{CP})}{r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})}. \quad (3.15)$$

onde :

- x e y são coordenadas de um ponto no referencial do espaço-imagem;
- X, Y e Z são coordenadas de um ponto no referencial do espaço-objeto;
- f é a distância focal;

- X_{CP}, Y_{CP} e Z_{CP} são coordenadas do centro perspectivo da câmera;
- r_{ij} são os elementos da matriz de rotação \mathbf{R} , com $(i, j) = 1, 2, 3$, como definida pela equação 3.8

As equações 3.15 e 3.14 são chamadas de equações de colinearidade diretas para casos genéricos. Considerando o caso particular onde $[X_{CP} \ Y_{CP} \ Z_{CP}] = [0 \ 0 \ 0]$ e $[\kappa \ \phi \ \theta] = [0 \ 0 \ 0]$ estas equações degeneram-se para as equações 3.6 e 3.5, respectivamente.

3.2.2 Equações de Colinearidade Inversas

Utilizando as equações 3.5 e 3.6, pode-se escrevê-las de forma inversa tal que:

$$X_i = \frac{Z_i x}{f} \quad \text{e} \quad (3.16)$$

$$Y_i = \frac{Z_i y}{f}. \quad (3.17)$$

Pode-se notar que as coordenadas X_i e Y_i no sistema referencial do espaço-objeto dependem, das coordenadas x e y no sistema referencial do espaço-imagem, da distância focal f e também de Z_i , que é uma variável no sistema referencial do espaço-objeto. Portanto, conclui-se que a partir de uma única imagem, não é possível determinar as coordenadas tridimensionais no referencial do espaço-objeto, sem que ao menos uma dessas coordenadas (X_i , Y_i ou Z_i seja conhecida, caracterizando uma não inversibilidade. Porém, as equações de colinearidade diretas, podem ser escritas inversamente como em 3.16 e 3.17, com a ressalva de dependerem de uma coordenada do sistema referencial do espaço-objeto.

Como nas equações 3.5 e 3.6, as equações 3.16 e 3.17 estão definidas para um caso muito específico, onde a origem do sistema referencial do espaço-imagem (CP) coincide com a origem do sistema referencial do espaço-objeto e não há rotações entre os eixos dos dois sistemas. Para tornar tais expressões genéricas pode-se escrever a seguinte equação paramétrica de reta.

$$\frac{X - X_{CP}}{X_i - X_{CP}} = \frac{Y - Y_{CP}}{Y_i - Y_{CP}} = \frac{Z - Z_{CP}}{Z_i - Z_{CP}} = t, \quad (3.18)$$

onde t é uma constante para esta reta.

Fazendo simples manipulações algébricas obtém-se:

$$X - X_{CP} = (Z - Z_{CP}) \frac{X_i - X_{CP}}{Z_i - Z_{CP}} \quad \text{e} \quad (3.19)$$

$$Y - Y_{CP} = (Z - Z_{CP}) \frac{Y_i - Y_{CP}}{Z_i - Z_{CP}}. \quad (3.20)$$

De forma análoga à equação 3.7, tem-se que:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda \cdot \mathbf{R} \cdot \begin{bmatrix} X_i - X_{CP} \\ Y_i - Y_{CP} \\ Z_i - Z_{CP} \end{bmatrix}. \quad (3.21)$$

Fazendo-se algumas manipulações matriciais, obtém-se:

$$\begin{bmatrix} X_i - X_{CP} \\ Y_i - Y_{CP} \\ Z_i - Z_{CP} \end{bmatrix} = \frac{1}{\lambda} \mathbf{R}^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (3.22)$$

Passando as coordenadas do centro perspectivo (CP) para o segundo termo da equação, resulta:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \frac{1}{\lambda} \mathbf{R}^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} X_{CP} \\ Y_{CP} \\ Z_{CP} \end{bmatrix}. \quad (3.23)$$

Sabendo-se que \mathbf{R} é ortogonal (propriedade das matrizes de rotação), sua inversa é igual a sua transposta e realizando o produto matricial, obtém-se:

$$X_i = \frac{1}{\lambda} r_{11}x + r_{21}y + r_{31}z + X_{CP}, \quad (3.24)$$

$$Y_i = \frac{1}{\lambda} r_{12}x + r_{22}y + r_{32}z + Y_{CP} \quad e \quad (3.25)$$

$$Z_i = \frac{1}{\lambda} r_{13}x + r_{23}y + r_{33}z + Z_{CP}. \quad (3.26)$$

Substituindo as expressões 3.24 e 3.26 em 3.19, e 3.25 e 3.26 em 3.20, tem-se:

$$X - X_{CP} = (Z - Z_{CP}) \frac{\left(\frac{1}{\lambda} r_{11}x + r_{21}y + r_{31}z + X_{CP} \right) - X_{CP}}{\left(\frac{1}{\lambda} r_{13}x + r_{23}y + r_{33}z + Z_{CP} \right) - Z_{CP}} \quad e \quad (3.27)$$

$$Y - Y_{CP} = (Z - Z_{CP}) \frac{\left(\frac{1}{\lambda} r_{12}x + r_{22}y + r_{32}z + Y_{CP} \right) - Y_{CP}}{\left(\frac{1}{\lambda} r_{13}x + r_{23}y + r_{33}z + Z_{CP} \right) - Z_{CP}}. \quad (3.28)$$

Após algumas manipulações matemáticas, obtém-se:

$$X = X_{CP} + (Z - Z_{CP}) \frac{r_{11}x + r_{21}y + r_{31}z}{r_{13}x + r_{23}y + r_{33}z} \quad e \quad (3.29)$$

$$Y = Y_{CP} + (Z - Z_{CP}) \frac{r_{12}x + r_{22}y + r_{32}z}{r_{13}x + r_{23}y + r_{33}z}, \quad (3.30)$$

onde $z = f$.

As expressões 3.29 e 3.30 são as equações de colinearidade inversas para casos genéricos. No caso específico de $[X_{CP} \ Y_{CP} \ Z_{CP}] = [0 \ 0 \ 0]$ e $[\kappa \ \phi \ \theta] = [0 \ 0 \ 0]$, as expressões degeneram-se para as expressões 3.16 e 3.17.

3.3 Calibração da Câmera

A etapa na qual são determinados os parâmetros de uma câmera é chamada de calibração. Existem dois tipos de parâmetros:

- os parâmetros extrínsecos, referentes à localização e orientação:

Orientação do sistema imagem (κ, ϕ, θ);

Posição do centro perspectivo (X_{CP}, Y_{CP}, Z_{CP});

- os parâmetros intrínsecos, que estão relacionados com a construção física da câmera:

Posição do ponto principal (centro da imagem);

Distância focal;

Coefficientes de distorções do sistema de lentes;

Fator de escala entre as dimensões horizontais e verticais do *pixel*.

Como os parâmetros intrínsecos estão relacionados com as características da câmera, estes podem ser considerados fixos, e sua determinação pode ser feita apenas uma vez. Os parâmetros extrínsecos devem ser determinados toda vez que a câmera sofre algum movimento de rotação ou translação.

As equações apresentadas na seção 3.2 só consideram os parâmetros extrínsecos. Para se considerar os parâmetros intrínsecos, os seguintes pontos devem ser considerados, entre outros:

- A projeção do centro perspectivo da câmera sobre o plano sensor não coincide com o centro da imagem (ponto principal);
- Distorções do sistema de lentes;
- Escalas diferentes do *pixel* nas direções x e y .

O deslocamento do ponto principal é dado pela distância entre este ponto e o centro da imagem.

As distorções do sistema de lentes ocorrem devido a impossibilidade prática de construir uma superfície ideal para as lentes, que deveriam ser perfeitos parabolóides de revolução. Esta dificuldade origina a distorção radial simétrica, normalmente considerada em aplicações que requerem alta precisão.

Escala diferentes do *pixel* nas direções x e y resultam em um *pixel* não mais quadrado, mas sim retangular. Este problema está associado à frequência de transmissão dos dados da câmera não ser necessariamente igual à frequência de transmissão do conversor A/D [18], e/ou devido a própria construção do *pixel* não ser quadrado em algumas câmeras, sendo que neste caso, esta diferença de escala é nominal, sendo fornecida pelo fabricante.

Outros parâmetros podem ser considerados dependendo da aplicação e da precisão necessária.

O seguinte modelo [19] pode ser utilizado para se considerar também os parâmetros intrínsecos da câmera.

$$x = x_R - c_x + (x_R - c_x)(k_1 r^2) + (x_R - c_x)ds_x \quad \text{e} \quad (3.31)$$

$$y = y_R - c_y + (y_R - c_y)(k_1 r^2), \quad (3.32)$$

onde:

- x e y são coordenadas no referencial do espaço-imagem, isentas de erros sistemáticos;
- x_R e y_R são coordenadas reduzidas ao centro aproximado da imagem;
- c_x e c_y são coordenadas do centro de distorção radial das lentes;
- k_1 é o coeficiente de primeira ordem de distorção radial simétrica;

- ds_x é o fator de escala entre as dimensões horizontais e verticais do *pixel*;
- $r^2 = x_R^2 + y_R^2$

Para se fazer a calibração, é necessário conhecer as coordenadas de alguns pontos no referencial do espaço-imagem e no referencial do espaço-objeto. A resolução de um sistema de equações determina os parâmetros, e pode ser realizada de forma linear ou não-linear. Um exemplo de modelo matemático que pode ser utilizado é o descrito pelas equações 3.33 e 3.34, que nada mais são do que as equações de colinearidade diretas (equações 3.14 e 3.15) com as considerações das equações 3.31 e 3.32:

$$\begin{aligned} x_R - c_x + (x_R - c_x)k_1r^2 + (x_R - c_x)ds &= \\ = f \left(\frac{r_{11}(X - X_{CP}) + r_{12}(Y - Y_{CP}) + r_{13}(Z - Z_{CP})}{r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})} \right) & \text{ e} \end{aligned} \quad (3.33)$$

$$\begin{aligned} (y_R - c_y) + (y_R - c_y)(k_1r^2) &= \\ = f \left(\frac{r_{21}(X - X_{CP}) + r_{22}(Y - Y_{CP}) + r_{23}(Z - Z_{CP})}{r_{31}(X - X_{CP}) + r_{32}(Y - Y_{CP}) + r_{33}(Z - Z_{CP})} \right), & \end{aligned} \quad (3.34)$$

onde:

- (X_{CP}, Y_{CP}, Z_{CP}) são coordenadas do centro perspectivo no sistema referencial do espaço-objeto;
- (X, Y, Z) são coordenadas de um ponto no sistema referencial do espaço-objeto;
- r_{ij} são elementos da matriz de rotação entre os sistemas referenciais do espaço-objeto e espaço-imagem, os quais são função dos ângulos κ , ϕ e θ (definidos na seção 3.2), para $i, j = 1, 2, 3$;
- f é a distância focal;
- demais parâmetros como definidos para as equações 3.31 e 3.32.

Este modelo é não-linear e necessita que os pontos no espaço-objeto sejam não coplanares. Esta não coplanaridade dos pontos no espaço-objeto é necessária para evitar que a correlação de Z e f , ocasione singularidade na matriz que é invertida.

Como cada ponto fornece duas equações, um total de seis pontos de coordenadas conhecidas é suficiente para resolver este sistema de maneira sobre-determinada, através, por exemplo, do método dos mínimos quadrados [20].

3.4 Técnicas de Reconstrução

Existem várias técnicas de reconstrução tridimensional empregadas em visão computacional. As principais técnicas são descritas a seguir.

3.4.1 *Shape From Shading*

A técnica *Shape From Shading* é uma técnica ativa de reconstrução tridimensional de uma cena baseada em fotometria, ou seja, na informação radiométrica contida em uma imagem. Esta técnica consiste em determinar a relação existente entre o vetor normal de um ponto de coordenadas (X,Y,Z) , sobre uma superfície no espaço-objeto e a irradiância de sua imagem, de coordenadas (x,y) , no espaço-imagem. Esta relação é dada por um modelo de iluminação.

Um dos modelos de iluminação mais utilizados é o modelo de Lambert, que postulam basicamente que a irradiância é proporcional ao cosseno do ângulo formado entre o vetor normal à superfície e o vetor direção da fonte de iluminação, multiplicado pela intensidade da fonte de iluminação e o coeficiente de reflexão difuso (albedo). A equação 3.35 é o modelo de iluminação proposto por Lambert.

$$E = \rho I \cos(\theta), \quad (3.35)$$

onde:

- E é a irradiância;
- ρ é o coeficiente de reflexão difuso;
- θ é o ângulo formado entre o vetor normal a superfície e o vetor direção da fonte de iluminação.

A principal vantagem dessa técnica é que ela é monocular, evitando assim, o problema de correspondência presente em algumas outras técnicas. Porém, o modelo de iluminação de Lambert, assume que a superfície é perfeitamente difusa (*Lambertiana*), o que não condiz com a natureza dos objetos e superfícies reais. As desvantagens para esta abordagem são: o coeficiente de reflexão difusa é de difícil determinação, e a necessidade de, no mínimo, uma fonte de iluminação com posições conhecidas.

As referências [5, 21, 22, 23] tratam sobre esse assunto com mais detalhes.

3.4.2 *Shape From Motion*

A técnica denominada *Shape From Motion* utiliza parâmetros geométricos para a determinação da estrutura tridimensional da cena, além da variável tempo. Os modelos matemáticos empregados nesta técnica, consideram o deslocamento em função do tempo, da imagem, de pontos 3-D no plano imagem, portanto, neste caso existe a necessidade de se resolver o problema de correspondência, a exemplo da técnica *Shape From Stereo*, mas de duas imagens tomadas em tempos distintos e não em posições distintas.

Como se pode deduzir, esta técnica é monocular e passiva e há a necessidade que a câmera, ou os objetos, ou ambos, estejam em movimento. O fato destes estarem em movimento e a necessidade de grandes seqüências de imagens para obter resultados precisos, pode ser uma desvantagem dessa técnica em algumas aplicações.

Esta técnica está habilitada à reconstrução tridimensional apenas de objetos rígidos 3-D (suas dimensões tridimensionais não variam em função do tempo) o que pode ser vantajoso ou desvantajoso conforme a aplicação.

As referências [6, 24, 25, 26, 27, 28, 29] trazem informações mais detalhadas sobre esse assunto.

3.4.3 *Shape From Texture*

A técnica *Shape From Texture* consiste basicamente em projetar um padrão de textura sobre uma cena de interesse e capturar a imagem desta. Nesta técnica, a posição e orientação da câmera, bem como as do projetor, devem ser conhecidos. Esta é obviamente uma técnica ativa e monocular, porém, o problema de correspondência entre a imagem capturada e o padrão projetado existe e é semelhante ao problema de correspondência do *Shape From Stereo*. A principal vantagem desta em relação ao *Shape From Stereo*, é que o padrão a ser projetado pode ser elaborado com inferências geométricas e fotométricas, que facilitem a correspondência com a imagem. A maior desvantagem desta técnica é ser uma técnica ativa, limitando a sua aplicação a ambientes onde se pode controlar as condições de iluminação.

A reconstrução tridimensional é obtida a partir das relações geométricas entre o padrão de textura projetado e a imagem deste padrão.

As referências [6, 8, 30, 31, 32] tratam este assunto mais detalhadamente.

3.4.4 *Shape From Focus*

O *Shape from Focus* é uma técnica de reconstrução tridimensional que permite determinar as coordenadas 3-D de uma cena, a partir da modelagem dos efeitos que os parâmetros intrínsecos de uma câmera exercem sobre as imagens capturadas, com pequena profundidade de campo (região do espaço-objeto onde as imagens de um objeto são nítidas dentro de um certo critério de aceitação) . Os parâmetros intrínsecos utilizados são: a distância focal e a abertura do diafragma.

O fato primordial desta técnica está em tomar várias imagens de uma mesma cena variando a distância focal ou movimentando a câmera ou a cena na direção do eixo ótico da câmera, mantendo a profundidade de campo pequena. Com isso,

obtém-se várias imagens com pontos focalizados e desfocalizados. Um ponto de uma certa imagem nitidamente focalizado, determina a distância focal ou a profundidade (se a distância focal for mantida fixa) correta para a formação daquela imagem para aquele ponto.

A desfocalização de uma imagem pode ser visto como um processo de filtragem passa-baixa, onde as componentes de altas frequências diminuem com o aumento da desfocalização, portanto, o operador de medida de foco deve ser um detector de altas frequências, ou seja, um filtro passa-alta. Uma vez determinado os pontos focalizados em cada imagem, pode-se associar a profundidade com esses pontos ou calculá-la através da equação de lentes, no caso em que a distância focal variou e a câmera e a cena permaneceram fixos durante as tomadas.

As principais vantagens desta técnica estão em ser passiva e monocular, evitando o problema de correspondência de outras técnicas.

Devido a necessidade de pequena profundidade de campo, a abertura do diafragma da câmera deve ser grande, impossibilitando a utilização do modelo de câmera de orifício (fácil utilização). Outras desvantagens desta técnica são: ser extremamente sensível as condições de iluminação, e necessitar de várias imagens, aumentando o custo computacional.

São sugeridas as referências [7, 33, 34] sobre este assunto.

3.4.5 *Shape From Stereo*

A primeira técnica de reconstrução tridimensional (a partir de imagens) e, ainda hoje, a mais difundida, é a “visão estéreo” (*Shape From Stereo*), que consiste basicamente em utilizar duas ou mais câmeras em posições distintas para obter a informação tridimensional da cena, em relação ao sistema de coordenadas adotado.

A visão estéreo utiliza-se da disparidade entre as imagens como a fonte de informação para a recuperação da estrutura tridimensional, uma vez que esta se relaciona de maneira inversamente proporcional à profundidade da cena.

Em um sistema típico de visão estéreo composto por duas câmeras, a disparidade associada a um ponto qualquer em uma imagem pode ser entendido como o deslocamento aparente entre este ponto e o seu ponto homólogo na outra imagem.

A maior dificuldade desta técnica de reconstrução é a determinação dos pontos homólogos entre as duas imagens. Devido a sua importância, esta dificuldade é denominada de “problema de correspondência”.

A solução do problema de correspondência pode ser representada na forma de um mapa de disparidade. Para o caso discreto, este mapa pode ser representado por uma matriz cujo seus elementos são os valores de disparidade. Quando esta matriz não possui valores de disparidade para todos os seus elementos gera-se um mapa de disparidade esparso.

O mapa de disparidade esparso é uma solução parcial para o problema de correspondência, pois, a partir deste, não é possível reconstruir a estrutura tridimensional da cena como um todo.

O mapa de disparidade denso é uma matriz que possui valores de disparidade para todos os seus elementos, tornando possível a reconstrução completa da estrutura tridimensional da cena. Entretanto, este tipo de mapa de disparidade, apesar de denso, é gerado não permitindo variações abruptas entre valores de disparidade vizinhos, na maioria das publicações relacionadas, limitando com isso, as aplicações desta técnica para apenas a reconstrução de superfícies suaves.

Portanto, em aplicações gerais de reconstrução tridimensional, o mapa de disparidade gerado deve ser denso e permitir variações abruptas entre valores de disparidade vizinhos.

As referências [17, 35] tratam do assunto com mais detalhes.

3.4.6 Luz Estruturada

Em face às dificuldades de se estabelecer a correspondência de *pixels* homólogos entre as imagens de maneira autônoma, de se determinar os parâmetros de posição

e orientação das câmeras e de se sincronizar os instantes das tomadas das câmeras, pode-se adotar uma variação da técnica *Shape From Stereo* substituindo uma das câmeras por uma fonte de luz com estrutura conhecida projetada sobre a cena.

A técnica de reconstrução por luz estruturada envolve o uso de um projetor e uma câmera, na qual o sistema de projeção é usado para projetar um padrão de luz sobre o objeto na cena e uma câmera captura a imagem dos padrões que são distorcidos pela superfície dos objetos da cena. Qualquer ponto projetado na cena com direção e identificação conhecidas pode ser reconstruído usando o princípio de interseção dos raios de luz.

Embora os sistemas de reconstrução por luz estruturada minimizem o problema de iluminação do ambiente ao projetar luz estruturada, podem ser necessárias várias operações de realce das imagens, até as etapas de reconhecimento e medição precisa dos alvos projetados.

Os padrões projetados podem ser linhas, faixas, malhas e pontos, entre outros. Características como ângulo de incidência e cor da luz também podem ser utilizadas como auxílio à identificação da estrutura projetada à reconstrução.

No capítulo seguinte, busca-se apresentar um modelo para reconstrução de superfícies baseadas nesta técnica de luz estruturada.

Capítulo 4

Um Modelo Para Reconstrução de Superfícies

Neste trabalho, a técnica adotada para a reconstrução tridimensional utiliza luz estruturada. Foi adotado um feixe retilíneo de luz projetado sobre a cena como luz estruturada. A fonte de luz projeta o feixe sobre a cena e a interseção entre o plano formado pelo feixe de luz e a superfície dos objetos da cena, em conjunto com a imagem da câmera, são suficientes para se obter as coordenadas tridimensionais.

A figura 4.1 mostra a projeção do feixe de luz retilíneo incidindo sobre os objetos da cena com uma inclinação dada pelo ângulo α . As figuras 4.2(a) e 4.2(c) mostram a vista superior e a imagem obtida pela câmera, respectivamente. A figura 4.2(b) mostra a vista lateral, evidenciando a interseção entre o plano XY e o plano formado pelo feixe de luz e seu ângulo de incidência.

Neste caso, pode-se estabelecer relações simples entre as coordenadas (X, Y, Z) do objeto em um referencial escolhido e suas respectivas imagens no plano de projeção da câmera.

De uma forma geral, as coordenadas que definem a posição do objeto com relação ao referencial do espaço-objeto (X, Y, Z) não são disponíveis de forma direta somente com uma imagem. Através da estrutura do feixe de luz incidindo sobre a cena, é possível determinar as coordenadas das superfícies dos seus objetos no referencial

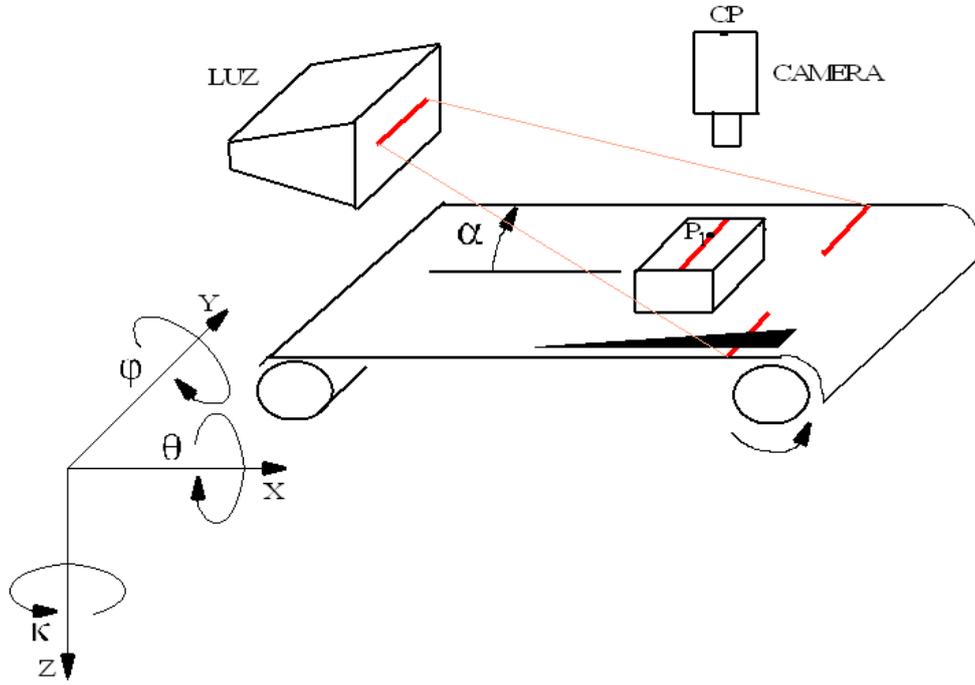


Figura 4.1: Esquema do modelo utilizado.

global.

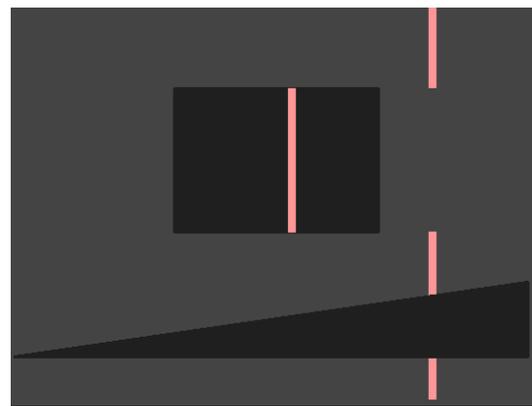
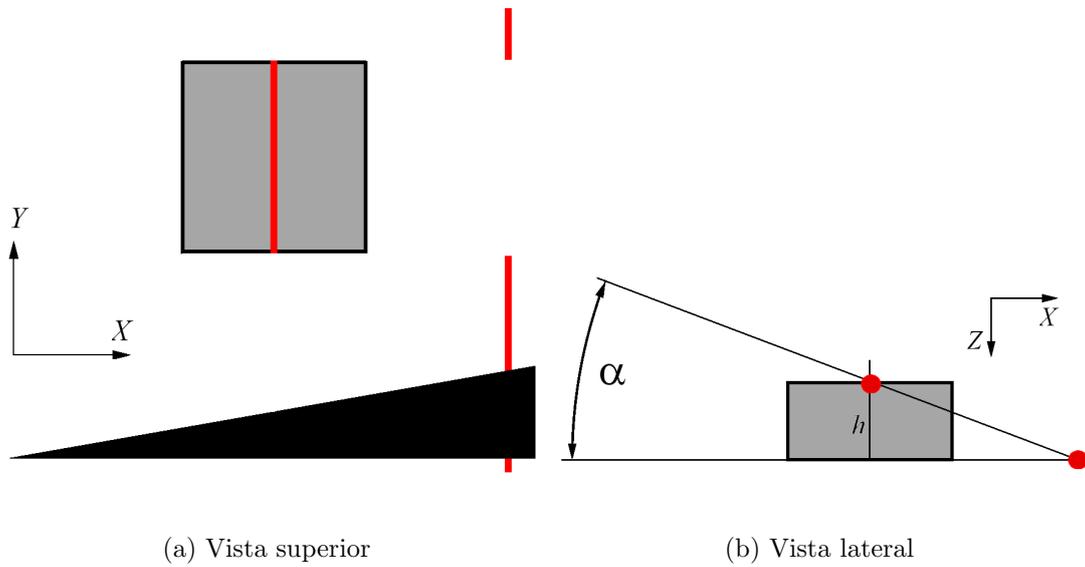
Seja um ponto P_1 genérico pertencente à superfície do objeto analisado atingida pelo feixe de luz de coordenadas (i_1, j_1) no referencial de tela (Figura 4.3).

As coordenadas x_1 e y_1 no referencial do espaço-imagem do ponto são obtidas aplicando sistema de equações 4.1, repetido abaixo:

$$\begin{cases} x = (i - N/2)E_x, \\ y = (M/2 - j)E_y, \\ z = f. \end{cases} \quad (4.1)$$

Uma vez conhecidas as equações 3.29 e 3.30, pode-se determinar as coordenadas X_1 e Y_1 de um ponto no referencial do espaço-objeto em função das suas coordenadas no referencial do espaço-imagem e de Z_1 , dadas pelas equações 4.2 e 4.3.

$$X_1 = X_{CP} + (Z_1 - Z_{CP}) \frac{r_{11}x_1 + r_{21}y_1 + r_{31}f}{r_{13}x_1 + r_{23}y_1 + r_{33}f}. \quad (4.2)$$



(c) Exemplo de imagem obtida pela câmara

Figura 4.2: Vistas e imagem obtida pela câmara.

$$Y_1 = Y_{CP} + (Z_1 - Z_{CP}) \frac{r_{12}x_1 + r_{22}y_1 + r_{32}f}{r_{13}x_1 + r_{23}y_1 + r_{33}f}. \quad (4.3)$$

Por sua vez, a coordenada Z_1 é dada pela interseção do plano descrito pelo feixe de luz com a superfície dos objetos da cena, dado pela equação 4.4 (Figuras 4.4 e 4.5).

$$Z_1 = \tan \alpha X_m. \quad (4.4)$$

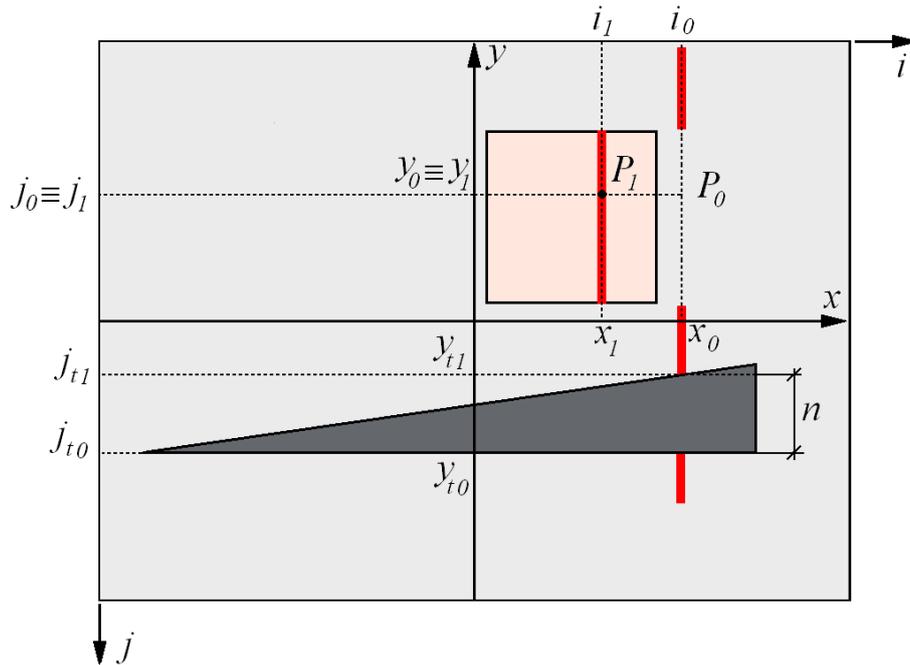


Figura 4.3: Determinação das coordenadas no referencial de tela e do espaço-imagem.

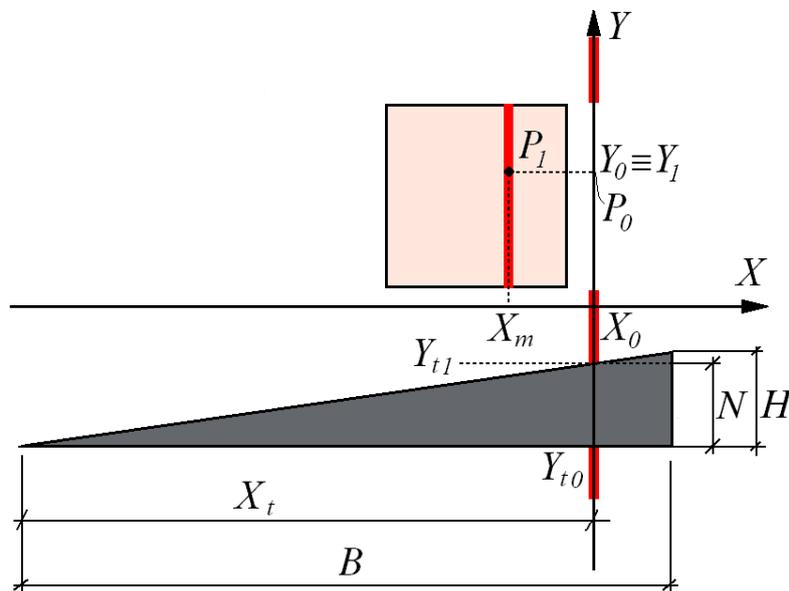


Figura 4.4: Determinação das coordenadas no referencial do espaço-objeto.

onde α é o ângulo de incidência do feixe de luz sobre o plano XY e Z_1 e X_m encontram-se indicado na figura 4.4.

Substituindo a equação 4.4 nas equações 4.2 e 4.3, e considerando que o plano sensor da câmera está paralelo ao plano XY , ou seja, não há rotação da câmera em

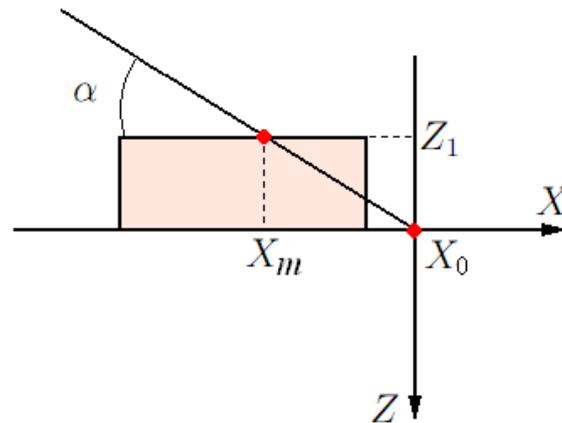


Figura 4.5: Detalhes da vista lateral.

relação ao referencial do espaço-objeto, obtemos as coordenadas X_m (equação 4.5) e Y_1 (equação 4.6) em função das coordenadas no referencial do espaço-imagem.

$$X_m = \frac{-X_{CP} + Z_{CP} \frac{x_1}{f}}{\tan \alpha \frac{x_1}{f} - 1} \quad \text{e} \quad (4.5)$$

$$Y_1 = Y_{CP} + \tan \alpha X_m - Z_{CP} \frac{y_1}{f}. \quad (4.6)$$

Já que o objetivo é obter as coordenadas de superfície dos objetos da cena, é necessário fazer o feixe de luz percorrer esta superfície. Isto é feito mantendo-se constante a luz estruturada que incide sobre o objeto analisado e aplicando-se deslocamentos horizontais ao conjunto formado pelo objeto analisado e o plano onde o mesmo se apóia. O controle dos deslocamentos horizontais impostos é feito através de um triângulo fixado no plano de sustentação do objeto que se movimenta de forma solidária ao mesmo. Através de relações de semelhança de triângulos, é possível calcular indiretamente os deslocamentos horizontais do objeto. Isto pode se tornar necessário quando estes deslocamentos não são uniformemente aplicados em função do tempo.

O artifício do triângulo permite a adoção da metodologia proposta em problemas dinâmicos como, por exemplo, uma esteira numa linha de produção. As coordenadas y_{t0} e y_{t1} mostradas na figura 4.3 são usadas para medir este deslocamento horizontal.

Aplicando novamente a equação 4.3 a y_{t0} e y_{t1} , obtêm-se Y_{t0} e Y_{t1} mostradas na figura 4.4:

$$Y_{t0} = Y_{CP} - Z_{CP} \frac{y_{t0}}{f}, \quad (4.7)$$

$$Y_{t1} = Y_{CP} - Z_{CP} \frac{y_{t1}}{f}. \quad (4.8)$$

A distância N , também mostrada na figura 4.4 é determinada pela equação 4.9 e o deslocamento X_t , que representa o deslocamento horizontal imposto ao objeto, pela equação 4.10

$$N = Y_{t1} - Y_{t0}. \quad (4.9)$$

$$X_t = N \frac{B}{H}. \quad (4.10)$$

Substituindo 4.9 em 4.10, tem-se:

$$X_t = (Y_{t1} - Y_{t0}) \frac{B}{H}. \quad (4.11)$$

Substituindo 4.7 e 4.8 em 4.11, tem-se:

$$X_t = \left[\left(Y_{CP} - Z_{CP} \frac{y_{t1}}{f} \right) - \left(Y_{CP} - Z_{CP} \frac{y_{t0}}{f} \right) \right] \frac{B}{H}, \quad (4.12)$$

que simplificando fica:

$$X_t = (y_{t0} - y_{t1}) \frac{Z_{CP} B}{f H}. \quad (4.13)$$

Sabendo-se que da figura 4.3 é possível relacionar a distância n às variáveis y_{t1} e y_{t0} :

$$n = y_{t1} - y_{t0}, \quad (4.14)$$

a equação 4.13 pode escrita como:

$$X_t = -n \frac{Z_{CP} B}{f H}. \quad (4.15)$$

Aplicando-se agora o deslocamento X_t ao ponto P_1 , tem-se a coordenada X_1 :

$$X_1 = X_m + X_t. \quad (4.16)$$

Substituindo então as equações 4.5 e 4.15 na equação 4.16, temos finalmente a coordenada X_1 :

$$X_1 = \frac{-X_{CP} + Z_{CP} \frac{x_1}{f}}{\tan \alpha \frac{x_1}{f} - 1} - n \frac{Z_{CP} B}{f H}. \quad (4.17)$$

Generalizando P_1 para um ponto P qualquer da superfície iluminada pelo feixe de luz, as suas coordenadas no referencial do espaço-objeto é dada pela solução do seguinte sistema de equações:

$$\begin{cases} X = \frac{-X_{CP} + Z_{CP} \frac{x_1}{f}}{\tan \alpha \frac{x_1}{f} - 1} - n \frac{Z_{CP} B}{f H}, \\ Y = Y_{CP} + (Z - Z_{CP}) \frac{y}{f} \text{ e} \\ Z = \tan \alpha X_m, \end{cases} \quad (4.18)$$

onde:

- (X, Y, Z) são as coordenadas no referencial do espaço-objeto;
- (x, y, z) são as coordenadas no referencial do espaço-imagem, com $z = f$;

- (X_{CP}, Y_{CP}, Z_{CP}) são as coordenadas do centro de perspectiva no referencial do espaço-objeto;
- X_m definido na equação 4.5;
- X_t definido na equação 4.15;
- demais parâmetros como já definidos anteriormente neste capítulo.

Se o plano sensor da câmera for considerado não paralelo ao plano XY , definem-se de forma análoga o sistema de equações 4.18 sem as simplificações adotadas a partir da equação 4.5

Repetindo-se os procedimentos usados nos cálculos das coordenadas (X, Y, Z) (equação 4.18) para cada ponto contido na parcela da luz estruturada que incide sobre o objeto analisado e para cada deslocamento horizontal imposto, é possível se fazer a reconstrução da superfície do objeto em estudo. Obviamente esta metodologia fica sujeita a problemas de sombras. Entretanto, podem ser minimizados usando-se diversos feixes de luz.

Capítulo 5

Validação da Metodologia

No intuito de verificar os resultados que podem ser obtidos com a técnica proposta, faz-se a análise detalhada de um exemplo preliminar: a reconstrução tridimensional de um cubo de 7cm de aresta que se move através de uma esteira rolante, sob um feixe de luz retilíneo e observado por uma câmera, semelhante ao esquema mostrado na figura 4.1. As etapas do processo de reconstrução são detalhadas nas seções a seguir.

5.1 Validação Através de Vídeos Gerados Artificialmente

5.1.1 Geração Artificial de Imagens

Visando a verificação do método proposto, foi utilizado o *software* Pov-Ray® [36] para a geração artificial das imagens do experimento. Desta forma são simuladas condições ideais de iluminação, posição e orientação da câmera e deslocamento dos objetos da cena, minimizando os possíveis ruídos e imprecisões durante as análises. O código escrito em linguagem própria do *software* Pov-Ray® mostrado na figura 5.1 foi usado para se gerar a imagem da figura 5.2.

```
#include "colors.inc" #include "shapes.inc"
//definição da câmera
camera{orthographic
    location <10,30,-5> //Centro perspectivo da câmera
    look_at <10,0,5>
    rotate<0,0,30>//kapa-> rotate<0,0,k> rotação em torno do eixo z
    rotate<0,30,0>//fi-> rotate<0,fi,0> rotação em torno do eixo y
    rotate<-50,0,0>//omega-> rotate<w,0,0> rotação em torno do eixo x
}
//definição da fonte luz
light_source {<0,20,-0> color White}
//definição do fundo
background { color rgb <0.3, 0.3, 0.3> }
//definição de um plano
plane {
    <0,1,0>,0
    pigment{color red 1 green 1 blue 1}
    no_reflection
}
//definição de um cubo amarelo de 7cm de aresta
box { <19,0,2>,<26,7,9>
    pigment {color Yellow}
}
```

Figura 5.1: Código do Pov-Ray® para geração da imagem da figura 5.2

O *software* adotado também é capaz de gerar os quadros de um vídeo. A figura 5.3 mostra três quadros do vídeo do cubo gerado artificialmente. Estas figuras apresentam a visão da câmera com luz ambiente e a correspondente imagem sem luz ambiente, sendo que esta última é usada no procedimento de reconstrução tridimensional.

Os quadros gerados pelo Pov-Ray® são transformados em vídeo através de outro *software* livre, o VirtualDub® [37], onde também pode ser controlada a taxa de

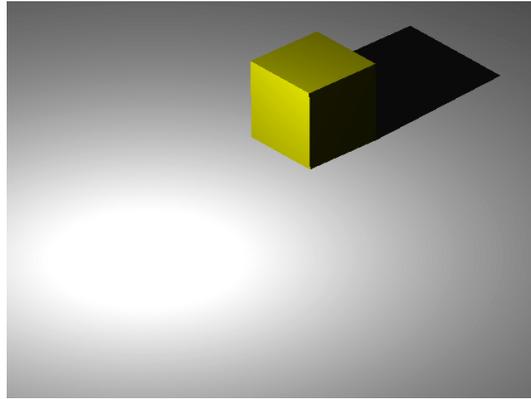


Figura 5.2: Imagem gerada artificialmente

amostragem, ou *frame rate*, quando for necessário.

Assim sendo, um vídeo simulando a imagem sem luz ambiente captada por uma câmera, para um cubo de 7cm de aresta que se move através de uma esteira rolante num esquema como o detalhado na figura 4.1, foi gerado artificialmente. Ao vídeo gerado foi aplicada a metodologia de reconstrução tridimensional proposta neste trabalho.

5.1.2 Calibração da Câmera

O modelo de formação das imagens utilizado no sistema proposto é o definido pelas equações de colinearidade diretas (equações 3.33 e 3.34) com as seguintes simplificações:

- A projeção do centro de perspectiva da câmera sobre o plano sensor coincide com o centro da imagem (ponto principal), ou seja, $c_x = 0$ e $c_y = 0$;
- Não existe distorção radial simétrica ($k_1 = 0$) e
- Dimensões horizontal e vertical do *pixel* iguais $ds_x = 1$.

Estas simplificações podem ser feitas pois o modelo da câmera utilizado na geração artificial de imagens utiliza as mesmas considerações.

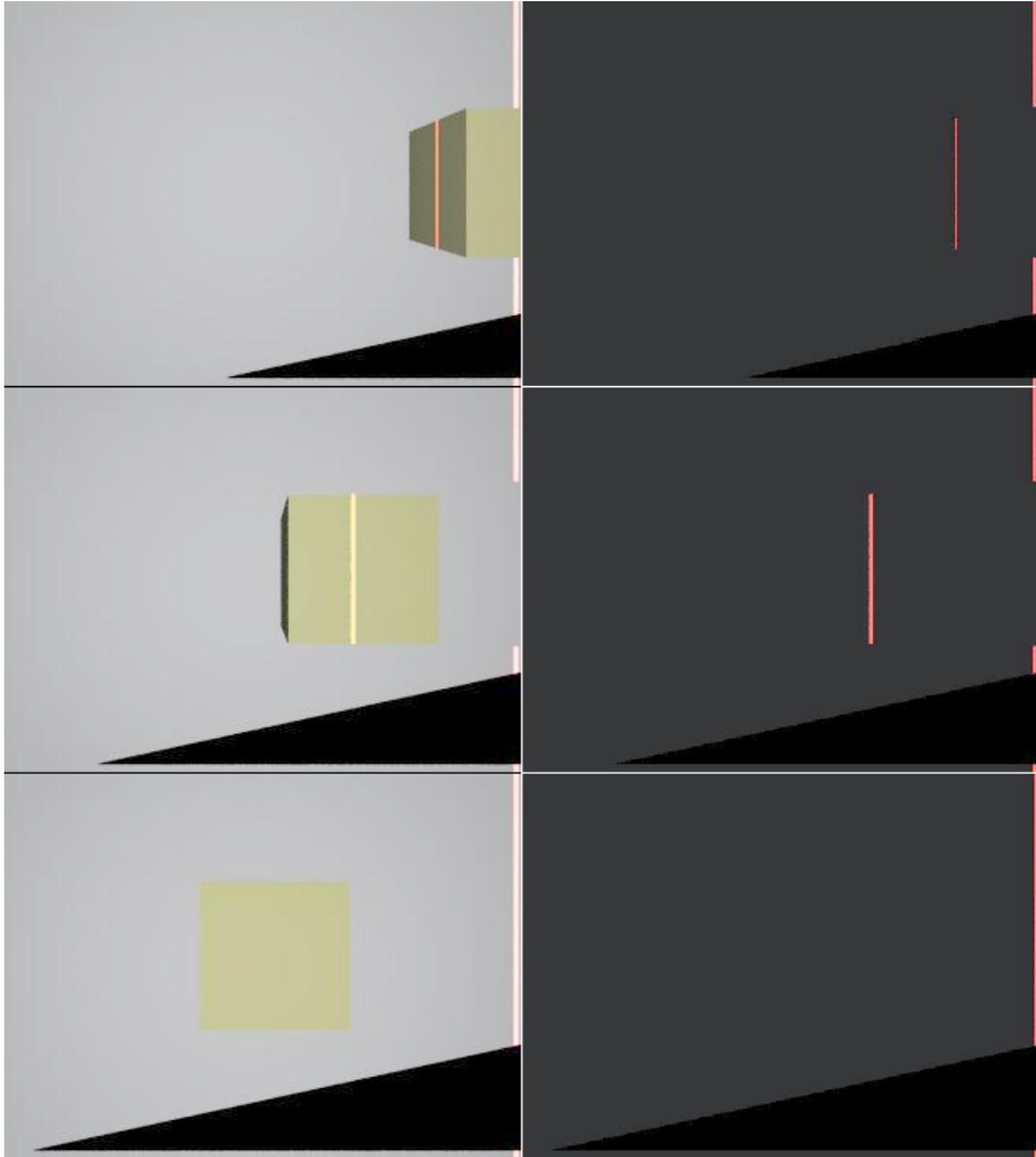


Figura 5.3: Sequência de imagens obtida artificialmente

Após as simplificações descritas anteriormente, as equações do modelo da câmera resultam em:

$$x = f \left(\frac{r_{11}(X_G - X_{CP}) + r_{12}(Y_G - Y_{CP}) + r_{13}(Z_G - Z_{CP})}{r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})} \right) \quad \text{e} \quad (5.1)$$

$$y = f \left(\frac{r_{21}(X_G - X_{CP}) + r_{22}(Y_G - Y_{CP}) + r_{23}(Z_G - Z_{CP})}{r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})} \right). \quad (5.2)$$

sendo:

$$\mathbf{R} = \begin{bmatrix} \cos \phi \cos \kappa & \cos \theta \sin \kappa + \sin \theta \sin \phi \cos \kappa & \sin \theta \sin \kappa - \cos \theta \sin \phi \cos \kappa \\ -\cos \phi \sin \kappa & \cos \theta \cos \kappa - \sin \theta \sin \phi \sin \kappa & \sin \theta \cos \kappa + \cos \theta \sin \phi \sin \kappa \\ \sin \phi & -\sin \theta \cos \phi & \cos \theta \cos \phi \end{bmatrix}, \quad (5.3)$$

como já definido no capítulo 3 onde a posição da câmera é definida pelas coordenadas do seu centro perspectivo CP (X_{CP}, Y_{CP}, Z_{CP}) no referencial do espaço-objeto e sua orientação pelas rotações κ (em torno do eixo Z), ϕ (em torno do eixo Y), e θ (em torno do eixo X). A distância focal f é a distância entre o CP e o plano sensor da câmera.

Para um dado conjunto de n pontos $(X_i, Y_i, Z_i), i = 1..n$ conhecidos, com as correspondentes coordenadas (x_i, y_i) determinadas e utilizando as equações 5.1 e 5.2 tem-se um conjunto de $2n$ equações onde as incógnitas são os parâmetros $(f, \phi, \kappa, \theta, X_{CP}, Y_{CP}, Z_{CP})$.

Assim, n deve ser no mínimo igual à metade do número de incógnitas a fim de determinar de maneira única os parâmetros do modelo matemático descrito nas equações 5.1 e 5.2. Geralmente o número de pontos observados é maior que a metade do número de incógnitas, tornando o sistema de equações sobre-determinado.

Utilizando 8 pontos obtidos de uma imagem de controle onde se arbitra através de comandos do *software* Pov-Ray® os parâmetros $f, \kappa, \theta, \phi, X_{CP}, Y_{CP}, Z_{CP}$ da câmera e para os quais as coordenadas no referencial do espaço-imagem e no referencial do espaço-objeto são conhecidas (Figura 5.4), foram obtidos os valores apresentados na tabela 5.1.

Esta tabela também apresenta os valores impostos durante a geração das imagens mostrando que a calibração obteve resultados semelhantes. As pequenas diferenças apresentadas são devidas a imprecisões quando se calculam as coordenadas dos pontos da superfície do cubo no referencial da câmera, já que estes são gerados em números inteiros de *pixels* e, por isso, sempre são arredondados para o *pixel* mais próximo. Outro fator que pode contribuir é a resolução da câmera, que define a

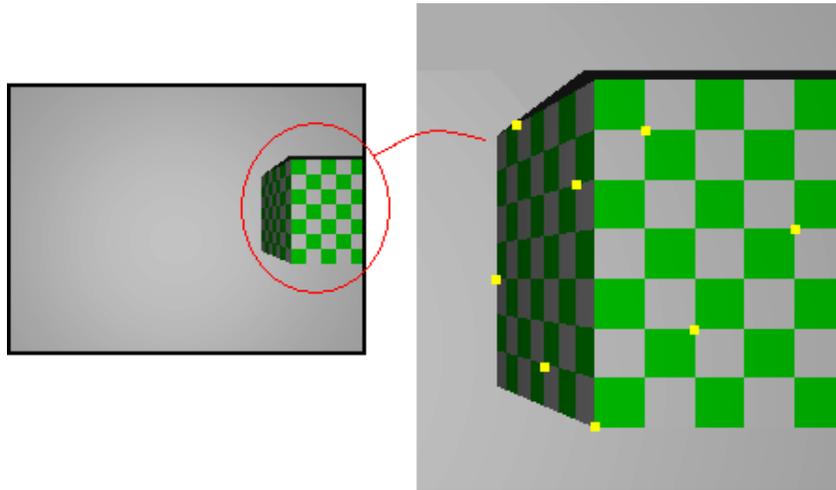


Figura 5.4: Imagem de controle com as coordenadas conhecidas

Tabela 5.1: Parâmetros obtidos pela calibração e impostos na geração das imagens

	Calibração	Impostos
f (cm)	1,002	1
κ (rad)	0,001	0
θ (rad)	-0,001	0
ϕ (rad)	0,001	0
X_{CP} (cm)	-16,342	-16,321
Y_{CP} (cm)	4,542	4,5
Z_{CP} (cm)	-24,992	-25

discretização da imagem.

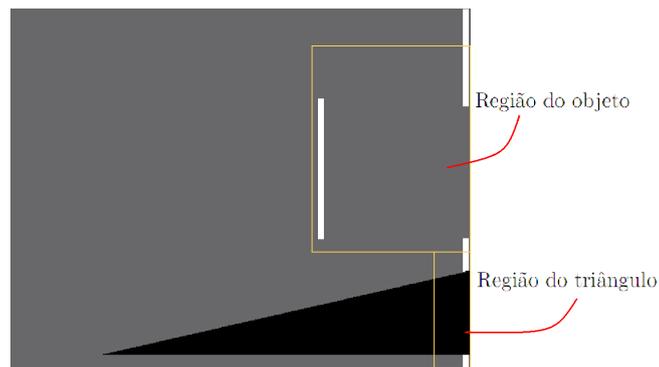
5.1.3 Processamento das Imagens

As etapas do processamento das imagens detalhadas a seguir são repetidas para cada quadro do vídeo. O algoritmo aplicado foi implementado no programa Matlab® [38] com a *toolbox* de processamento de imagens.

a) Seleção das Regiões de Interesse

Com o intuito de diminuir o tempo de processamento, nesta etapa é selecionada

uma região de interesse na imagem de forma que esta região seja a menor possível e forneça todas as informações necessárias. A figura 5.5(a) é um exemplo de um quadro do vídeo e as figuras 5.5(b) e 5.5(c) são as regiões de interesse extraídas da figura 5.5(a)



(a) Quadro do vídeo



(b) Região de processamento do objeto

(c) Região de processamento do triângulo

Figura 5.5: Seleção das regiões de interesse

b) Segmentação

Nesta etapa é identificada a interseção entre o plano formado pelo feixe de luz e os objetos da cena. A operação de segmentação utilizada é a limiarização com limiar global, segundo a regra definida pela equação 2.2. Como a imagem resultante é binária (0's e 1's), esta operação também pode ser chamada de binarização.

O limiar global é aplicado tanto na imagem da região de processamento do objeto quanto na do triângulo de controle (Figuras 5.5(b) e 5.5(c)). A figura 5.6 mostra o

resultado desta etapa.

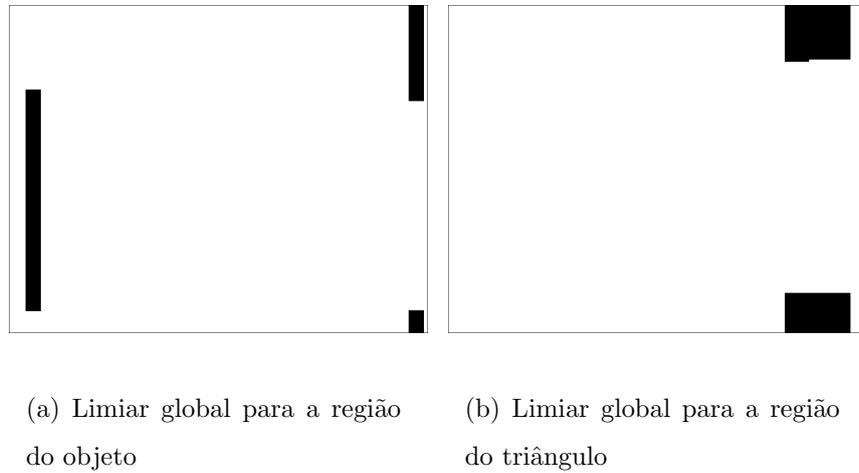


Figura 5.6: Resultados do processo de limiarização (cores invertidas)

c) Rotulação

Em algumas partes do algoritmo de processamento de imagens é necessário controlar se o laser está incidindo sobre o objeto e/ou sobre o triângulo. Este controle é feito pelo processo de rotulação, como definido na seção 2.5.2.

A figura 5.7 mostra o resultado da rotulação utilizando as imagens que passaram pelo processo de binarização. As cores diferentes indicam o “nome” dado a cada região pela rotulação em escala de cinza.

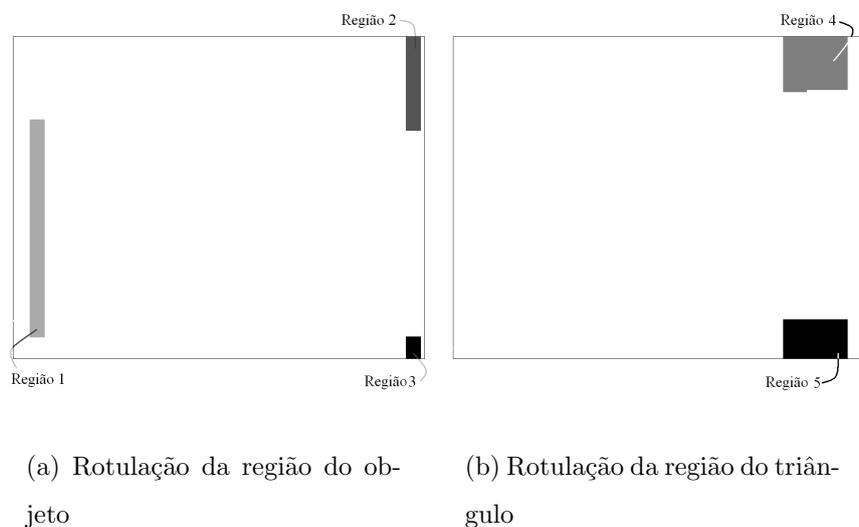


Figura 5.7: Resultados do processo de rotulação (cores invertidas)

d) Extração de Coordenadas no Referencial do Espaço-imagem

As coordenadas no referencial do espaço-imagem dos pontos de interesse são obtidas nesta etapa. Como a espessura do feixe de luz é maior que um *pixel*, é feita uma operação para encontrar a linha média do feixe (figura 5.8). Esta linha é o lugar geométrico definido pelos centróides de cada linha da região identificada da imagem.

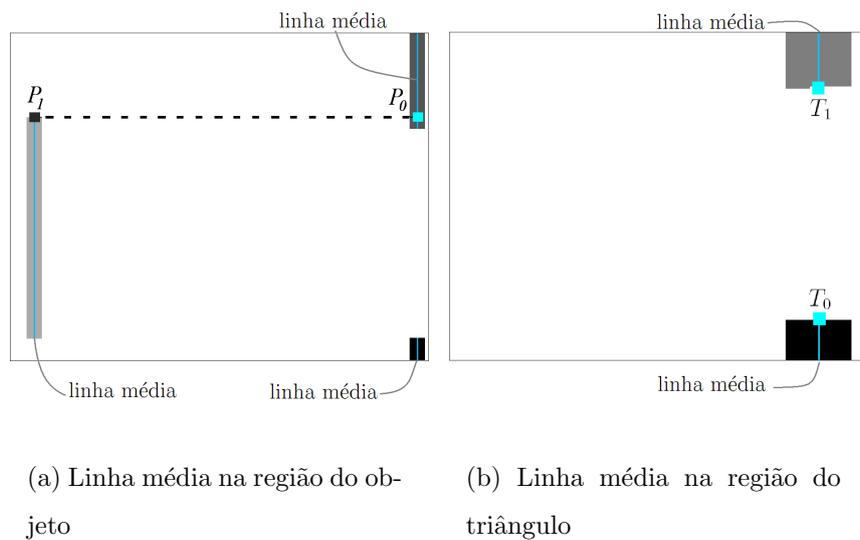


Figura 5.8: Cálculo da linha média (cores invertidas)

As coordenadas (i, j) no referencial de tela de todos os *pixels* da linha média na região do objeto são tomadas e a partir delas são obtidas as coordenadas no referencial do espaço-imagem aplicando a equação 4.1. Na região do triângulo são calculadas apenas as coordenadas y_{t0} e y_{t1} (pontos destacados na figura 5.8(b)).

Para exemplificar esta etapa, serão encontradas as coordenadas dos pontos P_0 , P_1 , y_{t0} e y_{t1} na figura 5.8(a). Primeiramente tomam-se as coordenadas no referencial de tela. A unidade utilizada neste referencial é o *pixel*:

$$\begin{aligned} i_{P_0} &= 635,5; & i_{P_1} &= 433,5; & j_{t_0} &= 463; \\ j_{P_0} &= 122; & j_{P_1} &= 122; & j_{t_1} &= 349,875. \end{aligned} \quad (5.4)$$

Para saber o quanto o ponto já foi deslocado, recorre-se à figura 5.8(b), extraíndo

as coordenadas no referencial de tela j_{t0} e j_{t1} dos pontos T_0 e T_1 , como definidas na figura 4.3.

Para a equação 4.1 ser aplicada, os fatores de escala devem ser definidos. Para as imagens geradas artificialmente, os seus valores são:

$$E_x = E_y = 480. \quad (5.5)$$

Sabendo-se que as dimensões das imagem produzem $N = 480$ e $M = 640$, as coordenadas no referencial do espaço-imagem podem ser calculadas com a equação 4.1:

$$\begin{aligned} x_0 = 0,6573 \quad x_1 = 0,2366 \quad y_{t0} = -0,2289 \\ y_0 = 0,2458 \quad y_1 = 0,2458 \quad y_{t1} = -0,4646 \end{aligned} \quad (5.6)$$

e) Reconstrução Tridimensional

Nesta etapa é feita a reconstrução propriamente dita. Por meio da equação 4.18 e das coordenadas no referencial do espaço-imagem definidas na etapa anterior, as coordenadas no referencial do espaço-objeto são obtidas.

Algumas coordenadas não são obtidas de forma direta apenas com a aplicação da equação 4.18. O cálculo do deslocamento através do triângulo de controle e sua aplicação ao objeto são feitas nesta etapa.

Continuando o exemplo do ponto P_1 , duas coordenadas no referencial do espaço-objeto são obtidas através do sistema de equações 4.18. As coordenadas X_0 e X_1 , São calculadas pela equação 4.5, com o resultado mostrado a seguir:

$$\begin{aligned} X_0 = 0,112 \\ X_1 = -12,055 \end{aligned} \quad (5.7)$$

Uma observação importante nesta etapa da metodologia é que X_0 deveria ser 0, de acordo com o referencial adotado na figura 4.4. Este fato não se verifica por

imprecisões devidas às simplificações e aproximações já comentadas. Então, fazendo-se o cálculo da equação 5.8 para ajustar a coordenada X_m ao referencial escolhido, tem-se

$$X_m = X_1 - X_0 = -12,167 \quad (5.8)$$

O cálculo do deslocamento X_t é feito pela equação 4.15, sabendo-se que as dimensões do triângulo são $B = 32$ e $H = 7$.

$$X_t = 26,935 \quad (5.9)$$

A coordenada X é dada por:

$$X = X_m + X_t = 14,768 \quad (5.10)$$

A coordenada Z é dada por:

$$Z = \tan \alpha X_m = -7,024 \quad (5.11)$$

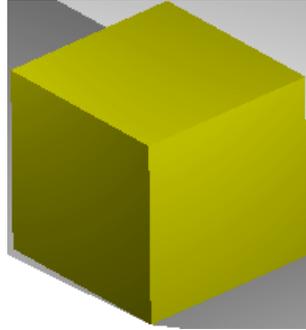
e finalmente, Y é dado por:

$$Y = Y_{CP} + (Z - Z_{CP}) \frac{y}{f} = 8,957 \quad (5.12)$$

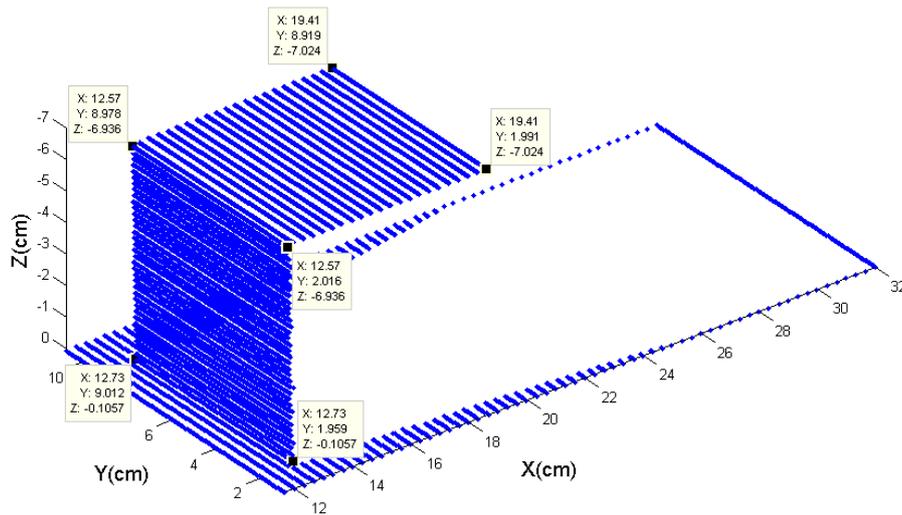
5.1.4 Apresentação dos Resultados

A apresentação dos resultados é feita com a simples impressão dos pontos com coordenadas no referencial do espaço-objeto obtidas pela aplicação da metodologia proposta para todos os pontos da luz estruturada que incide sobre o objeto e para todos os quadros do vídeo gerado artificialmente. A figura 5.9(b) mostra o resultado para o cubo, que, quando comparado com os valores esperados, observa-se que o resultado obtido é bastante próximo do cubo com 7cm de aresta mostrado na figura

5.9(a). Regiões não identificadas são devidas a partes do exemplo em estudo não atingidas pelo feixe de luz (sombras).



(a) Cubo



(b) Coordenadas do cubo 1

Figura 5.9: Resultado do processamento do cubo

Os pontos marcados na figura 5.9(b) são os vértices identificados. Sabendo-se a dimensão da aresta do cubo é de 7 cm, calcula-se as dimensões de a , b e c , respectivamente paralelas aos eixos X , Y e Z . Os erros absolutos e relativos são mostrados na tabela 5.2, onde a , b e c são as arestas paralelas aos eixos X , Y e Z , respectivamente.

Os erros cometidos durante a reconstrução podem ser melhorados aumentando a resolução das imagens geradas e diminuindo a velocidade do deslocamento dos

Tabela 5.2: Parâmetros obtidos pela calibração e impostos na geração das imagens

	Exato(cm)	Reconstrução(cm)	Erro absoluto(cm)	Erro relativo(%)
<i>a</i>	7	6,840	0,160	2,29
<i>b</i>	7	6,962	0,038	0,54
<i>c</i>	7	6,830	0,170	2,43

objetos da cena ou com aumento da taxa de amostragem da câmera.

5.2 Validação Através de Ensaio Experimental

5.2.1 Aquisição Experimental de Imagens

A aquisição experimental de imagens é feita através de um ensaio semelhante ao proposto pela geração artificial de imagens. O sistema experimental é composto pelos seguintes itens que seguem, alguns deles detalhados na figura 5.10

- Uma câmera digital. Resolução VHS (640 colunas e 480 linhas de *pixels*) e aquisição de 30 quadros por segundo;
- Uma fonte de luz retilínea.(Figuras 5.10 e 5.11). Esta fonte de luz foi obtida através da conexão de uma caneta apontadora *laser* num resfriador de processador para computador que, quando colocado em alta rotação, faz com que o apontador *laser* gere um plano¹.
- Objeto a ser estudado (Figura 5.12);
- Esteira para aplicar deslocamento aos objetos da cena (Figura 5.12 e 5.13);

¹A superfície gerada pela rotação da caneta apontadora é um plano desde que o ângulo entre o eixo de rotação do resfriador e a direção para qual o laser é apontado for de 90°. Para ângulos diferentes de 90°, a superfície gerada é cônica

- Triângulo de controle sobre a esteira para calcular de forma indireta o seu deslocamento (Figuras 5.12);
- Câmara escura para garantir que somente o feixe de luz refletido apareça na imagem (Figura 5.13).

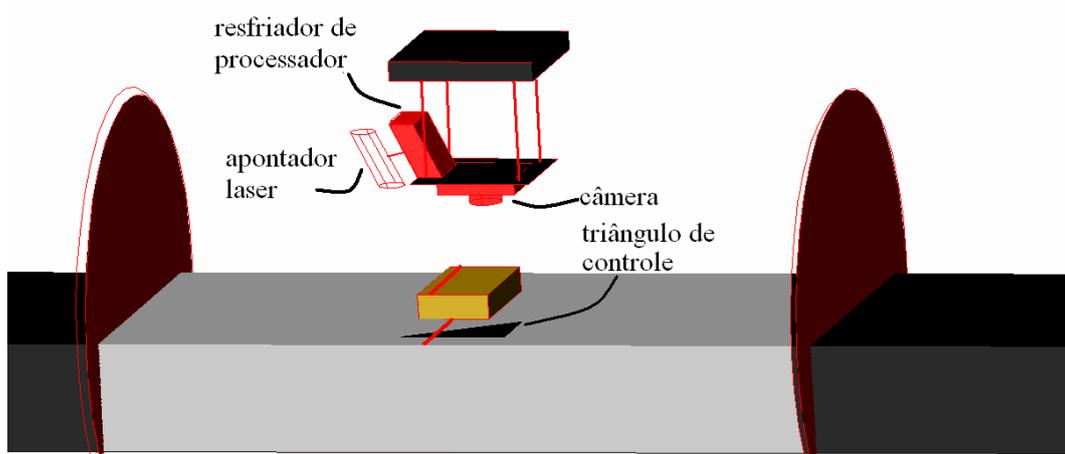


Figura 5.10: Detalhamento do sistema experimental para reconstrução tridimensional



Figura 5.11: Detalhe da fonte de luz - conexão entre o apontador *laser* e o resfriador de processador

A figura 5.14 mostra quadros do vídeo obtido pela câmera através de um ensaio que tenta reproduzir as mesmas condições propostas na seção 5.1.1, entretanto, como todo experimento real, fica sujeito à diversas fontes de ruído, à saber:

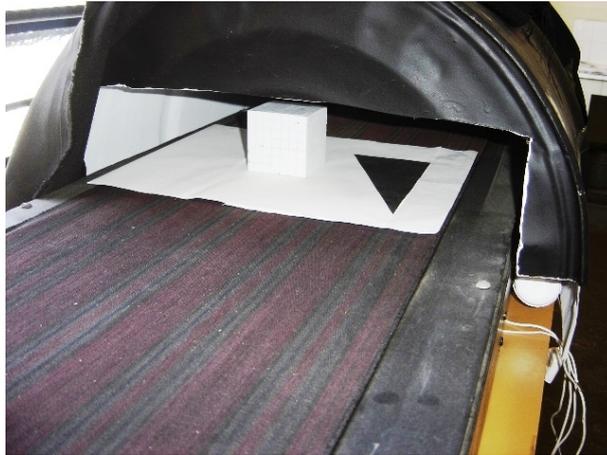


Figura 5.12: Detalhe da esteira, cubo utilizado no ensaio experimental e triângulo de controle



Figura 5.13: Câmara escura utilizada para controlar a iluminação do ensaio

- A velocidade da esteira não é constante. Esta fonte de ruído é minimizada através do triângulo de controle. Entretanto fica impossível assegurar uma distribuição espacial uniforme dos pontos obtidos na reconstrução tridimensional;
- A intensidade da iluminação do ambiente não é constante. Isso pode ocasionar problemas no procedimento de binarização, uma vez que o limiar adequado pode se alterar durante o ensaio;
- Objetos podem não refletir o *laser* ou refletí-lo em direção à esteira, inviabilizando as análises;

- O feixe de *laser* não gera uma superfície exatamente plana.

A figura 5.14 mostra quadros do vídeo obtido pela câmera, onde algumas das fontes de ruído podem ser observadas.

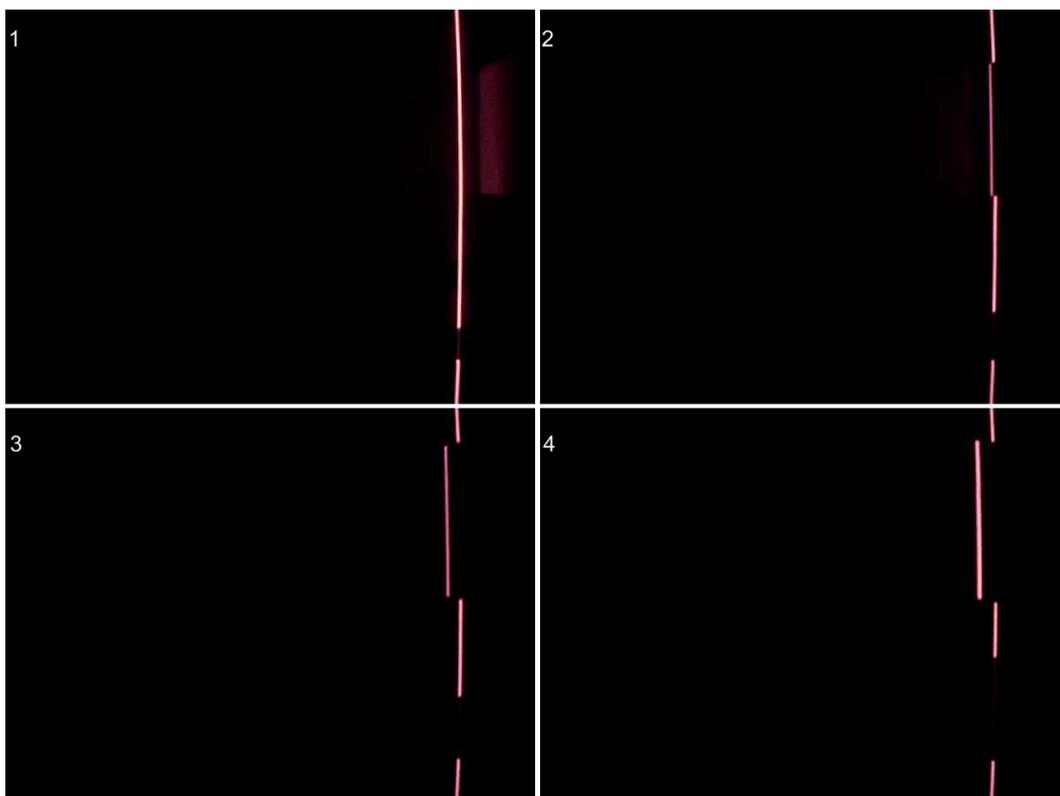


Figura 5.14: Sequência de imagens obtida experimentalmente

5.2.2 Calibração da Câmera

A calibração é feita de forma semelhante à descrita na validação através de imagens artificiais. A figura 5.15 mostra a imagem de controle submetida ao processo de calibração.

5.2.3 Processamento das Imagens

As seguintes etapas do processamento das imagens são aplicadas a cada quadro do vídeo obtido pela câmera, já detalhadas na seção anterior.

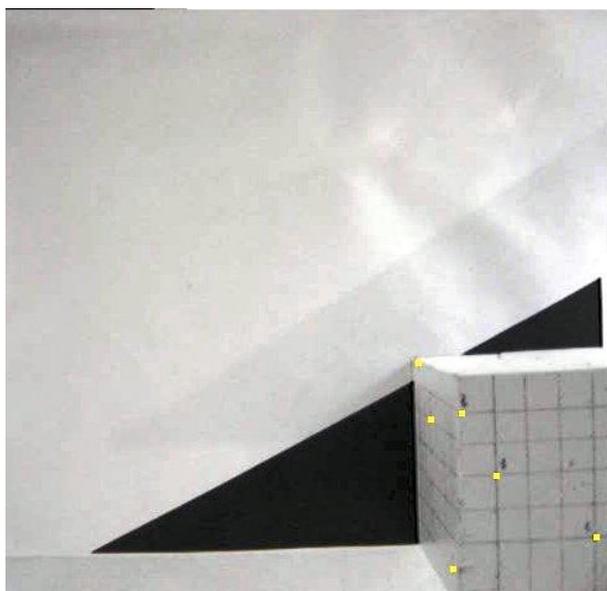


Figura 5.15: Imagem de controle com as coordenadas conhecidas

- Seleção das Regiões de Interesse;
- Segmentação;
- Rotulação;
- Extração de Coordenadas no Referencial do Espaço-imagem;
- Reconstrução Tridimensional;

5.2.4 Apresentação dos Resultados

A figura 5.16 exhibe o resultado experimental da reconstrução de um cubo com 7cm de aresta. Assim como no exemplo anterior, as regiões não identificadas são devidas à sombra.

Os pontos marcados na figura 5.16 são os vértices identificados. Sabendo-se a dimensão da aresta do cubo é de 7cm, calcularemos as dimensões de a , b e c , respectivamente paralelas aos eixos X , Y e Z . Os erros absoluto e relativo são mostrados na tabela 5.2.4

Os erros calculados na tabela 5.3, comparados aos da tabela 5.2, são ligeira-

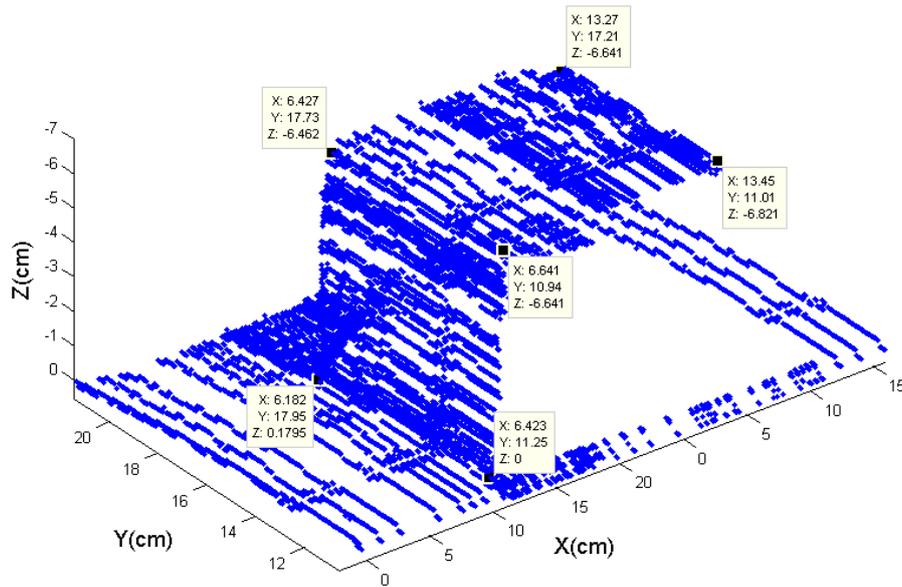


Figura 5.16: Resultado experimental do processamento do cubo

	Exato(cm)	Reconstrução(cm)	Erro absoluto(cm)	Erro relativo(%)
<i>a</i>	7	6,809	0,191	2,73
<i>b</i>	7	6,790	0,21	3,00
<i>c</i>	7	6,641	0,359	5,13

Tabela 5.3: Parâmetros obtidos pela calibração e impostos na geração das imagens

mente maiores. Percebe-se claramente na figura 5.16 a maior presença de ruídos que poderiam ser reduzidos caso alguma técnica de melhoramento de imagem, como uma filtragem, das imagens fosse adotada. Entretanto, visando apenas validar a metodologia proposta, nenhum processo diferente ao adotado para as imagens geradas artificialmente foi utilizado.

Outra fonte de erros é a não consideração de parâmetros da câmera, como os ângulos de rotação entre os eixos dos sistemas referenciais do espaço-imagem e do espaço-objeto e a razão de escala entre as dimensões horizontal e vertical dos *pixels*.

Capítulo 6

Resultados

De forma a ilustrar o funcionamento do método em casos mais gerais, outros exemplos de reconstrução tridimensional usando geração artificial de vídeos via Pov-Ray [®] foram propostos. Foram submetidos ao método de reconstrução um cubo em orientação diferente ao já mostrado no capítulo anterior, um cilindro em duas orientações diferentes, um cilindro vazado em duas orientações diferentes e a metade de um toróide reto.

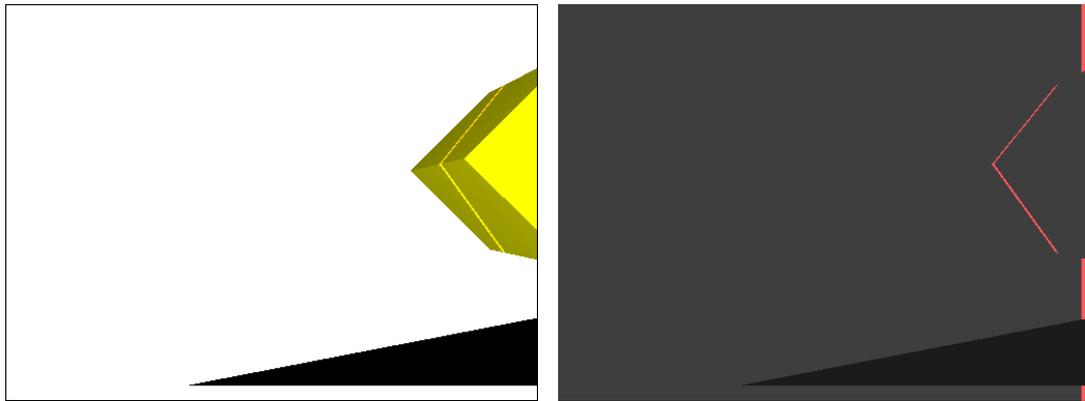
As etapas do processo de reconstrução para cada caso são as mesmas detalhadas anteriormente e uma nova calibração não foi necessária, pois a câmera foi mantida na mesma posição e com a mesma orientação.

A seguir cada exemplo é mostrado juntamente com seus resultados.

6.1 Cubo 2

Este exemplo utiliza um cubo de 7cm de aresta , com uma rotação de 45° em torno de um eixo paralelo ao eixo Z que passa pelo seu centróide. As figuras 6.1(a) e 6.1(b) mostram, respectivamente, a visão da câmera com luz ambiente e a imagem tomada pela câmera sem luz ambiente.

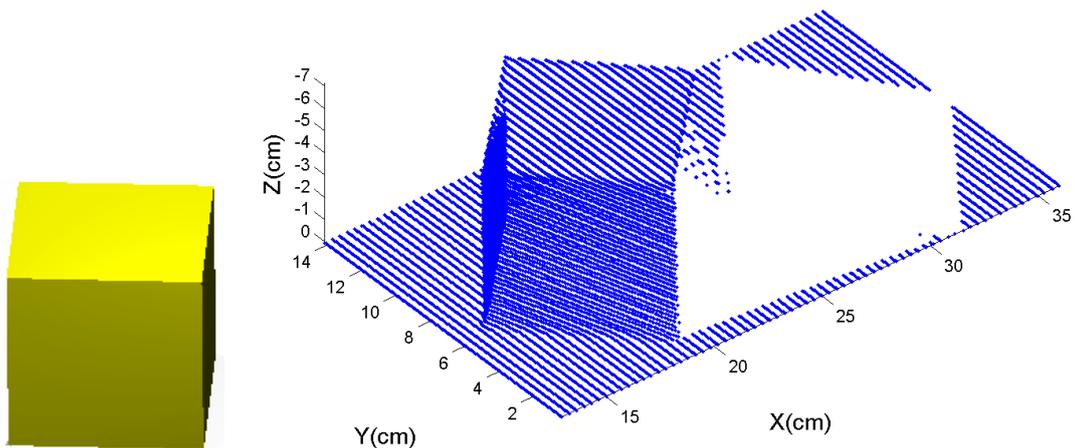
O resultado para o cubo da figura 6.2(a) é mostrado na figura 6.2(b)



(a) Vista superior com luz ambiente

(b) Imagem obtida pela câmera

Figura 6.1: Vista superior e imagem obtida pela câmera



(a) Cubo

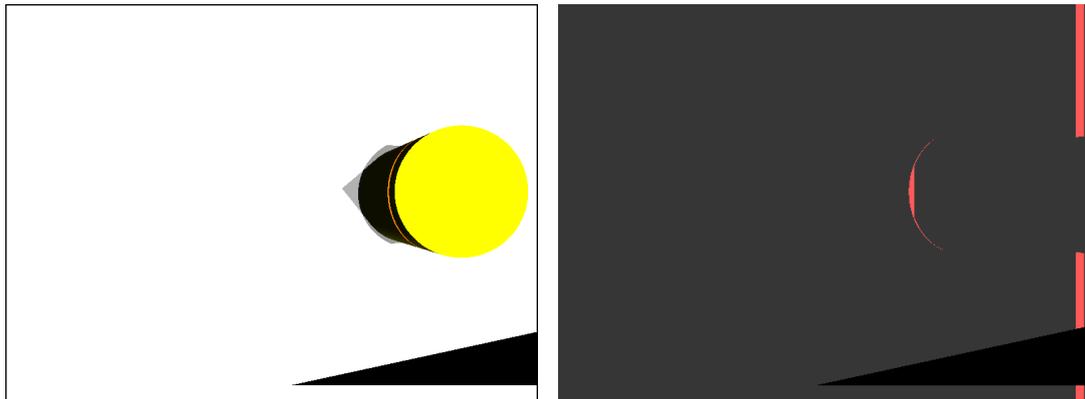
(b) Coordenadas do cubo 2

Figura 6.2: Resultados do processamento do cubo 2

6.2 Cilindro 1

Neste exemplo é reconstruído um cilindro de 5cm de raio e 7cm de altura, com eixo paralelo ao eixo Z . As figuras 6.3(a) e 6.3(b) mostram, respectivamente, a visão da câmera com luz ambiente e a imagem tomada pela câmera sem luz ambiente.

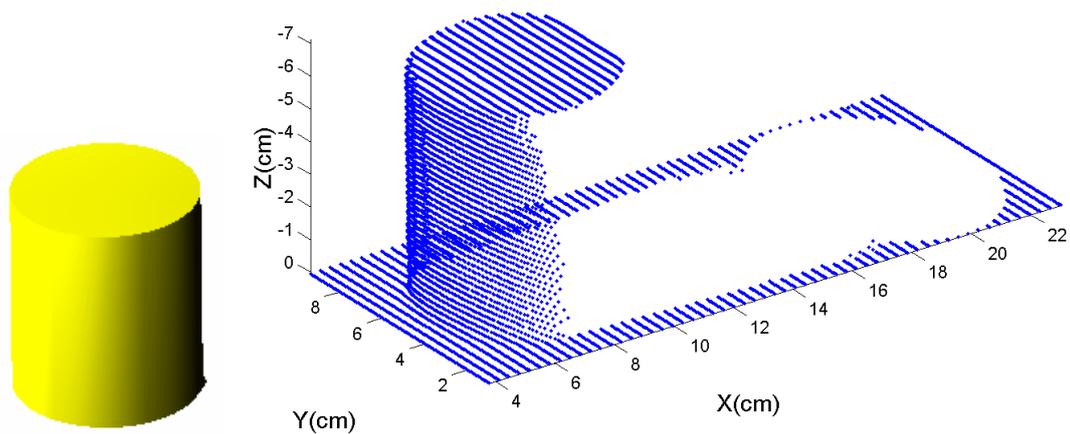
O resultado para o cubo da figura 6.4(a) é mostrado na figura 6.4(b)



(a) Vista superior com luz ambiente

(b) Imagem obtida pela câmera

Figura 6.3: Vista superior e imagem obtida pela câmera



(a) Cilindro

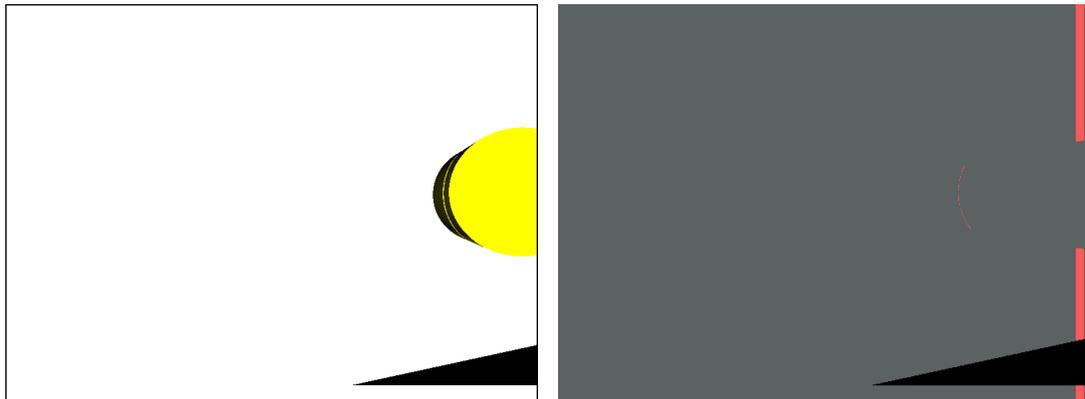
(b) Coordenadas do cilindro 1

Figura 6.4: Resultados do processamento do cilindro 1

6.3 Cilindro 2

Neste exemplo é reconstruído um cilindro de 5cm de raio e 7cm de altura, com eixo que passa pelo seu centróide rotacionado em 15° em relação ao eixo Z . As figuras 6.5(a) e 6.5(b) mostram, respectivamente, a visão da câmera com luz ambiente e a imagem tomada pela câmera sem luz ambiente.

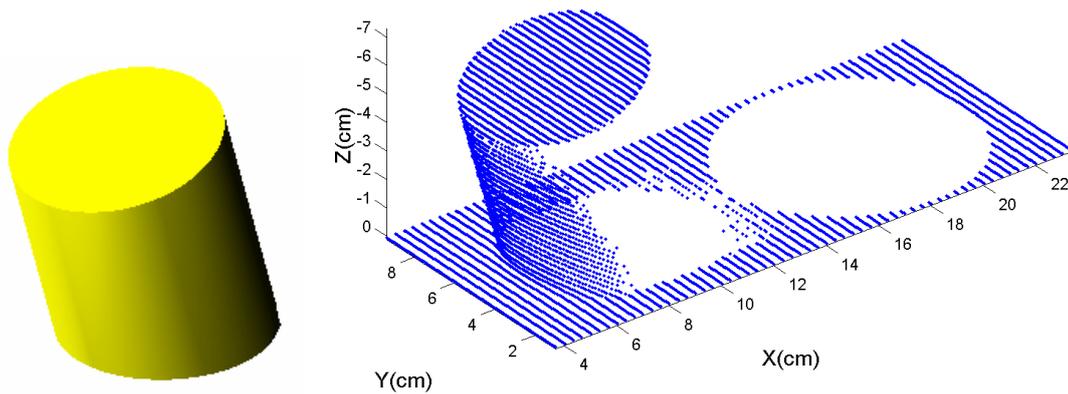
O resultado para o cubo da figura 6.6(a) é mostrado na figura 6.6(b)



(a) Vista superior com luz ambiente

(b) Imagem obtida pela câmera

Figura 6.5: Vista superior e imagem obtida pela câmera



(a) Cilindro

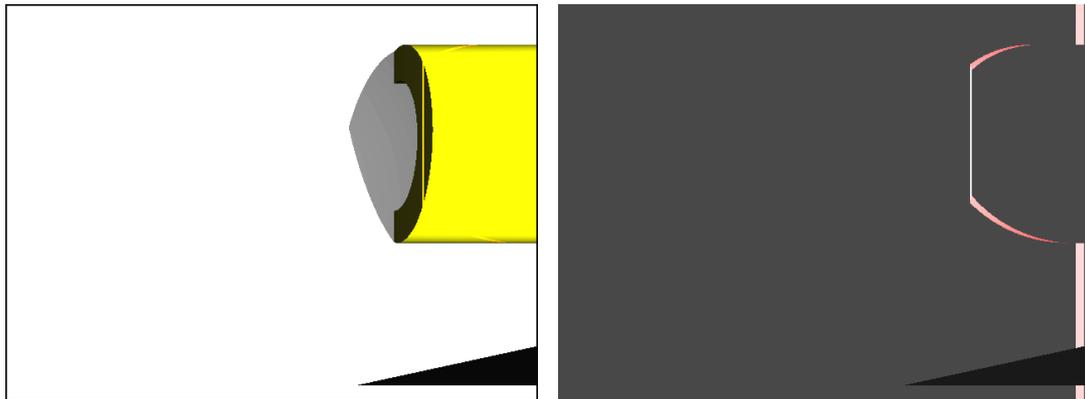
(b) Coordenadas do cilindro 2

Figura 6.6: Resultados do processamento do cilindro 2

6.4 Metade de um Cilindro Vazado 1

A metade de um cilindro vazado de 6cm de raio externo, 4cm de raio interno, 9cm de comprimento e eixo paralelo ao eixo X é reconstruído neste exemplo. As figuras 6.7(a) e 6.7(b) mostram, respectivamente, a visão da câmera com luz ambiente e a imagem tomada pela câmera sem luz ambiente.

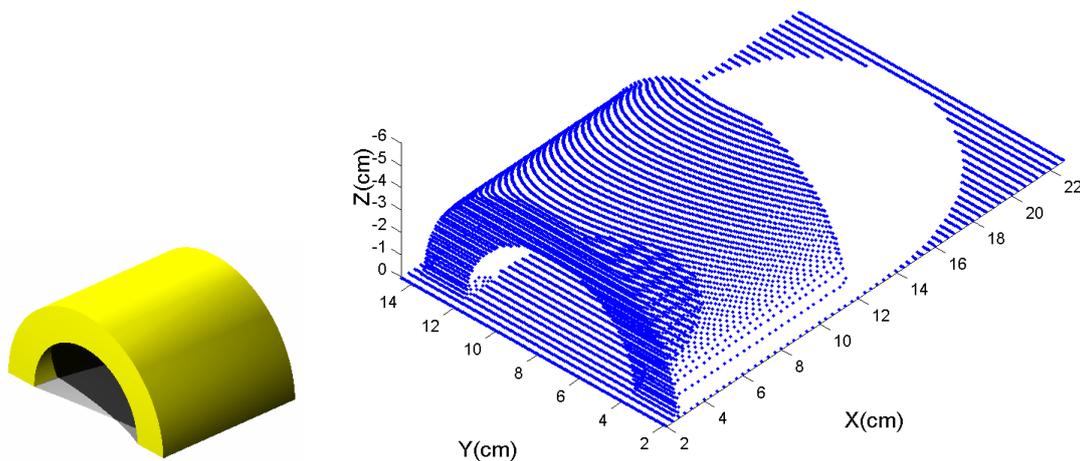
O resultado para o cubo da figura 6.8(a) é mostrado na figura 6.8(b)



(a) Vista superior com luz ambiente

(b) Imagem obtida pela câmera

Figura 6.7: Vista superior e imagem obtida pela câmera



(a) Cilindro Vazado

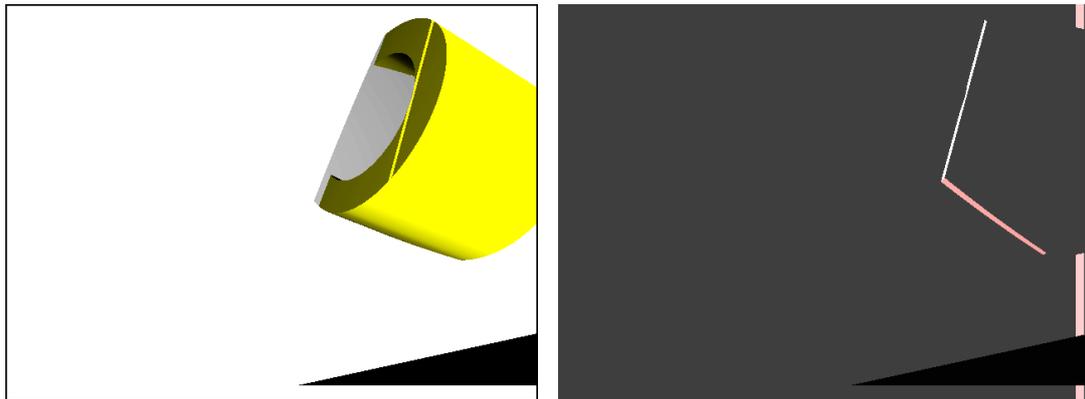
(b) Coordenadas do cilindro vazado 1

Figura 6.8: Resultados do cilindro vazado 1

6.5 Metade de um Cilindro Vazado 2

O cilindro vazado definido no exemplo anterior é analisada novamente, desta vez com rotação de 20° em relação a um eixo paralelo ao eixo Y e 20° em relação a um eixo paralelo ao eixo Z . As figuras 6.9(a) e 6.9(b) mostram, respectivamente, a visão da câmera com luz ambiente e a imagem tomada pela câmera sem luz ambiente.

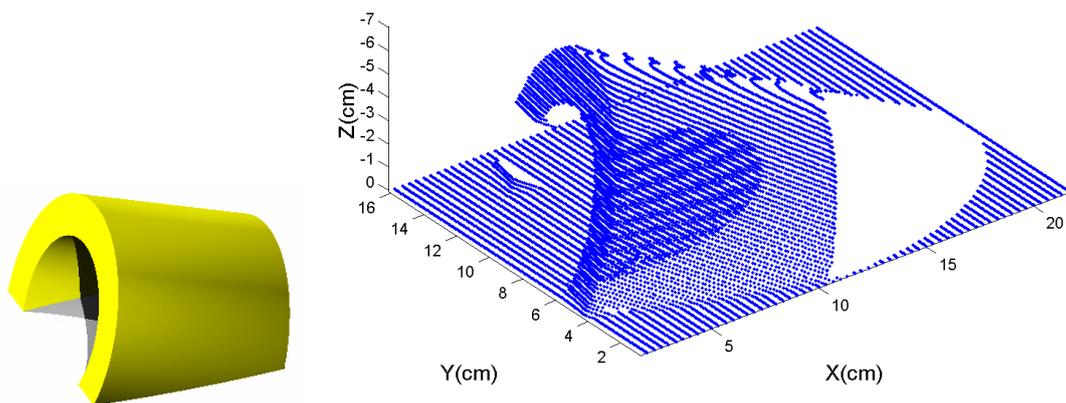
O resultado para o cubo da figura 6.10(a) é mostrado na figura 6.10(b)



(a) Vista superior com luz ambiente

(b) Imagem obtida pela câmera

Figura 6.9: Vista superior e imagem obtida pela câmera



(a) Cilindro Vazado

(b) Coordenadas do cilindro vazado 2

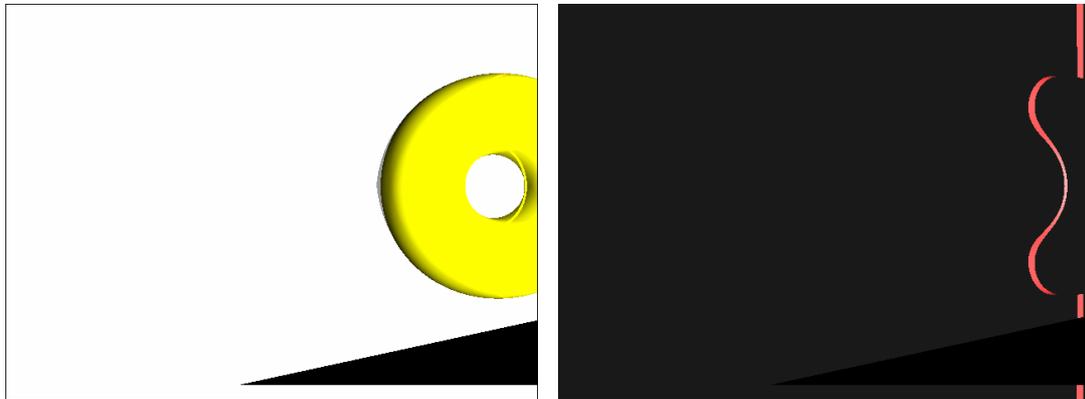
Figura 6.10: Resultados do cilindro vazado 2

6.6 Metade de um Toróide Reto

Neste exemplo é reconstruída a metade de um toróide reto de raio maior de 4,5cm e raio menor de 2,5cm. As figuras 6.11(a) e 6.11(b) mostram, respectivamente, a visão da câmera com luz ambiente e a imagem tomada pela câmera sem luz ambiente.

O resultado para o cubo da figura 6.12(a) é mostrado na figura 6.12(b)

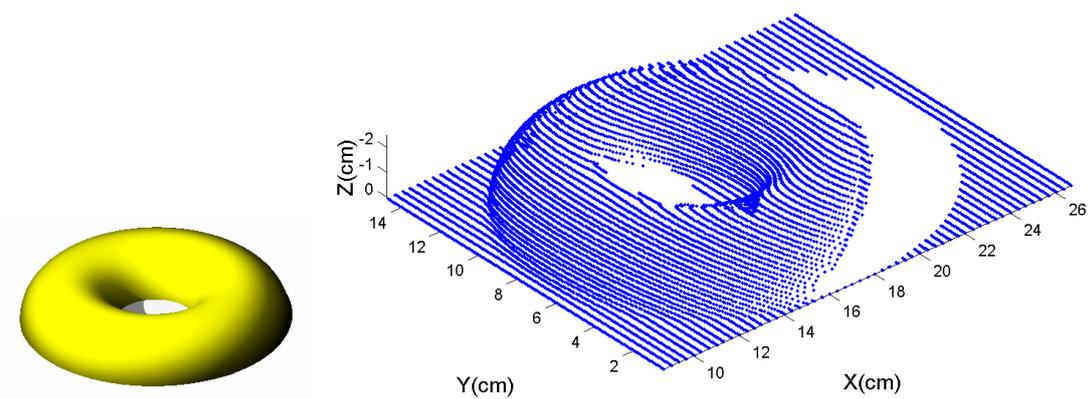
Pode-se observar que o problema da sombra está presente em todos os exemplos, como comentado anteriormente no capítulo 5



(a) Vista superior com luz ambiente

(b) Imagem obtida pela câmera

Figura 6.11: Vista superior e imagem obtida pela câmera



(a) Toróide

(b) Coordenadas do toróide

Figura 6.12: Resultados do toróide

Capítulo 7

Conclusões

Um modelo de reconstrução tridimensional utilizando um feixe de luz retilíneo como luz estruturada foi proposto baseando-se na técnica *Shape From Stereo*. Uma das câmeras desta técnica é substituída pela fonte de luz, e as coordenadas podem ser obtidas somente com a informação da imagem de uma câmera em conjunto com a estrutura formada pela luz incidindo sobre os objetos da cena.

Após a formulação do modelo, uma metodologia para a sua validação foi proposta. Inicialmente foram utilizadas imagens geradas artificialmente com o objetivo de eliminar eventuais fontes de ruídos, como os provocados por iluminação não uniforme e por distorção nas imagens. A validação em condições ideais foi necessária não só para comprovar o funcionamento do modelo, mas também para auxiliar a proposição de um procedimento experimental de forma a validar o modelo também em situações reais. Mesmo não adotando qualquer estratégia diferente da adotada para as imagens obtidas artificialmente, os resultados experimentais foram bastante razoáveis.

A metodologia apresentada identificou de forma satisfatória as formas tridimensionais analisadas e observou-se que o uso de um único feixe de luz retilíneo na reconstrução 3D acarreta problemas de identificação devido às regiões de sombra, gerando falhas na reconstituição. Estes problemas podem ser minimizados conjugando a metodologia apresentada com o uso de diversos lasers.

Como trabalhos futuros, propõe-se

- Incluir o tratamento de ruídos presentes nas imagens, como a aplicação de filtragens e de técnicas de limiarização mais elaboradas;
- Utilizar câmera capaz de obter vídeos com maior resolução, analisando a influência da resolução nos resultados que podem ser obtidos pela metodologia;
- Utilizar diversos feixes de luz incidindo em posições e ângulos diferentes para solucionar o problema das sombras;
- Otimizar o algoritmo para a reconstrução de superfícies em tempo real, onde as etapas de aquisição de imagens, calibração, processamento de imagens e visualização dos resultados seja feita em um único programa;
- Aplicar a metodologia em problemas de tomada de decisão em linhas de montagem para controle de qualidade.
- Propor técnicas de medição do deslocamento diferentes à adotada (triângulo) que possibilitem a monitoração contínua. Padrões que podem ser utilizados são senóides e gradientes, por exemplo.

Referências Bibliográficas

- [1] J. J. Lee and M. Shinozuka. A vision-based system for remote sensing of bridge displacement. *NDT&E International*, 39:425–431, 2006.
- [2] F. M. A. Nogueira, F. S. Barbosa, and L. P. S. Barra. Determinação de frequências naturais de estruturas sob vibrações tridimensionais a partir de uma câmera. In *SIMMEC 2006 - VII Simpósio de Mecânica Computacional*, 2006.
- [3] F. M. A. Nogueira, F. S. Barbosa, and L. P. S. Barra. Determinação de frequências naturais de estruturas sob vibrações tridimensionais a partir de imagens degradadas por borramento oriundas de uma câmera. In *CILAMCE 2006 - XXVII Iberian Latin-American Congress on Computational Methods in Engineering*, 2006.
- [4] T. Figueiro, N. Schuch, A. Soares, L. Guimarães, and A. Susin. Detecção automática de células sanguíneas. LaPSI - DELET - UFRGS, 1996.
- [5] B.K.P. Horn. Understanding image intensities. *Artificial Intelligence*, 8:201–231, 1997.
- [6] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, Inc., Englewood Cliffs, 1982.
- [7] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 16(8):824–831, 1994.
- [8] A. M. G. Tommaselli, M. H. Shimabukuro, P. A. P. Scalco, and F. M. A. Nogueira. Implementation of a photogrammetric range system. *International Archives of Photogrammetry and Remote Sensing*, 31(B2):368–373, 1996.

- [9] U. R. Dhond and J. K. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.
- [10] A. K. Dalmia and M. Trivedi. High-speed extraction of 3d structure of selectable quality using a translating camera. *Computer Vision and Image Understanding*, 64(1):97–110, July 1996.
- [11] M. A. G. Izquierdo, M. T. Sanchez, A. Ibañez, and L. G. Ullate. Sub-pixel measurement of 3d surfaces by laser scanning. *Sensors and Actuators*, 76:1–8, 1999.
- [12] G. Wang, Z. Hu, F. Wu, and H. Tsui. Implementtion and experimental study on fast object modeling based on muliple structured stripes. *Optics and Lasers in Engineering*, 42:627–638, 2004.
- [13] J. C. Russ. *The Image Processing Handbook*. IEEE Press, Raleigh, North Carolina, second edition, 1994.
- [14] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, first edition, 1992.
- [15] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, New Jersey, second edition, 2002.
- [16] D. L. B. R. Jurjo. *Desenvolvimento de um sistema de visão computacional para medição de deslocamentos em estruturas*. PhD thesis, COPPE/UFRJ, 2007.
- [17] F. M. A. Nogueira. Geração automática de mapas de disparidade em visão estéreo. Universidade Estadual de Campinas, 1998. Dissertação de Mestrado.
- [18] M. Galo. Calibração e aplicação de câmaras digitais. Universidade Federal do Paraná, 1993. Dissertação de Mestrado.
- [19] A.M.G. Tommaselli and C. L. Tozzi. Extração de linhas retas em imagens digitais para aplicações fotogramétricas. In *Anais do XVI Congresso Brasileiro de Cartografia*, 1993.

- [20] C. Gemael. *Introdução ao ajustamento de observações: aplicações geodésicas*. Editora UFPR, Curitiba, 1994.
- [21] R. J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Proceedings, 22nd International Symposium, Society of Photo-optical Instrumentation Engineers*, 1978.
- [22] M. Galo and C. L. Tozzi. Surface reconstruction using multiple light sources and perspective projection. In *Proceedings, IEEE - International Conference on Image Processing*, 1996.
- [23] J. K. Hasegawa. *Shape from shading com projeção perspectiva e calibração de câmara*. PhD thesis, FEEC/UNICAMP, 1997.
- [24] J. K. Aggarwal, L. S. Davis, and W. N. Martin. Correspondence processes in dynamic scene analysis. *Proceedings of IEEE*, 69(5):562–571, May 1981.
- [25] R.J. Schalkoff. *Digital Image Processing and Computer Vision*. John Wiley & Sons, New York, 1989.
- [26] J. Bigun, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 13:775–790, 1991.
- [27] A. Y. K. Ho and T. C. Pong. Cooperative fusion of stereo and motion. *Pattern Recognition*, 21(1):121–130, 1996.
- [28] A. D. Bimbo, P. Nesi, and J. L. C. . Optical flow computation using extended constraints. *EEE Transactions on Image Processing*, 5(5):206–218, 1996.
- [29] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 19(3):206–218, 1997.
- [30] R. Bajcsy and L. Lieberman. Texture gradient as a depth cue. *Computer Graphics and Image Processing*, 5(1):52–67, 1976.

- [31] N. Shrikhande and G. Stockman. Surface orientation from a projected grid. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 11(6):650–655, 1989.
- [32] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 12(2):148–163, February 1990.
- [33] J. Ens and P. Lawrence. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 15(2):97–108, 1993.
- [34] S. K. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 18(12):1186–1197, 1996.
- [35] F. M. A. Nogueira. *Reconstrução Tridimensional por Visão Estéreo e Detecção de Oclusões*. Tese de Doutorado, FEEC/UNICAMP, 2007.
- [36] Pov-Ray v3.6. Disponível para download em <http://www.povray.org/>. Site acessado em 25/07/2007, 2007.
- [37] VirtualDub 1.7.5. Disponível para download em <http://www.virtualdub.org/>. Site acessado em 15/06/2007, 2007.
- [38] Matlab 7.1. *The MathWorks Inc.*, Copyright 1984-2007.