



Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

*Reconstrução 3D em
Ambientes com Luzes Estruturadas*

Aluno: Eduardo Lourenço Apolinário (ela@cin.ufpe.br)

Orientadora: Judith Kelner (jk@cin.ufpe.br)

Co-Orientadora: Veronica Teichrieb (vt@cin.ufpe.br)

Recife, 4 de Outubro de 2006.

Resumo

A área de Reconstrução 3D tem recebido muita atenção nos últimos anos, seja por ser uma área com muitas aplicações potenciais, seja por apresentar ainda muitos problemas em aberto a serem resolvidos. Esse trabalho de graduação se propõe a esclarecer conceitos a respeito da área de Reconstrução 3D, apresentando as principais técnicas, além de problemas associados. Uma introdução à geometria projetiva e como esta se insere no contexto da reconstrução 3D também é apresentada. Com isso, o objetivo desse trabalho é apresentar e descrever a técnica ativa de reconstrução 3D proposta por Jean-Yves Bouguet, e sua implementação, ressaltando os resultados obtidos bem como as dificuldades encontradas durante o desenvolvimento.

Palavras-chave: Reconstrução 3D, Visão Computacional, Técnica Ativa de Bouguet

Agradecimentos

“Sooner or later you’re going to realize that there’s a
difference between knowing the path and walking the path”

Morpheus

É com uma felicidade imensa que finalizo mais essa etapa na minha vida. Sem dúvida, o caminho foi difícil, mas compensado pelo fato de não ter sido percorrido sozinho. Estive sempre bem acompanhado e fiz amizades que certamente durarão o resto da minha vida.

Primeiramente tenho que agradecer de coração aos meus pais Walter e Verônica, que sempre indicaram o melhor caminho, e mesmo quando minha decisão não era a mais acertada, me apoiaram. Agradeço também a meus irmãos Thiago e Renata que sempre torceram por mim. Obrigado por tudo que vivemos e que ainda vamos viver.

Agradeço também à Cinthia por toda a compreensão e companheirismo dedicados a mim durante todos esses anos. Amor, obrigado.

Aos meus amigos de todos os lugares, e de todas as épocas, todos os que vejo todos os dias, tanto os que por alguma infelicidade do destino, hoje não estou mais em contato. E que não vou enumerar para esses agradecimentos não se tornarem maiores que o próprio trabalho. Acreditem, não existem palavras para agradecer o que vocês fizeram por mim, só posso dizer um sonoro obrigado por acreditarem em mim.

À todo mundo do GPRT, Joma, Mouse, Luiz, Gabriel, Márcio, Pedro, enfim, a todos que trabalham aqui no GPRT/GRVM. Agradecimentos especiais a alguns artefatos que possibilitaram a finalização do presente trabalho: a cartola dos sinônimos, essencial para escrever o documento, o chapéu da inteligência, utilizado (e repartido) na implementação do protótipo e finalmente, Johnnes da Silva, que com suas poses nada convencionais, alegrou o dia-a-dia sempre inventivo do ambiente do GPRT.

À Verônica, ou melhor Vt, por aguentar todas as minhas dúvidas e incertezas e sempre estar pronta pra me acalmar e apontar a direção a seguir. Vt, obrigado do fundo do coração.

À profa. Judith por todo o apoio e crédito para a realização desse trabalho.

Obrigado.

Aos grandes mestres que tive durante a graduação, Ruy de Queiroz, André Santos, Sílvio Melo, Sílvio Meira, e todos os outros que de algum modo contribuíram para minha formação, meus sinceros agradecimentos.

Por fim, tenho que agradecer por tudo que vivi nesses quatro anos e meio de lutas, doces vitórias, amargas derrotas, grandes descobertas. Enfim, tudo que me fez querer sempre ser melhor do que eu era ontem e menos do que serei amanhã. Galera, obrigado.

Índice

1	INTRODUÇÃO.....	8
1.1	Objetivos.....	9
1.2	Estrutura do Documento.....	9
2	GEOMETRIA PROJETIVA.....	10
2.1	Ferramenta Ubíqua.....	10
2.2	Coordenadas Homogêneas.....	11
2.3	Homogeneidade.....	11
2.4	Geometrias Afim e Euclidiana.....	12
2.4.1	Geometria Afim.....	12
2.4.2	Geometria Euclidiana.....	13
2.5	Projeções de Câmera.....	15
2.6	Câmeras Calibradas.....	17
3	RECONSTRUÇÃO 3D.....	18
3.1	Modelos de Câmera.....	20
3.1.1	O Modelo de Projeção em Perspectiva.....	21
3.1.2	A Matriz de Calibração de Câmera.....	22
3.1.3	A Matriz de Parâmetros Extrínsecos.....	23
3.1.4	Estimativa da Matriz de Calibração da Câmera.....	23
3.1.5	Distorção de Lentes.....	24
3.2	Calibragem de Câmera.....	25
3.3	Técnicas de Reconstrução 3D.....	26
3.3.1	Técnicas Passivas.....	26
3.3.2	Técnicas Ativas.....	27
4	A TÉCNICA ATIVA DE BOUGUET.....	28
4.1	Formalismo Dual-B.....	28
4.1.1	Contexto.....	28
4.1.2	Espaços Duais.....	30

4.2	Descrição da Técnica de Bouguet	31
4.2.1	Localização da Fonte de Luz	33
4.2.2	Processamento Espacial e Temporal.....	34
4.2.3	Estimativa do Plano de Sombra e Triangulação	35
5	IMPLEMENTAÇÃO E RESULTADOS	36
5.1	Implementação.....	36
5.1.1	Calibração da Câmera.....	37
5.1.2	Calibração da Fonte de Luz	37
5.1.3	Cálculo da Profundidade	38
5.1.3.1	Estimativa do Plano de Sombra ou Processamento Espacial	39
5.1.3.2	Estimativa do Tempo de Sombra ou Processamento Temporal	39
5.1.3.3	Triangulação.....	39
5.1.4	Visualização da Nuvem de Pontos	40
5.2	Resultados.....	40
6	CONCLUSÃO E TRABALHOS FUTUROS.....	44
6.1	Trabalhos Futuros.....	44
7	REFERÊNCIAS.....	46

Índice de Figuras

FIGURA 1: LINHA DO HORIZONTE	13
FIGURA 2: PROJEÇÕES DE CÂMERA: (A) FORMAÇÃO DA IMAGEM; (B) TRANSFORMAÇÃO PROJETIVA LINEAR; (C) MAPEAMENTO DE PLANOS; (D) MOVIMENTAÇÃO DO CENTRO DA CÂMERA; (E) PONTOS COPLANARES.....	16
FIGURA 3: HOMEM DESENHANDO UM ALAÚDE (THE DRAUGHTSMAN OF THE LUTE), GRAVURA EM MADEIRA DE 1525, POR ALBRECHT DÜRER.....	20
FIGURA 4: O MODELO DE PROJEÇÃO EM PERSPECTIVA.....	21
FIGURA 5: <i>GRID</i> DE CALIBRAÇÃO USADA PARA ESTIMAR A MATRIZ DE CALIBRAÇÃO	23
FIGURA 6: DESFAZENDO A DISTORÇÃO NUMA IMAGEM. (A) DEVIDO À DISTORÇÃO, PROJEÇÕES DE LINHAS RETAS ESTÃO CURVADAS. (B) AS PROJEÇÕES DE LINHAS RETAS AGORA ESTÃO RETAS	24
FIGURA 7: QUADRO DE REFERÊNCIA DA CÂMERA.....	28
FIGURA 8: UTILIZAÇÃO DA TÉCNICA DE BOUGUET.....	31
FIGURA 9: PRINCÍPIO GEOMÉTRICO APLICADO NA TÉCNICA DE BOUGUET.....	32
FIGURA 10: <i>PIXELS</i> DA BASE DO LÁPIS E PONTA DA SOMBRA	37
FIGURA 11: PRINCÍPIO GEOMÉTRICO DA VARIAÇÃO DA TÉCNICA DE BOUGUET COM APENAS UM PLANO E FONTE DE LUZ CALIBRADA.....	38
FIGURA 12: LINHAS SUPERIOR E INFERIOR FORNECIDAS PELO USUÁRIO	38
FIGURA 13: INTERFACE DA FERRAMENTA BLENDER COM <i>SCRIPT</i> PYTHON EM DESTAQUE.....	40
FIGURA 14: CENA COM OBJETO CLARO E PEQUENO, RECONSTRUÍDA POR BOUGUET.....	41
FIGURA 15: (A) CENA MODELADA. (B) RESULTADO DE BOUGUET. (C) RESULTADO DA IMPLEMENTAÇÃO REALIZADA NESTE TRABALHO, (D) RESULTADO ENCONTRADO DE OUTRO PONTO DE VISTA	42

1 Introdução

Nos últimos anos, a Computação Gráfica obteve progressos tremendos na visualização de modelos 3D. Muitas técnicas atingiram um grau de maturidade bastante alto e estão sendo portadas para hardware especializado. A velocidade no aumento da capacidade de visualização 3D aumenta a uma taxa maior que a lei de Moore. Anos atrás, o que era conseguido com computadores de milhares de dólares, hoje, consegue-se poucas centenas de dólares, isso tudo com a possibilidade de visualização em tempo real.

Essa evolução vem suprir uma crescente demanda por modelos mais complexos. O problema é que mesmo com as ferramentas de modelagem 3D existentes no mercado, sintetizar modelos realistas é difícil e demorado, portanto, caro. Muitos objetos virtuais são inspirados em objetos reais, logo, faz bastante sentido ser possível capturar modelos a partir de objetos reais.

Pesquisadores vem investigando métodos de captura 3D de objetos por anos. No passado, a ênfase era em inspeção visual e definição de rotas para robôs. Hoje, isso está mudando, há mais e mais demanda por conteúdo 3D para computação gráfica, realidade virtual e comunicação. Isso resulta numa mudança na ênfase dos requisitos. A qualidade visual da informação se torna um dos pontos principais de atenção da aplicação. Assim, não somente as posições de uma malha com poucos pontos tem que ser calculadas com uma precisão grande, e sim a geometria e aparência de todos os pontos de uma superfície.

As condições de aquisição nesses novos domínios de aplicação podem não coincidir com os requisitos dos sistemas existentes. O que geralmente acontece é que esses sistemas necessitam de processos de calibração intrincados, o que implica em um requisito desejável ser a flexibilidade nos processos de aquisição, calibração. Além disso, os sistemas existentes, geralmente, são construídos utilizando um hardware bastante especializado (e.g. lasers), resultando em um alto custo final desses sistemas. Muitas aplicações, no entanto, precisam de sistemas robustos e de baixo custo. Isso estimula o uso de câmeras tanto de vídeo quanto de fotografias. Além disso, os progressos recentes em fotografia e vídeo digitais facilitam ainda mais esse fato.

Por conta da convergência desses diversos fatores, muitas técnicas foram desenvolvidas nos últimos anos. Muitas delas não necessitam de equipamento especializado para a captura dos modelos tridimensionais de objetos reais. Existem técnicas ativas e passivas. Basicamente, técnicas passivas controlam de alguma maneira a iluminação da cena (e.g. projeção de algum padrão de luz na cena), simplificando assim, o problema de reconstrução, ainda que restringindo a aplicabilidade, uma vez que nem sempre é possível alterar parâmetros de iluminação da cena. Já as técnicas passivas são mais flexíveis, mas computacionalmente mais caras e dependentes da estrutura da cena [38]. Existem muitas tentativas de atacar

o problema de reconstrução do ponto de vista de técnicas passivas, uma diferença fundamental entre as técnicas consiste nos níveis de calibração necessários e a interação, ou seja, o nível de automação necessário. Por muitos anos, fotogrametria [19] tem tratados dos problemas de extração de medidas precisas de imagens. Tais técnicas, em sua maioria, possuem um alto grau de calibração e quase não há automação. O processo de aquisição detalhada de modelos é bastante demorado.

1.1 Objetivos

Um dos objetivos propostos por esse trabalho é realizar a implementação de um sistema de reconstrução 3D utilizando a técnica proposta por Jean-Yves Bouguet em sua tese de doutorado [1]. Por implementação, entende-se codificação, análises de erro, testes com diversos conjuntos de cenas, enfim, realizar um estudo de caso da técnica proposta pelo autor, visando o entendimento das bases da área de reconstrução 3D para possíveis investigações futuras.

1.2 Estrutura do Documento

Esse trabalho, como dito anteriormente, pretende descrever a implementação de uma técnica de reconstrução 3D proposta por Jean-Yves Bouguet em [1]. Para justificar algumas das escolhas feitas para a implementação e melhor entendimento do contexto atual da área de reconstrução 3D, no capítulo 2, conceitos principais de geometria projetiva são mostrados, já no capítulo 3, as bases conceituais que fundamentam a área de reconstrução 3D são apresentadas. No capítulo 4, a técnica proposta por Bouguet é apresentada e uma discussão a cerca das escolhas feitas pelo autor é mostrada. Os dois últimos capítulos discorrem sobre a implementação e resultados alcançados pela implementação realizada, e por fim, uma conclusão sobre os progressos realizados nesse trabalho e trabalhos futuros é apresentada.

2 Geometria Projetiva

Neste capítulo serão apresentados os principais conceitos sobre geometria projetiva, além de aplicações na área de reconstrução 3D, com o objetivo de fundamentar este trabalho de conclusão de curso. A estratificação da geometria em projetiva, afim e euclidiana é apresentada. Outro ponto importante deste capítulo é a justificativa da escolha da geometria projetiva como sendo a base para tratar problemas de Visão Computacional. Essa justificativa é realizada mediante uma descrição do problema da calibração de câmera a partir de um modelo de geometria projetiva. Para um detalhamento mais sólido dos conceitos aqui discutidos, recomenda-se [2], [3] e [5].

2.1 Ferramenta Ubíqua

A maioria das pessoas está acostumada a utilizar transformações projetivas, mesmo desconhecendo seus conceitos. Quando se olha para uma figura, por exemplo, enxergam-se quadrados que não necessariamente são quadrados, ou até elipses que não são elipses. O tipo de transformação que mapeia esses objetos planares numa figura é a transformação projetiva.

Uma pergunta importante a ser feita é: quais as propriedades geométricas preservadas pelas transformações projetivas? Com certeza, o formato não é uma delas, já que uma elipse pode parecer um círculo. Tamanhos também não são preservados, já que os eixos de uma elipse são afetados de maneira diferente pelas transformações projetivas. Não são preservados ângulos, distâncias, proporção entre distâncias pelas transformações projetivas. No entanto, uma propriedade fundamental dos mapeamentos característicos de transformações projetivas é a retidão, ou seja, retas continuam sendo retas, após a aplicação da transformação [2].

Um problema simples que não é trivialmente resolvível utilizando apenas geometria euclidiana é a interseção de retas (restringindo-se ao caso bidimensional (2D)). Duas retas quase sempre se encontram num ponto, mas existem infinitos pares de retas que não se intersectam em ponto algum, chamadas retas paralelas. Um artifício lingüístico é dizer que elas se encontram “no infinito”. No entanto, o infinito não existe, é uma convenção matemática. Então, para resolver esse problema pode-se encarar esses pontos onde idealmente retas paralelas se encontram como pontos ideais. Uma vez que esses pontos são adicionados ao espaço euclidiano, este é transformado num espaço projetivo. Essa é uma simplificação bastante útil, uma vez que as propriedades do espaço Euclidiano, (com seus conceitos de distância, ângulos, tamanhos, retas, incidência, etc), são bastante familiares. Em resumo, o espaço projetivo é apenas uma extensão do espaço euclidiano onde retas paralelas se encontram num ponto no infinito chamado de ponto ideal.

2.2 Coordenadas Homogêneas

Um ponto em um espaço 2D pode ser representado por um par de números reais (x, y) . Pode-se adicionar outra coordenada a esse par, obtendo $(x, y, 1)$, que conceitualmente é o mesmo ponto. Particularmente no exemplo citado, é fácil convencer-se de que voltar à representação anterior é simples, bastando suprimir a última coordenada. Contudo, no caso de triplas do tipo $(x, y, 3)$ não se pode adotar a mesma abordagem. A definição da tripla genérica (kx, ky, k) representa o mesmo ponto (x, y) , independente do valor de k , logo, a tripla $(2x, 2y, 2)$ deveria representar o mesmo ponto que (x, y) .

Pontos são representados por classes de equivalência, chamadas de coordenadas homogêneas do ponto. Dada uma tripla (kx, ky, k) podem-se recuperar as coordenadas originais simplesmente dividindo por k as coordenadas da tripla, restando (x, y) . No entanto, a abordagem de classes de equivalência não funciona com triplas do tipo $(x, y, 0)$. Para esse tipo, deve-se dividir as primeiras coordenadas pela última, o que resultaria em $(x/0, y/0)$, e como não se define divisão por zero, diz-se que tais pontos estão no infinito. Essa é a origem dos pontos no infinito, que são pontos com coordenadas homogêneas onde a última é 0.

Resumindo, mostrou-se como estender o espaço euclidiano para um espaço projetivo pela simples adição de coordenadas homogêneas (operação possível em qualquer dimensão). Utilizando uma notação formal, o espaço euclidiano \mathbb{R}^n pode ser estendido para um espaço projetivo \mathbb{P}^n cujos pontos são vetores homogêneos, caracterizando a operação injetiva. Para detalhamento dos conceitos relativos a coordenadas homogêneas apresentados nessa seção, recomenda-se [7].

2.3 Homogeneidade

Em geometria clássica, por exemplo a euclidiana, não há distinção entre pontos, sendo toda a extensão do espaço considerada homogênea. A escolha do ponto de origem é arbitrária pois se pode realizar facilmente uma mudança de base escolhendo um novo sistema de coordenadas rotacionado e/ou transladado. O resultado dessa operação é uma transformação euclidiana. Outro tipo de transformação, mais geral, onde um operador linear é aplicado ao \mathbb{R}^n seguido de uma translação da origem do espaço, é conhecida como transformação afim.

Tanto em transformações afins quanto em euclidianas, pontos no infinito permanecem no infinito. Por analogia às transformações euclidiana e afim, pode-se definir uma transformação projetiva. Uma transformação linear no \mathbb{R}^n pode ser representada por uma multiplicação de matrizes aplicada às

coordenadas do ponto. Do mesmo modo, uma transformação projetiva do espaço \mathbb{P}^n é um mapeamento de coordenadas homogêneas representativas de um ponto (um vetor de $n + 1$ coordenadas), no qual o vetor de coordenadas é multiplicado por uma matriz não-singular. Sob tal mapeamento, pontos que estão no infinito (com coordenada final zerada) são mapeados em pontos arbitrários, ou seja, diferentemente de transformações afins e euclidianas, em mapeamentos projetivos, pontos no infinito não são preservados.

Em problemas de Visão Computacional, o espaço projetivo é usado como uma maneira conveniente de representar o mundo real 3D, estendendo-o para o espaço projetivo 3D. De maneira similar, imagens, geralmente formadas projetando-se o mundo em um espaço 2D, são estendidas por conveniência para que se pense que estas estejam imersas em um espaço projetivo 2D. Na realidade, o mundo real, e as imagens dele, não contêm pontos no infinito, e algumas vezes faz-se necessário distinguir quais são os pontos, as retas e os planos no infinito. Por essa razão, apesar de geralmente trabalhar-se com espaços projetivos, sabe-se a priori quais são as retas e planos no infinito. Isso vai de encontro ao espírito da geometria projetiva pura, apesar de ser bastante útil no tratamento prático dos problemas. Geralmente tratam-se todos os pontos do espaço como iguais, e somente quando se faz necessário, põe-se em evidência as retas e os planos no infinito

2.4 Geometrias Afim e Euclidiana

Na seção anterior foi visto que um espaço projetivo de dimensão n pode ser obtido adicionando um hiperplano no infinito (no caso 2D, uma reta, já no 3D, um plano). Nesta seção, o processo inverso será apresentado. Esse processo será apresentado devido a sua utilidade no processo de volta de um espaço Projetivo para um espaço Euclidiano. Por questões de simplicidade, apenas os casos 2D e 3D serão apresentados.

2.4.1 Geometria Afim

Partindo do pressuposto que, inicialmente, o espaço projetivo é homogêneo, não há o conceito de paralelismo entre retas, já que retas paralelas são aquelas que se encontram num ponto no infinito. Portanto, no espaço projetivo não há o conceito de pontos no infinito, ou seja, todos os pontos são iguais. Por isso, diz-se que em geometria projetiva não existe o conceito de paralelismo.

Para tal conceito fazer sentido, uma reta tem que ser concebida como a reta no infinito. Isso resulta no fato de que, apesar de todos os pontos serem criados igualmente, alguns são mais “iguais” que outros. Assim, tomando como exemplo uma folha de papel, e imaginando que esta se estende ao infinito (formando um espaço projetivo de dimensão 2), o “pedaço” de espaço que se vê parece, localmente, um plano euclidiano. Em seguida, desenha-se uma linha reta e diz-se que essa é a reta do infinito. Após isso, desenha-

se duas retas que se intersectam exatamente em um ponto nessa reta. Uma vez que elas se intersectam “no infinito”, define-se que elas são paralelas. Essa situação lembra um plano infinito. Para ilustrar, pode-se pensar numa foto tirada numa região plana do planeta Terra. Os pontos no infinito do plano da imagem aparecem como a reta no horizonte (veja Figura 1).

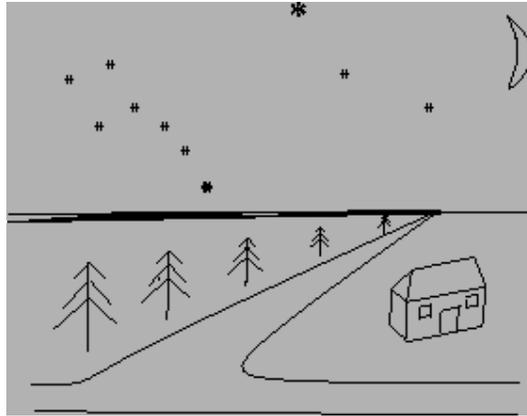


Figura 1: Linha do horizonte

Retas, como as linhas da estrada, se encontram na linha do horizonte. Pontos no infinito correspondem à linha do horizonte da imagem, e retas paralelas no mundo correspondem a retas se intersectando no horizonte. Do ponto de vista apresentado até aqui, o plano de mundo e sua imagem são apenas meios alternativos para visualizar a geometria de um plano projetivo. A geometria do espaço projetivo em conjunto com uma reta é conhecida como geometria afim e qualquer transformação projetiva que mapeia essa reta em um espaço para essa reta distinta de outro espaço é conhecida como transformação afim. Como visto, com essa reta “no infinito” pode-se definir paralelismo de retas no plano.

Em resumo, o conceito de paralelismo é ganho usando uma reta do espaço projetivo como a “reta no infinito” e, com ele, a geometria afim. Em outras palavras, a geometria afim é vista como sendo uma especialização da geometria projetiva, na qual uma reta é escolhida e apontada como sendo a “reta no infinito”.

2.4.2 Geometria Euclidiana

Como dito anteriormente, com uma reta apontada como sendo uma reta especial do espaço projetivo é possível operar com transformações ditas transformações afins. Além disso, é possível encontrar uma característica especial dessa reta, que faz com que seja possível formar a geometria euclidiana.

Em geometria afim, um círculo é um conceito que não existe, uma vez que transformações que

preservam a reta no infinito transformam círculos em elipses e vice-versa. Ou seja, em geometria afim, não há distinção entre elipses e círculos.

Em geometria euclidiana, no entanto, essas figuras são distintas, e possuem uma diferença importante. Algebricamente, uma elipse é uma equação de segundo grau, pertencente ao grupo de figuras geométricas conhecido por cônicas. Assim, é esperado, e verdadeiro, que duas elipses, num caso geral, se intersectem em quatro pontos distintos. No entanto, dois círculos não podem se intersectar em mais de dois pontos. Analiticamente, as interseções de duas curvas de grau dois estão sendo calculadas, logo, era de se esperar quatro soluções, no caso geral. A questão que surge é a seguinte: o que os círculos possuem de especial para, no caso geral, só produzirem duas soluções?

Na verdade, as outras duas soluções existem, apenas estão no plano complexo, ou seja, são pontos complexos. Além do mais, é fácil calculá-los, como mostrado a seguir.

A equação do círculo em coordenadas homogêneas (x, y, w) , apresentada na Equação 2.1, é da forma:

$$(x - aw)^2 + (y - bw)^2 = r^2 w^2 \quad 2.1$$

Neste caso o círculo é representado centrado no ponto $(x_0, y_0, w_0)^T = (a, b, 1)^T$. Além disso, é fácil verificar que os pontos $(x, y, w)^T = (1, \pm i, 0)^T$ pertencem a todos os círculos, o que implica em que para cada par de círculos, esses pontos pertencem ao conjunto de soluções. E, apesar de serem complexos, obedecem ao par de equações reais: $x^2 + y^2 = 0$ e $w = 0$. Estes são chamados os pontos circulares do plano.

Finalmente, a geometria euclidiana surge da geometria projetiva escolhendo-se uma reta no infinito e dois pontos pertencentes a essa reta, chamados pontos circulares.

Partindo desse ponto de vista, no plano, uma cônica pode ser definida por cinco pontos. Um círculo, por sua vez, pode ser definido por apenas três pontos, sendo visto como uma cônica especial, cujos outros dois pontos são os pontos circulares.

Uma vez que se define uma reta como sendo a reta no infinito, e os pontos circulares, junto com as cônicas, retas, e o conceito de paralelismo e distância, compõe-se a geometria euclidiana 2D. No caso 3D, basta adicionar o conceito de planos e definir, em vez de uma reta, um plano no infinito.

2.5 Projeções de Câmera

Numa projeção em perspectiva, diminui-se em um a dimensionalidade do espaço, indo-se de um espaço de 3 dimensões para um de 2 dimensões. A maneira mais comum de definir esse processo é a projeção central, onde se escolhe um ponto para ser o centro de projeção e , a partir daí, para cada ponto no espaço traça-se um raio que passa pelo centro de projeção. Esse raio intersecta um plano, chamado plano de imagem. A interseção do raio com o plano de imagem representa a imagem do ponto.

Esse modelo está de acordo com um modelo simples de câmera, onde um raio de luz de um ponto no mundo passa por uma lente e atinge o filme ou o CCD¹. Ignorando fatores tais como foco e distorção da lente, pode-se admitir que todos os raios passam por um único ponto, o centro de projeção.

Na projeção central, pontos em \mathbb{P}^3 são mapeados em pontos em \mathbb{P}^2 , logo, todos os pontos pelos quais um raio partindo do centro de projeção passa são projetados no mesmo ponto na imagem. Para os propósitos de projeção de imagem, é possível considerar todos os pontos ao longo desse raio como sendo iguais. Além disso, o conjunto de pontos na imagem é igual ao conjunto de raios passando pelo centro de projeção.

Se o raio que passa por $(0,0,0,1)^T$ e $(x,y,z,t)^T$ for representado pelas coordenadas $(x,y,z)^T$, é fácil ver que para qualquer constante k diferente de zero, o raio $k(x,y,z)^T$ representa o mesmo raio. Assim, os raios são representados por coordenadas homogêneas. De fato, eles formam um espaço 2D de raios. O conjunto de raios pode ser visto como sendo a representação do espaço de imagem \mathbb{P}^2 . Nessa representação, tudo que importa é o centro da câmera, uma vez que esse ponto determina qual conjunto de raios formam a imagem. Assim, duas imagens capturadas do mesmo ponto no espaço são projetivamente equivalentes. Somente quando se começa a medir pontos numa imagem é que se faz necessário escolher um quadro particular de coordenadas, especificando uma matriz de câmera. De modo resumido, com exceção do campo de visão (FOV – *Field Of View*), todas as imagens capturadas a partir de um mesmo centro de câmera são equivalentes.

¹ CCD (*Charge-Coupled Device*) ou Dispositivo de Carga Acoplado é um sensor para a captura de imagens, formado por um circuito integrado contendo um *array* de capacitores ligados (acoplados). Sob o controle de um circuito externo, cada capacitor pode transferir sua carga elétrica para um outro capacitor vizinho. Os CCDs são usados em fotografia digital e astronomia (particularmente em fotometria, óptica e espectroscopia UV e técnicas de alta velocidade).

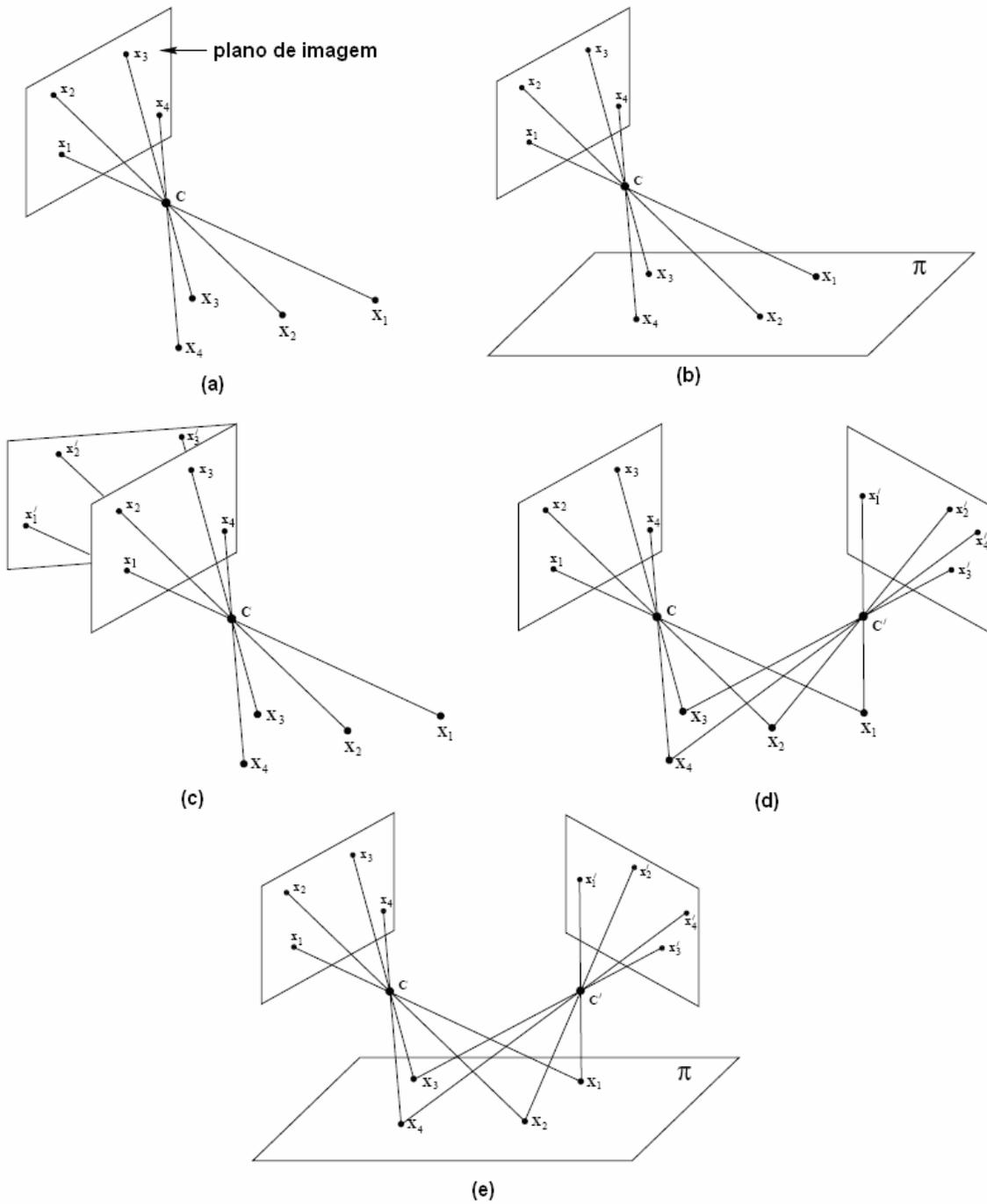


Figura 2: Projeções de câmera: (a) formação da imagem; (b) transformação projetiva linear; (c) mapeamento de planos; (d) movimentação do centro da câmera; (e) pontos coplanares

Essas características estão descritas na Figura 2. Na Figura 2.a os pontos de imagem x_i são as interseções do plano com os raios dos pontos no espaço X_i pelo centro da câmera C . Se os pontos no

espaço são coplanares, existe uma transformação projetiva linear entre os planos no mundo e da imagem (Figura 2.b). Todas as imagens da mesma câmera se relacionam por uma transformação projetiva linear. Comparando a Figura 2.b e a Figura 2.c, em ambos os casos, planos são mapeados um no outro por raios passando pelo centro de projeção. Em (b), o mapeamento é entre o plano em coordenadas de mundo e o plano da imagem, já em (c), entre dois planos da imagem. Se o centro da câmera se move (Figura 2.d), em geral, as imagens não estão relacionadas por uma transformação projetiva, a menos que os pontos sejam coplanares (Figura 2.e).

2.6 Câmeras Calibradas

Para entender completamente a relação entre imagem e mundo, é necessário expressar uma relação euclidiana entre ambos. A geometria euclidiana 3D é determinada especificando-se um plano em \mathbb{P}^3 como sendo o plano no infinito, e uma cônica no plano como sendo a cônica absoluta. Para uma câmera não localizada no plano no infinito, tal plano é mapeado univocamente no plano da imagem. Isso acontece porque qualquer ponto na imagem define um raio no espaço que intersecta o plano no infinito em apenas um ponto. Assim, o plano no infinito no mundo não diz nada a respeito da imagem. A cônica absoluta, por outro lado, por ser uma cônica no plano da imagem, deve ser mapeada numa cônica no plano da imagem. A imagem resultante é chamada de Imagem da Cônica Absoluta (ICA). Se a localização da ICA é conhecida na imagem, diz-se que a câmera está calibrada.

3 Reconstrução 3D

Por mais de duas décadas, o problema de reconstruir a estrutura 3D de uma cena a partir de um conjunto de imagens 2D foi objeto de estudo da comunidade de Visão Computacional. Navegação e manipulação de objetos estão entre as atividades que são melhor realizadas quando se tem conhecimento do ambiente 3D da cena.

O processo de extrair uma representação 3D a partir de um conjunto de imagens é uma das funções mais importantes do sistema visual humano. Acredita-se que essa tarefa é realizada associando-se vários itens visuais extraídos das imagens, tais como coloração, foco, deformação relativa a perspectiva, contornos de oclusão, brilhos, etc. [28]. Para uma discussão mais apurada a respeito desses itens, recomenda-se [32]. Em muitas circunstâncias, o sistema humano é capaz de, com apenas uma imagem, extrair uma boa representação 3D da cena observada.

Apesar dos mais de quarenta anos de pesquisa para tentar entender o sistema visual humano (da retina ao córtex visual), muito pouco é conhecido atualmente. Do ponto de vista biológico, a visão é provavelmente o sentido mais complexo, tanto pela sua arquitetura quanto funcionamento. Para mais detalhes, recomenda-se [27],[28] e [29].

Do ponto de vista da engenharia, o objetivo principal da comunidade de Visão Computacional é desenvolver sistemas artificiais que repliquem as funcionalidades do sistema visual humano. Em relação a esse problema, o sistema ideal seria capaz de extrair informações de geometria, de forma automática, apenas com o conjunto de imagens 2D fornecidas a ele.

As aplicações de tais dispositivos de modelagem 3D são inúmeras. As comunidades de Computação Gráfica e Multimídia vêm mostrando um interesse cada vez maior na construção e desenvolvimento de tais sistemas. Além delas, pode-se citar como áreas de aplicação o *design* industrial, visitas visuais a museus e catálogos comerciais *on-line*, entre outras.

Uma série de técnicas foram proposta para serem utilizada por essas aplicações. Entre elas, as técnicas baseadas em imagem consistem em gerar novos pontos de vista a partir de um conjunto de imagens, mas sem calcular explicitamente a geometria da cena. Tal tipo de técnica tem obtido bastante sucesso aplicada a cenas complexas onde a geometria é difícil de reconstruir. Um dos problemas potenciais dessa técnica é a grande quantidade de imagens requerida. Conseqüentemente, a quantidade de dados necessários a serem armazenados na memória faz com que esse tipo de técnica seja proibitivo para computadores de médio porte, impedindo, assim, sua utilização em aplicações de internet, por exemplo.

Por outro lado, técnicas baseadas em reconstrução explícita do modelo da cena permitem a geração de novos pontos de vista utilizando o mesmo modelo 3D, ou seja, bastando que ocorra uma nova renderização do modelo. Em comparação com técnicas baseadas em imagem, as técnicas baseadas em modelo podem ser vistas como uma maneira inteligente de comprimir o fluxo de imagem, ou o conjunto de imagens necessárias para o processo de reconstrução, uma vez que a informação contida no conjunto de imagens pode ser resumida como um único modelo 3D, renderizado de maneira eficiente com as técnicas existentes. Para uma apresentação de técnicas de síntese de imagem, recomenda-se [9].

Infelizmente, o sistema ideal ainda está muito longe de ser obtido. Apesar das tentativas, apresentadas em [32], para geração de imagens a partir de câmeras descalibradas, e dos resultados apresentados serem em geral bastante bons, existem muitos entraves para utilização de tais técnicas no problema de reconstrução 3D genérico. Desse modo, hoje, uma maneira prática de resolução desse problema são as técnicas ativas, descritas na Seção 3.3. A maioria desses sistemas consistem na combinação de um ou mais sistemas de imagem passivos (uma ou mais câmeras) e de dispositivos ativos (*lasers*, projetores LCD (*Liquid Cristal Display*)) calibrados um a um, i.e., cada câmera é calibrada com relação a outra.

O princípio de funcionamento desses sistemas é bastante intuitivo: o projetor emite um padrão de luz sobre a cena, que é detectado nas imagens capturadas pela câmera. A estrutura 3D da cena é calculada via triangulação geométrica (ou triangulação óptica). Deixando de lado o fato de que técnicas desse tipo são bastante insensíveis com relação à mudança de textura, elas têm a vantagem de alcançar resultados satisfatórios (erros no processo de reconstrução podem se reduzir a um em dois mil). O principal entrave à adoção desse tipo de técnica é o custo, como mencionado em [33], [34] e [35].

Um desafio interessante para os pesquisadores em Visão Computacional é construir sistemas nos quais a captura de imagens em condições normais de iluminação seja a condição suficiente para o processo de reconstrução 3D de uma cena. Essa classe de técnica que utiliza um dispositivo de projeção externo é chamado de técnicas passivas (maiores detalhes na Seção 3.3). Entre todos os itens passivos que contêm informação 3D (disparidade estereoscópica, textura, *motion parallax*, (des)foco, sombras, especularidades, contornos de oclusão e outras discontinuidades de superfície), hoje, somente disparidade estereoscópica fornece resultados razoáveis. Infelizmente, essa técnica tem problemas fundamentais: necessita de duas câmeras, aumentando, assim, o custo, e não trata de superfícies sem textura, apesar dos avanços alcançados segundo [37].

Uma extensão da técnica de estéreo com duas câmeras é a substituição do par de câmeras por apenas uma câmera. Nesse caso, a câmera captura duas imagens de pontos de vista diferentes, em diferentes instantes de tempo. A partir desse ponto, o processo de reconstrução é idêntico às técnicas com duas

câmeras. Uma das vantagens dessa extensão é o custo, já que se opera com apenas uma câmera. A outra vantagem é a ergonomia, já que a câmera pode ser movida livremente pela cena, sem preocupação com a calibração; para mais detalhes dessa técnica passiva com câmeras descalibradas veja [31]. O nome que se dá a esse tipo de técnica é Estrutura a Partir do Movimento (ou em inglês, *Structure from Motion*). No entanto, existem basicamente duas limitações nessa técnica. Primeiramente, uma vez que a câmera pode ser movimentada livremente, a disparidade de movimento deve ser sempre calculada. A segunda limitação ocorre pelo fato de que, já que as imagens são capturadas em instantes diferentes de tempo, é necessário que o mundo permaneça parado no intervalo. Apesar de existirem indícios em [2] e [36] de que essa hipótese pode ser relaxada, esse ainda é um problema em aberto.

3.1 Modelos de Câmera

Na literatura, encontram-se várias representações para os diversos modelos de câmera existentes, tais como *fisheye*, omnidirecional (ou catadióptrica), além de mais outros modelos em [36]. No entanto, neste documento, apenas o modelo de câmera estenopeica (ou *pinhole*[31]), acrescido de deformações radiais, será abordado. Nesta seção, esse modelo será descrito utilizando-se o formalismo de geometria projetiva e coordenadas homogêneas apresentado no capítulo anterior.

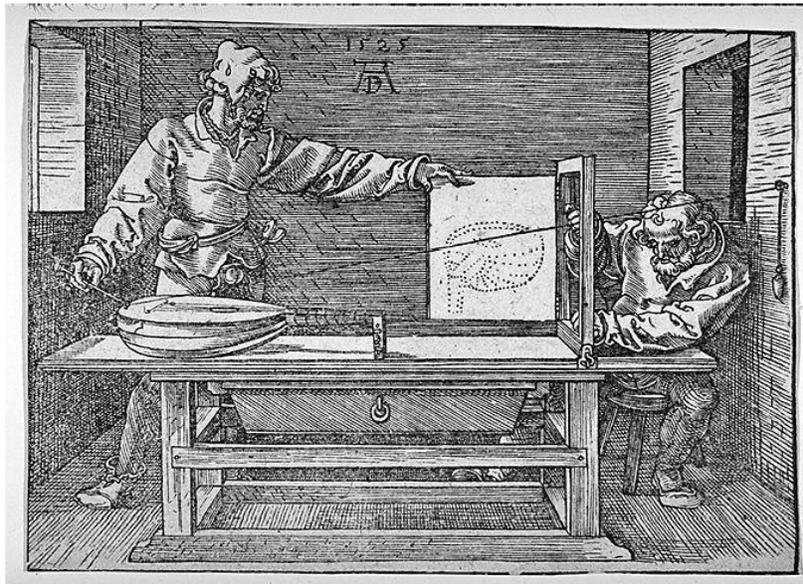


Figura 3: Homem desenhando um alaúde (The Draughtsman of the Lute), gravura em madeira de 1525, por Albrecht Dürer

O processo geométrico de formação de uma imagem foi ilustrado por Dürer (veja Figura 3), onde o processo de desenho da cena fica completamente determinado uma vez que se escolha qual o plano de

projeção e qual o centro da figura.

3.1.1 O Modelo de Projeção em Perspectiva

Bem antes dos pesquisadores da área de Visão Computacional atacarem o problema de estimativa de poses e parâmetros de câmera, fotogrametristas já tinham pesquisado o mesmo extensivamente. Para detalhes adicionais sobre modelos de câmera e um estudo mais aprofundado das propriedades geométricas e numéricas da calibração de câmera, veja [14] e [15].

Matematicamente, a formação da imagem pode ser definida como a projeção do espaço 3D sobre o plano da imagem, como mostrado na Figura 4. $(O, \vec{i}, \vec{j}, \vec{k})$ é o sistema de coordenadas mundial, $(C, \vec{i}_c, \vec{j}_c, \vec{k}_c)$ é o sistema de coordenadas da câmera, M é um ponto 3D e m é sua projeção no plano da imagem.

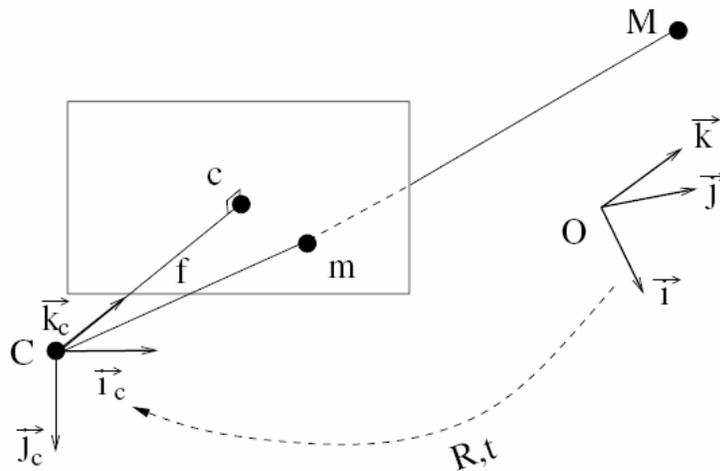


Figura 4: O modelo de projeção em perspectiva.

As coordenadas do ponto 3D $M = [X, Y, Z]^T$, expressas num sistema de coordenadas euclidiano, e seu ponto 2D correspondente $m = [x, y]^T$ na imagem são relacionados pela Equação 3.1,

$$s\tilde{m} = P\tilde{M} \quad 3.1$$

onde s é um fator de escala, $\tilde{m} = [x, y, 1]^T$ e $\tilde{M} = [X, Y, Z, 1]^T$ são as coordenadas homogêneas (para mais detalhes a cerca deste assunto, consulte a Seção 2.2) dos pontos m e M , e P é uma matriz de projeção 3×4 . P é definida, exceto por um fator de escala, dependendo de 11 parâmetros. É, de fato, aceito na

comunidade que essa matriz descreve realisticamente uma câmera real, mantendo a simplicidade do ponto de vista matemático. Tal matriz de projeção pode ser decomposta como apresentado na Equação 3.2:

$$P = K[R|t] \quad 3.2$$

Onde K é a matriz de calibração dependente dos parâmetros intrínsecos da câmera, tais como distância focal, e $[R|t]$ é a matriz de parâmetros extrínsecos, e corresponde a transformação euclidiana de coordenadas mundiais para o sistema de coordenadas da câmera; R representa uma matriz de rotação e t uma translação.

Abaixo segue uma descrição mais detalhada sobre ambas matrizes, de calibração e de parâmetros extrínsecos.

3.1.2 A Matriz de Calibração de Câmera

A matriz de calibração de câmera K contém os parâmetros intrínsecos da câmera, também chamados de parâmetros internos. A matriz K apresenta a forma expressa na Equação 3.3:

$$K = \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad 3.3$$

Onde α_u e α_v são, respectivamente, os fatores de escala nas direções u , e v . Esses fatores são proporcionais a distância focal f da câmera: $\alpha_u = k_u f$ e $\alpha_v = k_v f$, onde k_u e k_v são os números de *pixels* por unidade de distância nas direções u e v . $c = [u_0, v_0]^T$ representa as coordenadas, no sistema da imagem, da interseção entre o eixo principal e o plano da imagem, a esse ponto dá-se o nome de ponto principal. s é diferente de 0 somente se as direções u e v não forem perpendiculares, o que é bastante incomum nas câmeras modernas.

Devido a fatores tais como distorção de lentes e tamanho das lentes, escolher o ponto principal c para ser o centro da imagem é apenas uma aproximação; para uma revisão completa dos problemas associados a essa escolha, veja [16]. De modo similar, se os *pixels* forem quadrados, α_x e α_y podem ser iguais. A partir disso, é dito que uma câmera está calibrada quando seus parâmetros intrínsecos forem conhecidos.

3.1.3 A Matriz de Parâmetros Extrínsecos

A matriz $[R|t]$ define a orientação e a posição da câmera. É formada por R , a matriz de rotação, e t , o vetor de translação, também conhecido como pose da câmera. Várias aplicações [14] assumem que a matriz de calibração K é conhecida, logo, os esforços são concentrados em estimar R e t , ou, equivalentemente, estimar as posições dos objetos na cena e suas orientações com relação ao sistema de coordenadas da câmera.

Mais formalmente, a matriz corresponde a transformação euclidiana do sistema de coordenadas mundial para o sistema de coordenadas da câmera: um ponto 3D representado pelo vetor M_w no sistema de coordenadas mundial será representado pelo vetor $M_c = RM_w + t$ no sistema de coordenadas da câmera. Dessa relação, pode-se recuperar o centro óptico C no sistema de coordenadas mundial. Este deve satisfazer $O = RC + t$, o que implica em $C = -R^{-1}t = -R^T t$.

3.1.4 Estimativa da Matriz de Calibração da Câmera

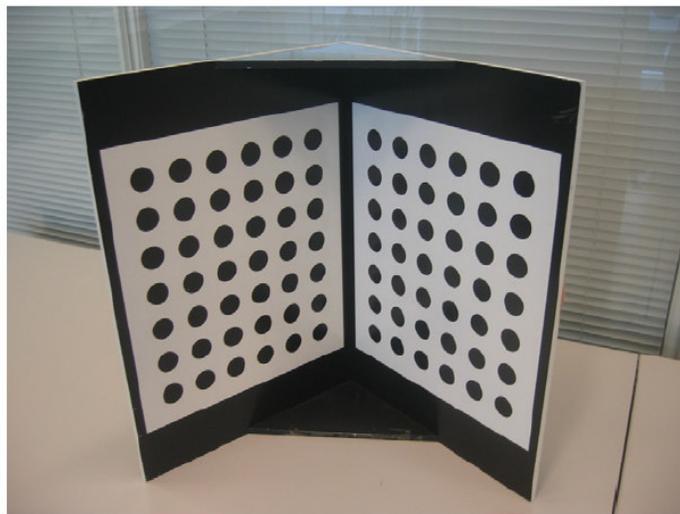


Figura 5: *Grid* de calibração usada para estimar a matriz de calibração

Em muitos sistemas de reconstrução 3D [21], assume-se que os parâmetros intrínsecos da câmera são fixos e conhecidos, o que na prática significa que a câmera não oferecerá a função de *zoom*, por ser difícil distinguir uma mudança na distância focal de uma translação ao longo do eixo principal da câmera. Esses parâmetros podem ser estimados numa etapa de calibração *offline*, a partir das próprias imagens. Os métodos clássicos de calibração de câmera fazem uso de um padrão de calibração de tamanho conhecido.

Algumas vezes uma grade 3D com algum padrão regular pintado, tal como os círculos pretos da Figura 5, é utilizado para estimar a matriz de calibração.

Nesse exemplo, as coordenadas 3D dos cantos dos quadrados brancos com respeito ao canto superior esquerdo são conhecidas com exatidão. É relativamente fácil encontrar tais cantos na imagem e as correspondências entre os pontos 3D e os pontos 2D da imagem, a fim de calcular os parâmetros da matriz. Nos últimos anos, [18] e [12] introduziram técnicas similares de calibração baseadas em uma grade planar simples vista de várias posições. Tais técnicas são bastante flexíveis, visto que os padrões podem ser impressos e colados em algum objeto planar. Implementações podem ser encontradas em [12] e [3].

3.1.5 Distorção de Lentes

O modelo de projeção em perspectiva nem sempre é suficiente para representar todos os aspectos da formação da imagem, uma vez que ele não leva em consideração as possíveis distorções ocasionadas pela lente da câmera. Existem várias situações de poses de câmera onde as distorções não podem ser negligenciadas (por exemplo, em câmeras de ângulo aberto²).

Felizmente, as distorções de câmera podem ser modeladas através de deformações 2D da imagem. Como mostrado na Figura 6, dado uma estimativa dos parâmetros de distorção, o efeito de distorção causado pela perspectiva pode ser removido da imagem de uma maneira eficiente, utilizando uma look-up table. Na seqüência, a projeção em perspectiva pode ser aplicada.

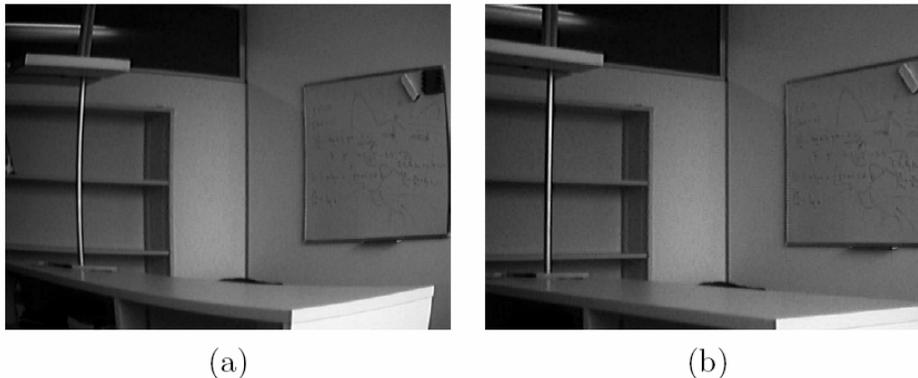


Figura 6: Desfazendo a distorção numa imagem. (a) Devido à distorção, projeções de linhas retas estão curvadas. (b) As projeções de linhas retas agora estão retas

² *Wide angle cameras*

Uma representação comum dos parâmetros de distorção é dada a seguir. Sejam $\hat{u} = [\tilde{u}, \tilde{v}]^T$ as coordenadas do *pixel* na imagem distorcida, e $\hat{x} = [\tilde{x}, \tilde{y}]^T$ as coordenadas normalizadas correspondentes tais que $\tilde{u} = u_0 + \alpha_u \tilde{x}$ e $\tilde{v} = v_0 + \alpha_v \tilde{y}$, onde u_0 , v_0 , α_u e α_v são os parâmetros intrínsecos da câmera, descritos na Equação 3.3. Sejam $\bar{u} = (u, v)$ e $\bar{x} = (x, y)$ os valores correspondentes após a distorção ser desfeita. A distorção pode ser expressa como uma soma de dois componentes tais que $\hat{x} = x + dx_{radial} + dx_{tangencial}$. Segundo [18], a distorção radial pode ser aproximada pela seguinte série de potências, apresentada na Equação 3.4:

$$dx_{radial} = (1 + k_1 r^2 + k_2 r^4 + \dots)x \quad 3.4$$

Onde $r = \sqrt{x^2 + y^2}$. A distorção tangencial $dx_{tangencial}$ tem menos influência e pode ser expressa pela Equação 3.5:

$$dx_{tangencial} = \begin{bmatrix} 2p_1 xy + p_2(r^2 + 2x^2) \\ p_1(r^2 + 2y^2) + 2p_2 xy \end{bmatrix} \quad 3.5$$

Mas, em muitos casos, a parcela relativa a distorção tangencial pode ser ignorada.

A biblioteca OpenCV [3] implementa o método derivado em [17] e [12]. Esse é um método bastante conveniente para sistemas *desktop*. Em áreas como fotogrametria, outros métodos são utilizados, como o Método das câmeras métricas [19] e câmeras estereométricas [20]. Sem distorção, a imagem de uma linha reta será uma linha reta e, reciprocamente, os parâmetros de distorção podem ser estimados a partir de linhas retas medindo desvios de retidão.

3.2 Calibragem de Câmera

Basicamente, o processo de calibração de câmera consiste na recuperação da distância principal f e do ponto principal $(x_0, y_0)^T$, ou equivalentemente, a recuperação da posição do centro de projeção $(x_0, y_0, f)^T$ no sistema de coordenadas da imagem. Em fotogrametria [21], dá-se o nome de orientação interior a esse processo.

Um alvo de calibração pode ser colocado na cena para ajudar na correspondência entre pontos na imagem e no espaço. Contudo, é impraticável posicionar o alvo de calibração utilizando meios puramente mecânicos. Como resultado disso, é necessário recuperar a relação entre os sistemas de coordenadas da imagem e da cena. Em fotogrametria, dá-se o nome de orientação exterior a esse processo [20].

Uma vez que as lentes das câmeras contêm distorções geométricas que não podem ser desprezadas no processo de calibração de câmera se faz necessário incluir a recuperação dos coeficientes da série de potências dessas distorções.

É importante observar que no processo de calibração de câmera o objetivo principal é recuperar as transformações, baseando-se em medidas de coordenadas, onde geralmente se usam transformações conhecidas para estabelecer a relação entre um sistema de coordenadas e outro.

O método implementado na biblioteca OpenCV [3], basicamente recupera as orientações interior e exterior, os coeficientes da série de potências relativa a distorção das lentes e um fator de escala que melhor se adapta às coordenadas da imagem correspondentes aos pontos com coordenadas 3D conhecidas. Isso é realizado em estágios, começando com uma fórmula fechada de mínimos quadrados para estimativa dos parâmetros [36], seguida de uma otimização iterativa não-linear de todos os parâmetros usando as estimativas como valores iniciais.

3.3 Técnicas de Reconstrução 3D

Nesta seção, serão apresentadas algumas técnicas que representam o estado-da-arte da pesquisa em reconstrução 3D. Primeiramente, a classe de técnicas passivas será apresentada e, posteriormente, as técnicas ativas serão descritas. O presente trabalho se baseia nessa última categoria, sendo relacionado com a técnica ativa de reconstrução proposta em [1], chamada Técnica de Bouguet. Além disso, diversas aplicações e resultados obtidos por ambas as classes de técnicas serão mostrados brevemente.

3.3.1 *Técnicas Passivas*

Essa classe de técnicas se baseia na captura de imagens da cena sob condições naturais de iluminação. Na maior parte das técnicas, um conjunto de pontos salientes é observado e suas localizações são calculadas usando triangulação. Esse conjunto de pontos constitui o modelo 3D reconstruído, e para efeitos de visualização, conecta-se essa nuvem de pontos formando uma malha. Para conseguir estimar a forma do objeto, também é necessário calcular a trajetória percorrida pela câmera no espaço. Para maiores detalhes sobre esse assunto recomenda-se [32].

Basicamente, técnicas passivas oferecem três vantagens. A primeira vantagem refere-se ao fato de que apenas uma câmera é necessária para a captura das imagens. Além disso, como o movimento da câmera é livre, existe uma probabilidade maior de uma reconstrução ter um nível de erro menor. Finalmente, a terceira vantagem advém da segunda, pois como o movimento é livre, não há a necessidade de *hardware* especial (como um braço de robô). Essa característica traz uma vantagem ergonômica para essa classe de

técnicas, uma vez que o conjunto final de equipamentos para captura de imagens é bastante pequeno, bastando apenas uma câmera (sensor CMOS³ e lentes) e um computador.

3.3.2 Técnicas Ativas

O grande problema na utilização de técnicas passivas é que elas assumem que os objetos da cena a ser reconstruída possuem textura. Apesar de em [37] existirem indícios que apontam para a possibilidade de utilização das técnicas passivas para cenas com objetos sem textura, as hipóteses apresentadas nesse trabalho não permitem reconstrução de ambientes abertos, além de outros problemas. Logo, em geral, ainda se admite que técnicas passivas não conseguem reconstruir cenas sem textura.

Uma opção para resolver esse problema é adicionar fisicamente uma textura a cena, inserindo pontos na superfície. Porém, o fato é que nem sempre essa abordagem é possível (e.g., museus não podem colar marcadores nas estátuas).

As técnicas ativas são baseadas no princípio de projeção de padrões de luz na cena, e consequente observação de como tais padrões se deformam para calcular a forma 3D dos objetos. Pode-se perceber que essencialmente uma textura artificial é mapeada ao objeto. Essa textura produz características salientes da imagem (i.e., pontos, retas, curvas, malhas), possibilitando uma reconstrução 3D densa.

Independentemente de qual técnica ativa está sendo usada, a principal vantagem dessa categoria de técnicas é que um conjunto denso de pontos pode ser extraído, mesmo para objetos sem textura. Em alguns casos, o brilho especular da superfície dos objetos pode ser tratado até um certo nível (objetos totalmente especulares não são tratáveis usando triangulação tradicional). Por outro lado, a maioria das técnicas de *scanning* só produz visões 3D parciais da cena, o que é chamado de *range data* [31] e [5]. Para conseguir um modelo 3D completo é necessário fazer a fusão dos modelos gerados, segundo [31]. Para a realização dessa etapa não existe uma solução totalmente automática e, além disso, está-se sujeito a erros de calibração para cada imagem. Segundo [32] o menor erro introduz diferenças no processo de alinhamento das imagens, e [26] diz que mais de 80% do tempo total de modelagem se divide entre registro das imagens e criação da malha global.

³ (*Complementary Metal Oxide Semiconductor* ou Semicondutor de Metal Óxido Complementar) Um *chip* que guarda as informações essenciais do computador, como data, hora, informações sobre o *winchester* e etc. Quando a máquina está desligada, ele é alimentado por uma bateria.

ortogonais e definem um sistema de coordenadas de mão-direita (o eixo (O_c, Z_c) foi escolhido para ser perpendicular ao plano da imagem). Recomenda-se [9] para mais detalhes sobre sistemas de coordenadas. Veja a Figura 7; no quadro de referência da câmera $F = (O_c, X_c, Y_c, Z_c)$, um ponto P no espaço tem coordenadas $\bar{X} = [X, Y, Z]^T$. A projeção em perspectiva de p sobre o plano da imagem tem coordenadas $\bar{x} = [X/Z, Y/Z]^T$.

Assim, de acordo com a Equação 4.1, um plano Π no espaço pode ser representado por um conjunto de pontos P de coordenadas homogêneas \bar{X} tais que:

$$\langle \bar{\pi}, \bar{X} \rangle = 0 \quad 4.1$$

Onde $\bar{\pi} = [\pi_x, \pi_y, \pi_z, \pi_t]^T$ é um vetor de coordenadas homogêneas que parametriza um plano. O operador $\langle \cdot \rangle$ é o de produto interno, usualmente utilizado na matemática.

Seja p a projeção no plano de imagem de um ponto P de coordenadas $\bar{X} = [X, Y, Z]^T$, e denote por $\bar{x} = [x, y]^T$ o vetor de coordenadas associado ao plano de imagem. De acordo com a descrição dada na Seção 2.5, esses pontos podem ser relacionados de acordo com a equação da câmera estenoica (ou *pinhole*), apresentada na Equação 4.2:

$$\bar{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} X \\ Y \end{bmatrix} \quad 4.2$$

Segundo [22], em problemas de Visão Computacional algumas vezes é válido tratar o plano de imagem como um espaço projetivo de dimensão 2. Nessa representação, um ponto p no plano de imagem é representado por um vetor de coordenadas homogêneas $\bar{x} = [x, y, 1]^T$.

A grande vantagem de utilizar geometria projetiva nesse caso é o fato de que o operador de projeção, definido na Equação 4.2 agora é linear, como apresentado na Equação 4.3:

$$\bar{x} = P\bar{X} \text{ com } P = [I_{3 \times 3}, 0_{3 \times 1}] \quad 4.3$$

Onde \bar{X} e \bar{x} são os vetores de coordenadas homogêneas associados aos pontos P e p , respectivamente, $I_{3 \times 3}$ é a matriz identidade e $0_{3 \times 1}$ é um vetor de zeros. Assim, o vetor \bar{x} é igual a \bar{X} ,

apenas pela multiplicação de um escalar (Equação 4.4), ou seja:

$$\bar{x} \simeq \bar{X} \quad 4.4$$

Assim, pode-se definir que o vetor \bar{x} é o vetor de direção associado ao ponto P .

Além de um plano, define-se uma reta λ no plano de imagem como sendo o conjunto de vetores de coordenadas homogêneas, conforme Equação 4.5, tais que:

$$\langle \bar{\lambda}, \bar{x} \rangle = 0 \quad 4.5$$

Onde $\bar{\lambda} = [\lambda_x, \lambda_y, \lambda_z]^T$ é um vetor de coordenadas homogêneas que definem a reta λ .

Existem várias relações entre retas no plano de imagem e planos no espaço. Em [1] uma nova parametrização dessas relações é descrita e chamada de espaço dual-B, ou somente, espaços duais.

4.1.2 Espaços Duais

Como mostrado na sub-seção anterior, um plano Π pode ser representado por um vetor de coordenadas homogêneas $\bar{\pi} = [\pi_x, \pi_y, \pi_z, \pi_t]^T$ num quadro de câmera $F = (O_c, X_c, Y_c, Z_c)$. Pode-se afirmar ainda que, caso o plano não contenha a origem do quadro da câmera O_c , pode-se defini-lo como sendo $\bar{\omega} = [\omega_x, \omega_y, \omega_z]^T$, conforme a Equação 4.6:

$$\langle \bar{\omega}, \bar{X} \rangle = 1 \quad 4.6$$

Seja $(\Omega) = \mathbb{R}^3$. Uma vez que todo ponto $\bar{\omega} \in (\Omega)$ define um plano único no espaço euclidiano (E) , define-se que (Ω) é o espaço de plano ou espaço dual-B. Existe uma relação bastante simples entre os planos descritos em geometria projetiva e no novo formalismo de espaços duais, apresentada na Equação 4.7:

$$\bar{\omega} = \frac{-1}{\pi_t} \begin{bmatrix} \pi_x \\ \pi_y \\ \pi_z \end{bmatrix} \text{ se } \pi_t \neq 0 \quad 4.7$$

Percebe-se que nesse sentido, a geometria de espaço dual-B não é um conceito novo. Originalmente,

no espaço tridimensional, um plano pode ser representado por um vetor de coordenadas homogêneas $\bar{\pi} = [\pi_x, \pi_y, \pi_z, \pi_t]^T$. Um ponto P de coordenadas homogêneas $\bar{X} = [X, Y, Z, 1]^T$ está em um plano genérico Π de coordenadas $\bar{\pi}$ tal que $\langle \bar{\pi}, \bar{X} \rangle = 0$. A grande contribuição do formalismo de espaço dual-B é a nova parametrização que facilita a relação de objetos no espaço Euclidiano (planos, retas e pontos) e suas projeções no plano de imagem (retas e pontos). Basta lembrar que com essa nova parametrização, todos os planos que não intersectam a origem estão modelados, e essa não é uma restrição na técnica, como será mostrado a seguir.

4.2 Descrição da Técnica de Bouguet

A Técnica de Bouguet consiste basicamente em lançar um conjunto de sombras, com uma vareta, sobre a cena a qual deseja-se reconstruir, e estimar a profundidade pelas sombras deformadas. Além disso, existem duas variações da técnica, uma aplicando um plano vertical e outra não. A diferença entre ambas é que na variação sem plano vertical a fonte de luz é calibrada, enquanto que no caso dos dois planos a fonte de luz não necessariamente está calibrada. A Figura 9 ilustra a utilização da técnica; pode-se observar que na figura central existe o plano vertical, enquanto que nas outras apenas o plano horizontal é utilizado.



Figura 8: Utilização da Técnica de Bouguet

Basicamente, a técnica consiste em determinar a profundidade de cada *pixel* da imagem. É fácil ver que, dada a localização da fonte de luz e a posição da vareta, para cada imagem, forma-se um plano. A fronteira da sombra lançada na cena é a interseção desse plano com a superfície do objeto. No processo de reconstrução, as inclinações na fronteira da sombra são exploradas para calcular a profundidade de cada *pixel*.

A Figura 9 mostra o princípio geométrico aplicado na técnica.

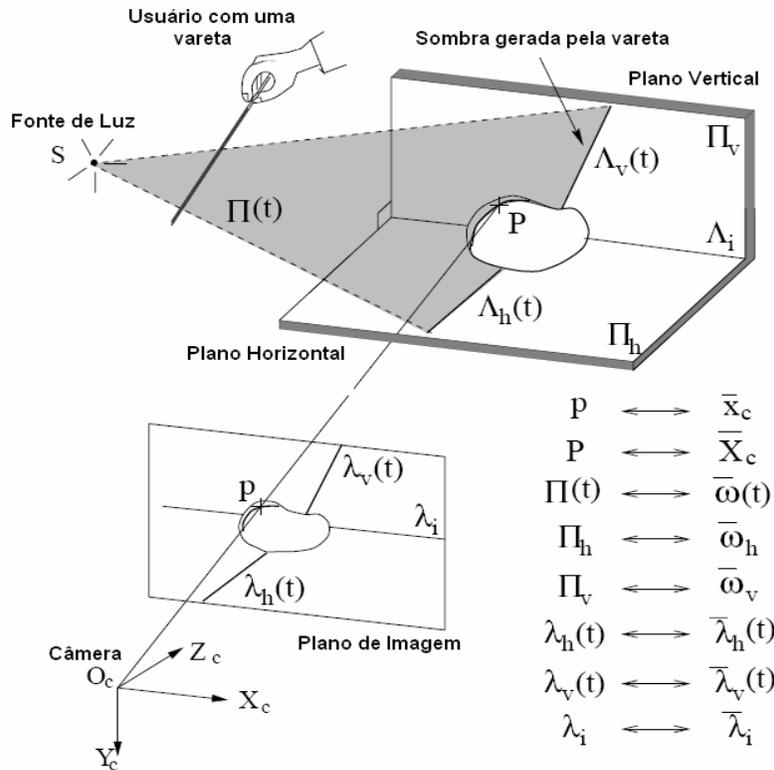


Figura 9: Princípio geométrico aplicado na Técnica de Bouguet

A seguir, para fazer com que o problema seja tratável do ponto de vista matemático, serão feitas várias simplificações; para um tratamento mais rigoroso do ponto de vista geométrico, [1] e [2] são recomendados.

A fonte de luz é aproximada como sendo pontual e de localização S , o plano horizontal (e.g., a mesa onde o objeto está apoiado) é representado por Π_h , e Π_v representa um plano ortogonal a Π_h . Assume-se, *a priori*, que a posição de Π_h é conhecida devido à calibragem da câmera (mais detalhes na Seção 5.1.1). Os parâmetros de π_v são facilmente inferidos, dada a projeção λ_i da reta de interseção entre os planos Π_h e Π_v . O objetivo é estabelecer uma estimativa da localização de cada ponto P no espaço correspondente a cada *pixel* p de coordenadas \bar{x}_c na imagem. Chama-se de t o tempo, medido em qualquer unidade de tempo, em que a sombra passou por um dado *pixel* \bar{x}_c ; o tempo t é chamado de tempo de sombra. Denota-se por $\Pi(t)$ o plano de sombra correspondente ao tempo t . Assume-se, sem perda de generalidade, que porções da sombra projetada sobre os planos Π_h e Π_v podem ser visualizadas e possuem coordenadas $\lambda_h(t)$ e $\lambda_v(t)$. Após a extração dessas duas retas, deduz-se a localização das retas

correspondentes $\Lambda_h(t)$ e $\Lambda_v(t)$, simplesmente intersectando os planos $(O_c, \lambda_h(t))$ e $(O_c, \lambda_v(t))$ com Π_h e Π_v , respectivamente. Uma vez feito isso, o plano $\Pi(t)$ é definido pelas duas retas não-colineares $\Lambda_h(t)$ e $\Lambda_v(t)$. Para uma prova desse fato, recomenda-se a leitura da Seção 6.2.5 em [1].

Como dito anteriormente, existem duas variações da técnica, e no caso da variação sem o plano ortogonal π_v , o plano $\Pi(t)$ é inferido diretamente da reta $\lambda_h(t)$ e da posição S da fonte de luz, bastando que nesse caso a fonte de luz esteja calibrada.

Finalmente, o ponto P pode ser calculado intersectando o raio (O_c, p) com o plano $\Pi(t)$. Essa etapa é chamada de triangularização.

A Técnica de Bouguet pode ser resumida em quatro passos: (a) estime o tempo de sombra $t_s(\bar{x}_c)$ para cada *pixel* \bar{x}_c (a essa etapa dá-se o nome de processamento temporal); (b) localize as retas de referência $\lambda_h(t)$ e $\lambda_v(t)$ para cada instante t (processamento espacial); (c) calcule o plano de sombra; e, finalmente, (d) triângule e calcule a profundidade.

4.2.1 Localização da Fonte de Luz

Um passo fundamental dessa técnica de reconstrução é a localização da fonte de luz. Em [1], esse problema está resolvido utilizando o formalismo proposto pelo autor, mas restrito às seguintes hipóteses: (a) fonte de luz pontual; (b) ponto de luz não coincide com ponto da câmera; (c) o ponto de base e a altura do objeto utilizado (no caso, um lápis) são fornecidos pelo usuário; e (d) objeto utilizado sempre está visível nas imagens de calibração.

Com a hipótese (c) satisfeita, juntamente com os parâmetros extrínsecos da câmera, fornecidos pela calibração da câmera, é trivial calcular a coordenada 3D do ponto de luz. Sabe-se que o ponto luminoso tem que estar no raio formado pela extremidade do lápis e a extremidade da sombra. Logo, com dois ou mais desses raios candidatos, pode-se achar sua interseção, isso é, a localização do ponto luminoso.

Caso mais de duas imagens sejam utilizadas, pode-se usar a média das localizações encontradas em cada par de imagens. São gerados $(N*(N-1))/2$ pares de imagens e correspondentes localizações de pontos luminosos, onde N é o número de imagens de entrada. Essa média ameniza os erros associados e também compensa o fato de que a luz não é uma fonte de luz pontual.

4.2.2 Processamento Espacial e Temporal

Um passo fundamental é a detecção das retas de interseção do plano de sombra $\Pi(t)$ com os dois planos Π_h e Π_v . Um jeito simples de extrair $\bar{\lambda}_h(t)$ seria assumir, sem perda de generalidade, que existem porções da imagem que estão livres de objetos (e.g., as primeiras e as últimas linhas da imagem). Partindo desse pressuposto, existem dois passos a serem realizados: (a) localizar as arestas de sombra que estão projetadas nos planos a cada tempo t , o que resultará no conjunto de todos os planos de sombra; (b) estimar o tempo de sombra $t_s(\bar{x}_c)$ associado a cada *pixel* $\bar{x}_c = (x, y)$. Em [26], é apresentada uma justificativa para o uso de processamento espaço-temporal, relatando que o mesmo é apropriado para tratar de descontinuidades na localização das arestas, além de reduzir problemas de distorções de faixa.

Ambas tarefas consistem na localização da aresta de sombra, mas a busca atua em domínios diferentes: uma opera em um domínio espacial (coordenadas de imagens) e a outra no domínio temporal. Apesar de, à primeira vista, parecer que as tarefas são independentes, elas precisam estar relacionadas e restritas a seguinte condição: se em um dado tempo t_0 a aresta de sombra passa pelo *pixel* $\bar{x}_c = (x, y)$, ambas as buscas devem achar o mesmo ponto (x_c, y_c, t_0) .

Assim, é necessário estabelecer um critério único de descrição das arestas de sombra no domínio espacial e temporal. A técnica aplicada, chamada limitante espaço-temporal, é simples e satisfaz a propriedade esperada. Consiste na observação do seguinte fato: uma vez que as sombras são geradas e projetadas na cena, cada *pixel* vê sua intensidade de brilho variar de seu valor normal $I_{\max}(x, y)$, ou seja, quando a sombra não passou ainda naquele *pixel*, para um valor mínimo $I_{\min}(x, y)$, e voltando ao seu valor normal.

Para qualquer *pixel* $\bar{x}_c = (x, y)$, define-se $I_{\max}(x, y)$ e $I_{\min}(x, y)$ como sendo o maior e o menor valor de brilho encontrados durante toda a seqüência de imagens. Define-se a aresta de sombra como sendo as localizações onde o *pixel* da imagem $I(x, y, t)$ intersecta a imagem limiar $I_{\text{sombra}}(x, y)$, definida como sendo a média aritmética dos valores $I_{\max}(x, y)$ e $I_{\min}(x, y)$.

Intuitivamente, *pixels* correspondentes a regiões ocluídas na cena não oferecem informação a respeito da sua profundidade. Assim, *pixels* com um valor de contraste $I_{\text{contraste}}(x, y) = I_{\max}(x, y) - I_{\min}(x, y)$ menor que um valor fornecido I_{limiar} são desconsiderados. Esse parâmetro de contraste deve ser controlado de acordo com o nível de ruído encontrado na imagem.

Uma vez que todos esses valores tenham sido computados ao longo da linha de referência (uma das linhas nas quais a busca pelas arestas será realizada), percorre-se a linha mais uma vez, procurando pelo *pixel* que tem seu valor de brilho menor que $I_{\text{sombra}}(x, y)$. É necessário fazer uma interpolação linear simples entre o *pixel* imediatamente anterior ao que está sendo atravessado e o *pixel* imediatamente posterior ao qual a aresta de sombra está atravessando na linha de referência. Uma vez que está acontecendo uma interpolação, valores do nível de *sub-pixel* podem porventura aparecer.

Por outro lado, computar o tempo no qual uma aresta de sombra atravessa um dado *pixel* é feito de uma maneira bastante similar ao processamento espacial. Contudo, nesse caso, valores de *pixels* ao longo do tempo estão sendo procurados, com a posição fixada. No caso anterior, a posição variava mas o tempo permanecia fixado.

Após a primeira varredura das imagens, faz-se uma segunda e acha-se o primeiro quadro no qual a imagem ultrapassa o limiar $I_{\text{sombra}}(x, y)$. Interpola-se esse quadro e o quadro imediatamente anterior a esse para achar o tempo exato no qual a aresta de sombra atravessou o *pixel*. Mais uma vez, esse valor é dado num nível abaixo da taxa de amostragem do quadro, ou seja, em um nível de sub-quadro.

4.2.3 Estimativa do Plano de Sombra e Triangulação

A partir da calibração da câmera obtém-se a normal do plano de referência e a distância perpendicular desse plano para o centro óptico c da câmera (a origem do sistema de coordenadas).

A fim de encontrar a coordenada 3D de um *pixel* no plano, procura-se pelo tempo de sombra para aquele *pixel* (pré-calculado para todos os *pixels*). Percorre-se a lista de pontos de referência calculados para cada imagem, para achar o tempo exato no qual a aresta de sombra atravessou o *pixel*. Já que o tempo se encontra em um nível de sub-quadro, faz-se outra interpolação linear dos pontos de referência, que foram calculados para cada quadro. Esses pontos de referência são pontos no plano de imagem. Uma vez que um raio é disparado através desses pontos de referência e intersecta essa lista com o plano de referência, consegue-se as coordenadas 3D dos pontos de referência para um dado tempo de sombra. Esses pontos de referência, em conjunto com a posição da fonte de luz (que já foi calculada na calibração da fonte de luz), formam um novo plano, o plano de sombra.

Com as coordenadas do plano de sombra e do raio associado ao *pixel* em questão, basta calcular a interseção desse raio com o plano de sombra para saber a profundidade do *pixel*. A esse passo dá-se o nome de triangulação.

5 Implementação e Resultados

Conforme descrito no Capítulo 4, existem duas variações da Técnica de Bouguet: (1) com dois planos de referência e uma fonte de luz descalibrada, e (2) com apenas um plano de referência e uma fonte de luz calibrada. Ambas variações foram implementadas neste trabalho, no entanto, (1) apresentou erros em uma das etapas fundamentais do processo, a localização do plano de referência vertical, além de problemas na estimativa da fonte de luz. A partir desses fatos, a variação (2) foi escolhida para ser implementada.

Dito isto, seguem abaixo um detalhamento da implementação e os resultados obtidos.

5.1 Implementação

Uma das principais fontes de algoritmos de visão computacional implementados existente hoje é a biblioteca OpenCV (*Open Source Computer Vision*) da Intel [3]. Esta biblioteca foi utilizada neste trabalho em praticamente todas as etapas do processo, desde a calibração de câmera, que é uma funcionalidade implementada como um conjunto de funções de baixo nível na biblioteca, até o processamento de imagens espacial e temporal. Nessa última etapa, foram utilizadas funções de processamento de imagens disponíveis na biblioteca.

Devido a restrições de tempo, neste trabalho todos os testes foram realizados com imagens geradas sinteticamente. A implementação foi testada com uma variedade razoável de cenas, desde cenas simples a cenas mais complexas contendo modelos de faces. Apesar de parecer um grande risco o fato da aplicação não ter sido testada com fotos reais, a quantidade de cenas do conjunto de testes oferece resultados bastante satisfatórios, que apontam que a implementação está correta.

Basicamente, o projeto consiste de três partes distintas, descritas nas próximas sub-seções:

1. Calibração da câmera;
2. Calibração da fonte de luz;
3. Cálculo da profundidade de cada *pixel* na cena.

A técnica não especifica uma maneira específica de visualização, uma vez que a saída do algoritmo é uma nuvem de pontos. Logo, pode-se optar por criar desde malhas de triângulos até de quadrados, ou até utilizar uma técnica de síntese de imagem mais realista, como o Raytracing [25] ou o Photon Mapping [24]. Por questões de simplicidade e agilidade, um programa de modelagem bastante conhecido, o Blender [10], foi utilizado para visualização do modelo reconstruído, gerando malhas quadrangulares.

5.1.1 Calibração da Câmera

Como dito anteriormente, utilizou-se a biblioteca OpenCV [3] para implementação da Técnica de Bouguet, descrita no capítulo anterior. Essa biblioteca oferece funções de baixo nível para encontrar parâmetros intrínsecos, extrínsecos e matrizes fundamentais. Apesar de existirem outras alternativas de mais alto nível para atacar o problema de calibragem de câmera, tais como [11] e [12], optou-se por utilizar as funções oferecidas pela biblioteca para calibração de câmera. Uma descrição completa da técnica implementada pela biblioteca pode ser encontrada em [23].

5.1.2 Calibração da Fonte de Luz

A localização da luz é calculada usando geometria básica, como descrito na Seção 4.2.1. Devido ao fato do algoritmo de calibração de fonte de luz só tratar pares de imagens, surgiu uma dificuldade com relação a escolha das imagens de entrada, por conta do erro associado a cada imagem. Foi visto que as nuvens de pontos finais eram bastante robustas à localização da fonte de luz, tanto quanto o valor esteja próximo do correto. Esse fato foi notado durante os testes. No entanto, usar a média dos candidatos ajudou a amenizar os erros, fazendo com que uma melhor estimativa da localização da fonte de luz fosse conseguida.

Um problema do esquema proposto por Bouguet é a premissa de que a fonte de luz é pontual, uma vez que não há intervenção humana na calibração da fonte de luz, bastando apenas que sejam fornecidas a altura do lápis, as localizações do ponto da base do lápis e qual o pixel correspondente a ponta da sombra. Nos testes realizados, essas medidas foram fornecidas antes do processo de calibração, isto é, para cada imagem utilizada na calibração de fonte de luz está associada uma tripla com os valores da altura do lápis, pixel da base e pixel da sombra do lápis, a figura Figura 10 abaixo ilustra o que foi descrito:

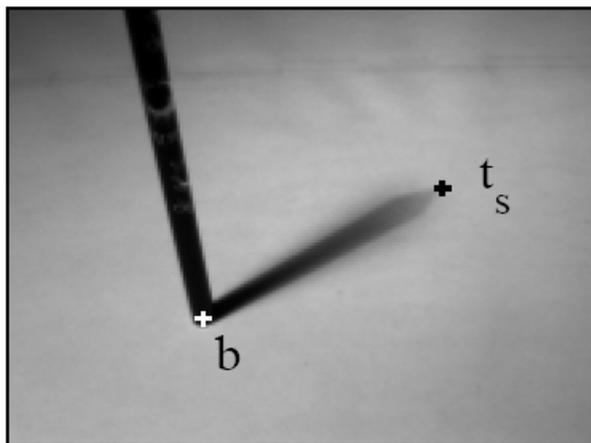


Figura 10: *Pixels* da base do lápis e ponta da sombra

5.1.3 Cálculo da Profundidade

A Figura 11 ilustra o princípio geométrico utilizado pela Técnica de Bouguet com um plano e com fonte de luz calibrada, que, como dito anteriormente, foi a variação escolhida para ser implementada neste trabalho.

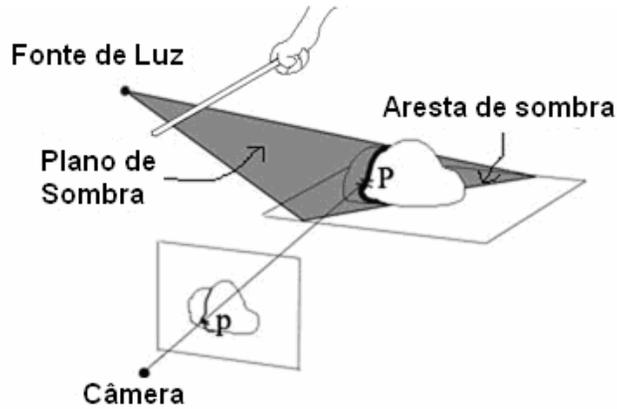


Figura 11: Princípio geométrico da variação da Técnica de Bouguet com apenas um plano e fonte de luz calibrada

Antes de começar a tratar as imagens, todas são transformadas em imagens em escala de cinza. Além disso, devem ser fornecidas duas linhas, uma superior e outra inferior, em coordenadas de imagem, para poder diminuir o espaço de busca relativo a procura dos planos de sombra, como mostrado na figura Figura 12 abaixo:

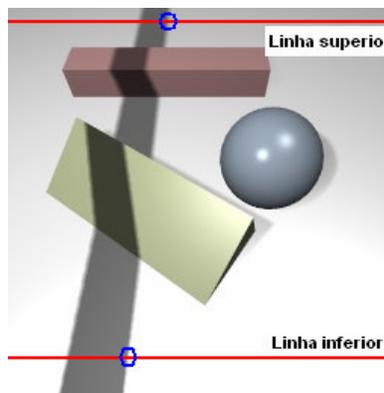


Figura 12: Linhas superior e inferior fornecidas pelo usuário

Esse procedimento pode ser descrito por três passos: estimativa do plano de sombra ou processamento espacial, estimativa do tempo de sombra ou processamento temporal, e triangulação. Esses

passos são detalhados a seguir.

5.1.3.1 Estimativa do Plano de Sombra ou Processamento Espacial

Esse passo, denominado localização da aresta de sombra espacial, envolve encontrar, para cada linha da imagem, o *pixel* pelo qual a sombra passa. Mais especificamente, um par de *pixels* tem que ser identificado com as intensidades mudando de “normal” para “escuro”. Dada uma outra localização, em outro tempo, um critério universal é utilizado para essa classificação:

$$\begin{aligned}I_{\min} &= \min(I(x, y, t)) \\ I_{\max} &= \max(I(x, y, t)) \\ I_{\text{sombra}} &= (I_{\min} + I_{\max}) / 2\end{aligned}$$

Onde x, y são as coordenadas do *pixel* e t é o tempo. Qualquer intensidade maior que I_{sombra} é considerada normal, e intensidades abaixo dessa são consideradas sombra.

Uma vez que as arestas de *pixels* foram coletadas, uma reta 2D é associada aos *pixels* pelo método de mínimos-quadrados. A projeção dos pontos 3D dessa reta é então calculada usando a equação do plano de referência, que é conhecida por conta da calibração da câmera. O plano de sombra é definido como sendo o plano que contém a aresta de sombra e a fonte de luz. A partir da aresta de sombra e as duas linhas fornecidas (uma superior e outra inferior), pode-se calcular dois pontos pertencentes a tais linhas. E com esses dois pontos e o ponto da fonte de luz, forma-se o plano de sombra associado a um dado *pixel*.

5.1.3.2 Estimativa do Tempo de Sombra ou Processamento Temporal

Esse passo requer analisar a intensidade de cada *pixel* ao longo dos quadros, achando os pares de quadros onde um dado *pixel* varia de normal para escuro. O critério utilizado para essa busca baseia-se no que foi apresentado na sub-seção anterior.

A grande dificuldade dessa etapa é relacionar o *pixel* que se quer triangular com o plano de sombra, num determinado tempo de sombra. *A priori*, os valores encontrados para os planos de sombra são encontrados em uma frequência maior que a taxa de amostragem das imagens, o que força o tempo de sombra ser uma interpolação de dois tempos de sombra.

5.1.3.3 Triangulação

As coordenadas 3D de cada *pixel* são obtidas simplesmente intersectando o raio óptico originado do

centro da câmera, que passa pelo *pixel* no plano de imagem, e o plano de sombra associado àquele *pixel* num dado tempo de sombra.

5.1.4 Visualização da Nuvem de Pontos

Para a visualização dos pontos gerados pelo algoritmo, foi utilizada a ferramenta Blender [10]. O Blender é um *software open source* de modelagem 3D, animação, renderização, pós-produção, entre outras funcionalidades. É bastante utilizado e foi escolhido por ser relativamente fácil de usar e adequado para a visualização da nuvem de pontos gerada pelo algoritmo desenvolvido neste trabalho. A Técnica de Bouguet não se preocupa em oferecer uma malha triangularizada dos pontos, e sim somente os pontos, havendo a necessidade de gerar uma malha visível externamente ao algoritmo.

A solução encontrada foi usar a triangularização oferecida pelo Blender. O programa em questão oferece um modo de programação na linguagem de *script* Python, o que faz com que seja fácil gerar malhas para visualização. A interface do Blender é ilustrada na Figura 13, e um trecho de código em *script* é destacado em vermelho.

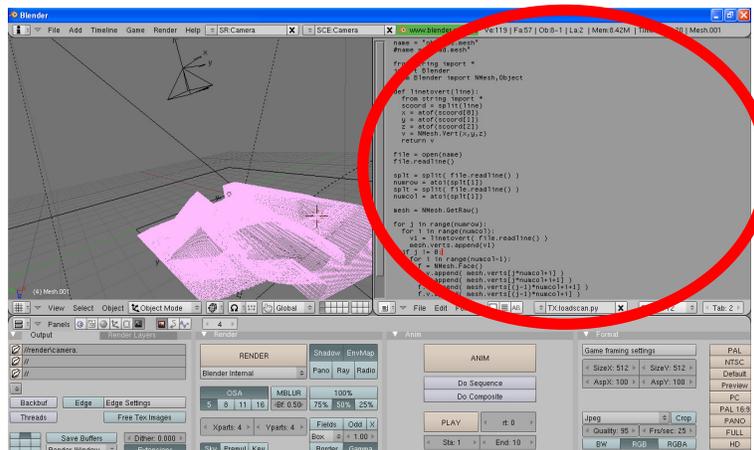


Figura 13: Interface da ferramenta Blender com *script* Python em destaque

5.2 Resultados

Nesta seção, os resultados alcançados pela implementação da variante (2) da técnica de reconstrução 3D proposta por Jean-Yves Bouguet em [1] são apresentados, bem como algumas dificuldades encontradas.

As imagens utilizadas para reconstrução 3D foram geradas sinteticamente, no Blender. Para cada cena foram geradas aproximadamente 250 imagens. E, para visualização da malha final, também se utilizou o Blender.

Primeiramente testou-se a calibração da fonte de luz, a qual apresentou erros bastante frequentes devido a espinhos que apareciam por conta de falhas no cálculo dos planos de sombra associados a cada imagem. Esse resultado ressalta o fato de que a calibração da fonte de luz é, realmente, uma etapa primordial na implementação correta dessa variante da Técnica de Bouguet.

Um outro ponto é que a nuvem final de pontos extraídos dos dados apresenta um certo ruído. A justificativa encontrada para o ruído encontra-se na dificuldade em encontrar um valor limiar razoável para a sombra, além de variações nas superfícies, e de problemas de arredondamento relativos a extração da interseção do raio com o plano de sombra.

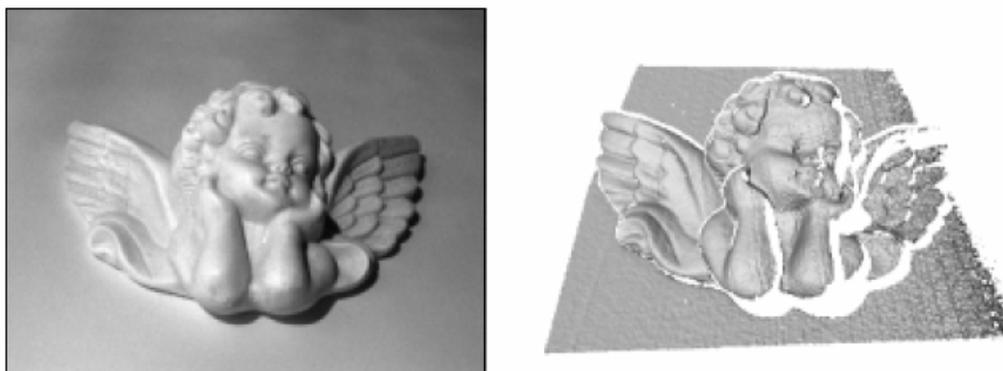


Figura 14: Cena com objeto claro e pequeno, reconstruída por Bouguet

A análise temporal funcionou bem, visto que não foram encontrados muitos problemas com relação às arestas de sombra para cada imagem. Essa etapa do processo não influenciou no ruído encontrado nos resultados, pois foram realizados testes para verificar se o cálculo do tempo de sombra estava de acordo com as imagens analisadas, e de fato, foi encontrado que em mais de 95% dos testes, as imagens com sombras foram encontradas corretamente.

A análise espacial, utilizada para encontrar os pontos de referência, foi bastante problemática. Olhando-se algumas imagens, percebeu-se que geralmente ocorria uma penumbra na aresta de sombra, provocada pelo limiar de sombra, calculado para estar no meio de uma penumbra muito grande. Visualmente, esses pontos de referência mostraram que, geralmente, eles eram encontrados no meio da penumbra e nunca no valor definido para ser sombra. Para piorar, a penumbra, aparentemente, mudou um pouco de tamanho à medida que a vareta foi movimentada ao longo do objeto, o que ocasionou problemas no cálculo do plano de sombra, por estes não serem consistentes entre uma imagem e outra.

Infelizmente, por escassez de tempo, apenas um módulo virtual da técnica foi implementado, o que

acarretou dificuldades no cálculo da margem de erro dos resultados apresentados neste trabalho. Bouguet cita que conseguiu uma margem de erro de 1.5% em cenas com objetos claros e pequenos, como os ilustrados na Figura 14 acima.

Em termos de desempenho, não é possível realizar uma comparação entre os resultados apresentados por Bouguet e a implementação aqui apresentada, visto que a capacidade de processamento dos computadores atualmente é bastante superior aos da data em que Bouguet realizou sua implementação. Num conjunto de 250 imagens com resolução de 256×256 *pixels*, a implementação realizada demora cerca de 20 segundos calculando os planos de sombra e mais 10 a 12 segundos no cálculo da profundidade dos *pixels*. Em outras palavras, leva-se em média 30 segundos para calcular a malha final.

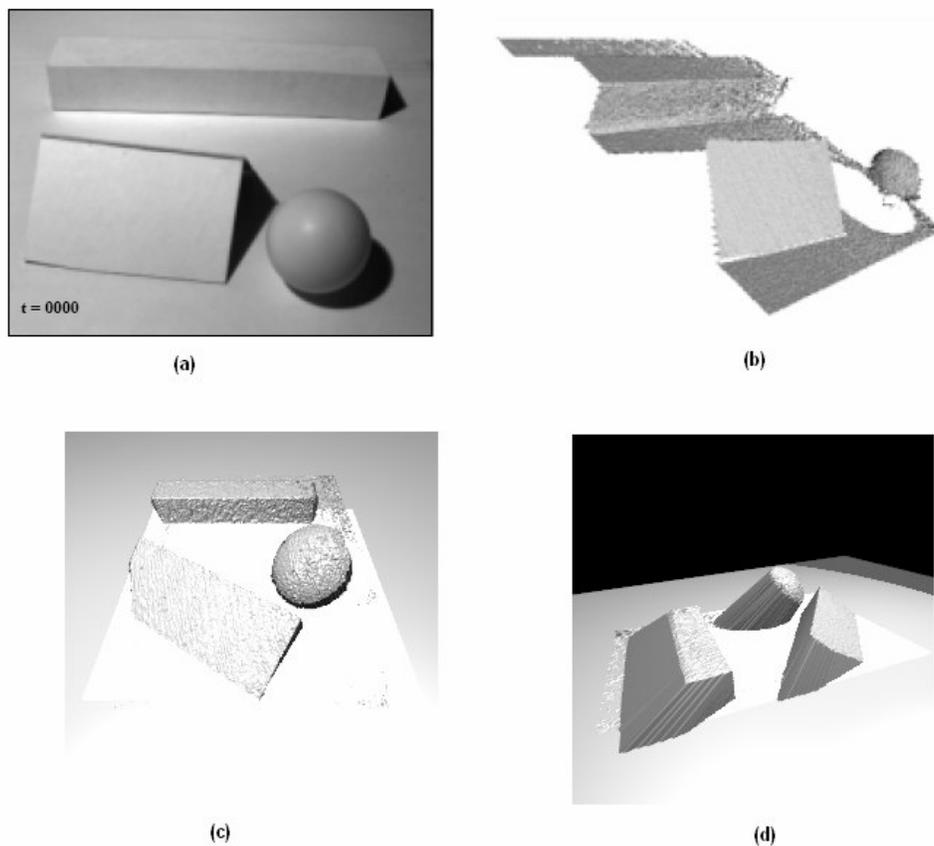


Figura 15: (a) Cena modelada. (b) Resultado de Bouguet. (c) Resultado da implementação realizada neste trabalho, (d) Resultado encontrado de outro ponto de vista

A Figura 15 mostra um dos resultados encontrados, em (a) encontra-se a cena modelada com 2 sólidos e 1 esfera dispostos em um plano, já em (b) está o resultado produzido por Bouguet em um ponto de vista diferente do apresentado em (a), em (c) o mesmo ponto de vista da cena modelada é reproduzido pela

implementação realizada, em (d) a mesma cena é visualizada de outro ponto de vista.

Na Figura 15 pode-se perceber uma diferença fundamental entre a implementação de Bouget e a realizada no presente trabalho: na implementação de Bouguet os raios que atingem partes escuras não tem profundidade associada, porém na implementação realizada simplesmente projeta-se os raios até atingir o plano de referência. Essa escolha foi realizada por razões de simplicidade.

Uma das limitações desse método é o fato de que, diferentemente de outras técnicas ativas, essa técnica baseia-se no bloqueio da luz, ao invés da projeção de um padrão de luz na cena. Uma grande dificuldade é reconstruir objetos com muitas partes escuras, pois neste caso torna-se impossível a definição de um limiar aceitável, o que acarreta erros perceptíveis na reconstrução de tais objetos.

6 Conclusão e Trabalhos Futuros

Nesse trabalho de conclusão de curso de graduação foram apresentados conceitos relacionados com áreas correlatas a de Reconstrução 3D, entre eles, geometria projetiva, e conceitos relacionados a fotogrametria, utilizados por técnicas de Visão Computacional e como estas se relacionam com a área de Reconstrução 3D.

As dificuldades associadas a cada categoria de técnica foram elucidadas. Concluiu-se que técnicas ativas têm um grande peso quando se quer modelar cenas com grande eficácia, rapidez e relativa facilidade de implementação. Isto se deve ao fato de que, normalmente, neste caso trabalha-se com ambientes controlados. Já as técnicas passivas possuem um grau maior de flexibilidade, apesar do conjunto menor de restrições, uma vez que o processo de reconstrução se dá utilizando câmeras comuns de baixo custo e sem o controle da iluminação do ambiente.

O presente trabalho se propôs a implementar uma técnica ativa de reconstrução 3D, proposta por Jean-Yves Bouguet e os resultados obtidos, apesar de mais restritos, estão de acordo com os relatados por Bouguet. Chegou-se a conclusão de que o principal ponto fraco dessa técnica são as cenas com objetos escuros ou com sombra projetada em suas superfícies.

Finalmente, esse trabalho está inserido no contexto de um projeto maior, já que serviu de base para os trabalhos na área de reconstrução 3D do Grupo de Pesquisa em Realidade Virtual e Multimídia – GRVM. Este projeto tem por objetivo utilizar técnicas passivas para reconstrução de cenas de ambientes abertos.

6.1 Trabalhos Futuros

Como trabalhos futuros, pode-se citar a utilização de mais que uma fonte de luz, o que implica na combinação dos resultados obtidos com cada fonte de luz. Bouguet oferece um método simples de fazer a combinação de mais de uma malha final.

Outra extensão seria a utilização de texturas na malha final. Bouguet também cita um método simples de realizar essa operação.

Uma outra extensão que poderia ser realizada é a implementação da técnica em GPU⁴, devido ao alto

⁴ GPU (*Graphics Processing Unit*, ou Unidade de Processamento Gráfico), é um microprocessador especializado em processar gráficos. São utilizadas em placas de vídeo para fazer computação gráfica.

grau de paralelismo oferecido pela mesma. Essa abordagem permitiria a implementação da técnica em tempo-real, para possível aplicação em Realidade Aumentada [4].

7 Referências

- [1] BOUGUET, J. Visual Methods for Three-dimensional Modeling, California Institute of Technology, 1999. Ph.D. Thesis.
- [2] FAUGERAS, O. Three-Dimensional Computer Vision. Primeira Edição, MIT Press, Massachusetts 1993, 695 p.
- [3] OPENCV. Open Source Computer Vision Library. URL: <http://www.intel.com/technology/computing/opencv/>. Consultado em 11 set. 2006.
- [4] BIMBER, O.; RASKAR, R., Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters, Ltd., Massachusetts, 2005, 392p.
- [5] HARTLEY, Richard; ZISSERMAN, Andrew, Multiple View Geometry in Computer Vision, Segunda Edição, Cambridge University Press, Cambridge, 2004, 672 p.
- [6] BEARDSLEY, P.A.; REID, I.D., Affine Calibration of Mobile Vehicles. Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, pág. 214-221 1995.
- [7] BLOOMENTHAL, J.; ROKNE, J. Introduction to Homogeneous Coordinates. Department of Computer Science. The University of Calgary, The Visual Computer, 1994.
- [8] TRIGGS, B. Projective Geometry for Image Analysis. URL: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/MOHR_TRIGGS/isprs96.html. Consultado em 15 set 2006.
- [9] FOLLEY, J.D.; VAN DAM, A. Computer Graphics – Principles and Practice. Segunda Edição, Addison Wesley, 1995.
- [10] BLENDER. URL: <http://www.blender.org>. Consultado em 10 set de 2006.
- [11] VLADIMIR V., OpenCV and MatLab Camera Calibration Toolboxes Enhancement. URL: <http://graphics.cs.msu.su/en/research/calibration/index.html>. Consultado em 18 de set de 2006.
- [12] ZHANG, Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.
- [13] TSAI, R Y. A Versatile Camera Calibration Technique for high-accuracy 3D Machine Vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal of Robotics and Automation, 3(4):323-344, 1987.
- [14] GRUEN, A.; HUANG, T. Calibration and Orientation of Cameras in Computer Vision. Vol. 34 de Springer Series in Information Sciences, Springer-Verlag, 2001.
- [15] MCGLOVE, C.; MIKHAIL, E. e BETHEL, J. Manual of Photogrammetry. American Society for Photogrammetry and Remote Sensing, Quinta Edição, American Society of Photogrammetry and Remote Sensing, 2004, 563 p..
- [16] WILLSON, R.G.; SHAFER, S. A. What is the Center of the Image? Journal of Optical Society of America, Vol. 11, No. 11, pp.2946-2955, 1994.

- [17] HEIKKILA, J.; SILVEN, O. A four-step camera calibration procedure with implicit image correction. Conference on Computer Vision and Pattern Recognition, pp. 1106–1112, 1997.
- [18] STURM, P.; MAYBANK, S. On plane-based camera calibration: A general algorithm, singularities, applications. Conference on Computer Vision and Pattern Recognition, pp. 432–437, 1999.
- [19] BROWN, D. Close range camera calibration. Photogrammetric Engineering, vol. 37, no. 8, pp. 855–866, 1971.
- [20] FRYER, J.; GOODIN, D. In-flight aerial camera calibration from photography of linear features. Photogrammetric Engineering and Remote Sensing, vol. 55, no. 12, pp. 1751–1754, 1989.
- [21] DELLAERT, F.; SEITZ, S. M. e THORPE, C. E. Structure from Motion Without Correspondences. Proc. Computer Vision and Pattern Recognition Conf. (CVPR) , pp 327-337, 2000
- [22] BIRCHFIELD, S. An Introduction to Projective Geometry for Computer Vision. Nota não-publicada. Stanford University, 1998.
- [23] BOUGUET, J. Camera Calibration Toolbox for Matlab, URL: http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/ref.html. Acesso em 20 de setembro de 2006.
- [24] JENSEN, H. W. Realistic image synthesis using photon mapping. Primeira Edição, A.K. Peters. Massachusetts, 2001, 200 pág.
- [25] SHIRLEY, P. e MORLEY, K., Realistic Ray Tracing. Segunda Edição, A.K. Peters. Massachusetts, 2003, 235 pág.
- [26] CURLESS, B. e LEVOY, M. Better optical triangulation through spacetime analysis. Proc. 5th Int. Conf. Computer Vision, Boston, USA, pp 987-993, 1995.
- [27] SPILLMAN, L.; WERNER, J.S. Visual Perception - The Neurophysiological Foundations, Academic Press, 1990, 198 p.
- [28] WANDELL, B. Foundations of Vision, Sinauer Associates, Inc., Sunderland, Massachusetts, 1995, 245 p.
- [29] HUBEL, D. H. Eye, Brain and Vision, Scientific American Library, New York, 1995, 95 p.
- [30] PHOTOSYNTH. Microsoft Live Labs. URL: <http://labs.live.com/photosynth>. Acesso em 20 de setembro de 2006.
- [31] POLLEFEYS, M. Self-calibration and metric 3D reconstruction from uncalibrated image sequences. KU Leuven, 1999, Ph. D. Thesis
- [32] POLLEFEYS, M.; KOCH, R.; VERGAUWEN, M.; VAN GOOL, L. *Automatic Generation of 3D Models from Photographs*. Proceedings Virtual Systems and Multimedia. pp-232-240, 1998.
- [33] TROBINA, M., *Error model of a coded-light range sensor*. Technical Report BIWI-TR-164, ETH-Zentrum, 1995, 76 p.

- [34] WANG, Y.F. *Characterizing three-dimensional surface structures from visual images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(1):52-60, 1991.
- [35] KLETTE, R.; SCHLUNS, K. e KOSCHAN, A. *Computer Vision : Three-Dimensional Data from Images*, Springer Verlag, 1998, 392 p.
- [36] HORN, B. *Robot vision*, MIT press, Massachusets, 1986, 480 p.
- [37] SUNDARAM, H.; NAYAR, S. *Are textureless scenes recoverable?* Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 814-820, 1997.
- [38] CRIMINISI, A. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. Dept. Engineering Science University of Oxford, Springer, 1997, 184 p.

Judith Kelner

Eduardo Lourenço Apolinário