

2 Trabalhos Relacionados

Nesta seção são expostos os principais trabalhos utilizados como base para esta pesquisa. O objetivo é introduzir os conceitos que foram desenvolvidos nesses trabalhos, usados como referência, e expor os principais assuntos que se relacionam diretamente com esta dissertação.

2.1 Calibração de Câmera

Uma câmera calibrada é aquela na qual se pode relacionar um ponto qualquer em seu plano de projeção ao respectivo raio que liga seu centro óptico a esse mesmo ponto [Hartley and Zisserman, 2003]. Matematicamente, isso equivale a encontrar uma matriz K de calibração que representa uma transformação entre esse ponto x e a direção do raio $\mathbf{d} = K^{-1}x$.

A calibração consiste na recuperação de parâmetros internos da câmera — chamados parâmetros *intrínsecos* — que são usados para gerar tal matriz K . Esses parâmetros não medem propriedades relativas à localização ou à orientação da câmera, já que essas são mapeadas por parâmetros ditos *extrínsecos*. Alguns dos métodos usados para recuperá-los serão analisados na Seção 2.2 deste mesmo capítulo.

Existem muitos trabalhos que tratam do problema de calibração da câmera a partir de uma ou mais fotos. Em [Hartley and Zisserman, 2003] são introduzidos conceitos fundamentais da geometria projetiva que permitem a criação de certos dispositivos simples para a obtenção da matriz de calibração. Por exemplo, através de homografias obtidas entre os cantos (quinas) de três quadrados não coplanares no espaço da cena, e a projeção desses pontos no espaço da imagem, pode-se obter restrições algébricas que permitem que se calcule K .

Os trabalhos propostos por Zhengyou Zhang e Roger Tsai em [Zhang, 1998] e [Tsai, 1986], respectivamente, tiveram grande repercussão na área de visão computacional por sua simplicidade e eficiência. Tais trabalhos apresentam métodos para calibrar uma câmera utilizando padrões planares e tridimensionais e também propõem maneiras de estimar a rotação e a

translação da câmera em relação a esses padrões.

Ao relacionar pontos do mundo com pontos da imagem, é preciso considerar que a imagem pode não ser precisa em relação a uma projeção geométrica ideal dos pontos do mundo no plano de projeção da câmera. Isso ocorre porque as imagens podem conter distorções e cisalhamentos oriundos dos dispositivos eletrônicos que a geraram. Para resolver esse problema pode-se incluir parâmetros de distorção radial no cálculo de calibração para estimar tal anomalia e calibrar a camera de forma mais precisa. Roger Tsai aborda esse problema em [Tsai, 1986] e fornece um método para solucioná-lo de forma integrada com o processo de calibração.

O processo de correção da distorção radial é feito, em alguns casos, em um passo anterior ao da calibração. Em [Kang, 2000; Pajdla et al., 1997] são propostas técnicas para resolver esse problema sem o conhecimento da estrutura tridimensional da cena ou dos parâmetros da câmera. O trabalho aqui desenvolvido, no entanto, não aborda o problema da distorção radial e assume que o erro de distorção causado pelos dispositivos de captura de imagem é baixo, já que a magnitude de tal distorção é baixa em câmeras digitais de boa qualidade. Apenas em dispositivos simples como em *webcams*, a magnitude de distorção é alta, pois suas lentes são simples e seus dispositivos eletrônicos têm pouca qualidade.

Em [Luong and Faugeras, 1999], Olivier Faugeras apresenta muitos conceitos importantes relacionados ao problema da calibração de câmera e da reconstrução tridimensional que são considerados centrais na visão computacional. Nesse trabalho Faugeras adota uma abordagem que envolve o cálculo de homografias entre múltiplas imagens, tiradas de uma mesma cena em posições distintas, para recuperar a matriz fundamental e estimar tanto a estrutura da cena como os parâmetros internos da câmera. Esse processo não requer o conhecimento prévio da estrutura da cena, mas exige que uma grande quantidade de pontos de associação entre as imagens seja fornecida. Embora existam algoritmos especializados em detectar automaticamente esses pontos relacionados em diferentes imagens, como o conceituado detector de características relacionadas **SIFT** (*Scale Invariant Feature Transform*) proposto e patenteado por David G. Lowe [Lowe, 2004], essa tarefa geralmente envolve grande complexidade computacional.

Algumas das abordagens de calibração mencionadas exigem que se tenha em mão muitos pontos da imagem cuja respectiva posição no mundo é conhecida. Em [Zhang, 1998] por exemplo, são necessários no mínimo dois grupos de oito pontos da imagem associados a um padrão planar. Esses grupos de pontos são obtidos de fotos tiradas de um padrão de calibração em diferentes

posições. Em alguns casos isso pode não representar uma desvantagem, uma vez que os padrões utilizados são simples e de fácil rastreamento na imagem. Porém, em casos onde as imagens de entrada possuem uma grande quantidade de ruído e o padrão de calibração utilizado tem geometria complexa, não é simples obter um conjunto grande de pontos na imagem que tenham posição conhecida no mundo. Considerando-se que este trabalho trata justamente deste tipo de situação, optou-se aqui por um método que permite que se tenha pouco conhecimento a respeito da geometria da cena. Esse método é o método de calibração por pontos fuga que será analisado a seguir.

2.1.1

Calibração por Pontos de Fuga

Pontos de fuga são mapeamentos projetivos de pontos no infinito para coordenadas da imagem. Na geometria projetiva, pontos do mundo que estão muito distantes, chamados pontos no infinito, são projetados em uma posição finita do plano de projeção da câmera. Esses pontos marcam direções no mundo em relação à orientação da câmera, e não dependem de sua posição, apenas de sua rotação em relação à cena [Hartley and Zisserman, 2003]. A Figura 2.1 ilustra o processo de formação de pontos de fuga.

Por representarem direções da cena, esses pontos fornecem restrições fortes em relação à orientação e à geometria da câmera (parâmetros intrínsecos). Essas restrições permitem calcular a matriz de calibração K e uma matriz de rotação R , quando se assume que não há deformação ou distorção radial na imagem utilizada [Cipolla et al., 1999b; Cipolla et al., 1999a].

O processo de calibração por pontos de fuga possui, entretanto, uma limitação, pois seu funcionamento é restrito às imagens em que pontos de fuga podem ser detectados. Para tanto, em tais imagens devem existir segmentos que correspondam a três direções ortogonais conhecidas da cena — frequentemente as direções principais x , y e z . A Figura 2.2 ilustra pares de segmentos paralelos a cada uma das direções principais.

Se esses segmentos não forem visíveis ou passíveis de serem detectados na imagem, os pontos de fuga não podem ser calculados de forma simples. Isso porque é necessário que haja três pontos de fuga para calcular com precisão a matriz K de calibração e cada um deles é obtido a partir de segmentos paralelos a uma das três direções ortogonais. Como em geral, as direções principais x , y e z são mais bem definidas e mais facilmente detectadas em uma cena (por exemplo, arestas das interseções entre paredes de um quarto, arestas da fachada de uma edificação, etc), elas são naturalmente usadas para encontrar os pontos de fuga. Isso significa que as três dimensões de objetos devem aparecer

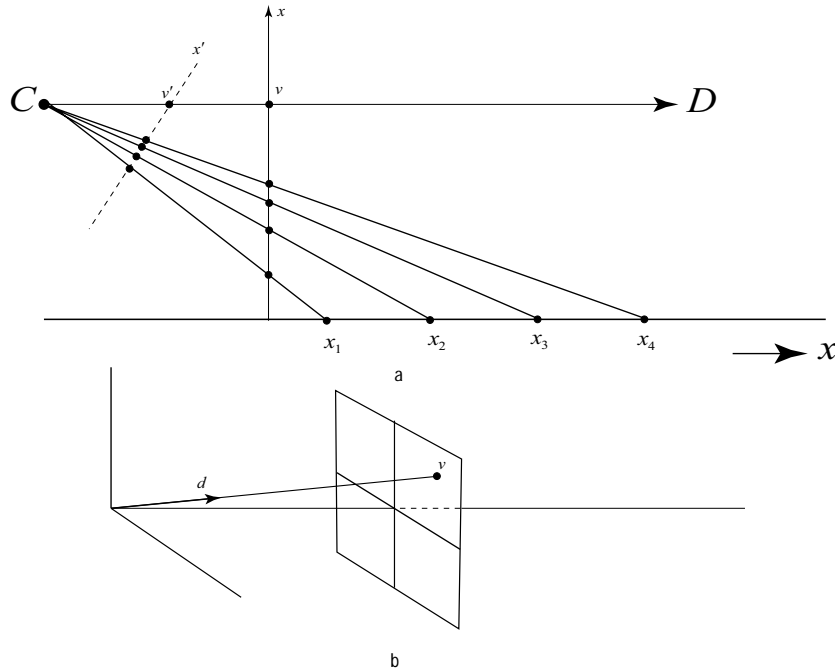


Figura 2.1: Processo de formação de pontos de fuga. A figura **a** ilustra que pontos igualmente espaçados no mundo são projetados para pontos cada vez mais próximos na imagem. Um ponto muito distante na cena é projetado em um ponto chamado ponto de fuga, representado pelo ponto v na reta x ou v' no plano de projeção da câmera. A figura **b** mostra que um ponto de fuga corresponde a uma direção no mundo construída a partir de um vetor que vai do centro de projeção da câmera até esse ponto na imagem.

na imagem de entrada — no caso de objetos tridimensionais — para que seja possível calibrar a câmera usando este método. Quando apenas dois pontos de fuga são detectados na imagem, ainda é possível estimar a matriz de calibração utilizando restrições algébricas expostas em [Rother, 1999]. Contudo, não é possível calibrar a câmera a partir de apenas um único ponto de fuga. Nesta dissertação assume-se que todas as fotos de entrada exibem segmentos nas três direções principais do objeto de referência e que, portanto, sempre é possível calibrar por pontos de fuga.

Em [Rother, 1999; Schaffalitzky and Zisserman, 2000; McLean and Kotturi, 1995] e [Gamba et al., 1996] são desenvolvidas técnicas para detecção automática de pontos de fuga, mas funcionam apenas em casos em que há pouco ruído e em que há bons segmentos de reta detectáveis na imagem. Além disso, a probabilidade de falso positivo é grande, ou seja, freqüentemente segmentos que não pertencem a nenhuma das três direções são detectados como tal, já que retas perpendiculares do mundo podem ser projetados quase de forma paralela na imagem.

Nesta dissertação a técnica de calibração por pontos de fuga é utilizada

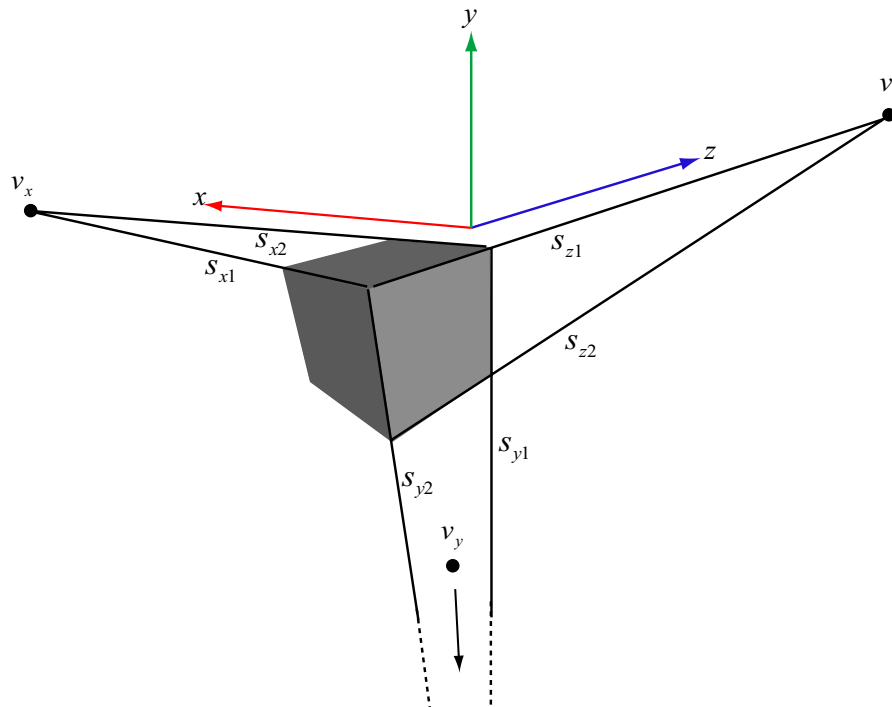


Figura 2.2: A figura ilustra marcações de segmentos no espaço da imagem ($s_{x1}, s_{x2}, s_{y1}, s_{y2}, s_{z1}, s_{z2}$) que correspondem a segmentos paralelos às direções principais da cena (x, y, z) (espaço do mundo). Para cada direção existem dois segmentos cuja interseção é exatamente o ponto de fuga.

de forma manual. Para tanto, segmentos correspondentes a direções conhecidas do mundo são marcados pelo usuário na foto. A posição dos pontos é então calculada a partir da interseção de dois desses segmentos por direção principal da cena, totalizando seis segmentos de reta necessários, no mínimo.

Devido à possibilidade de grande imprecisão nas marcações, provocada por ruído nas imagens ou falha do usuário, o cálculo da posição dos pontos de fuga pode se tornar mais exato se for utilizado um conjunto maior de segmentos por direção, ao invés de apenas dois. Isso faz com que o erro se distribua entre esses segmentos marcados, tornando mais precisa a posição do ponto de fuga obtido, mas exige também que tal posição seja calculada através de um processo mais complexo, envolvendo técnicas de minimização. Isso porque a interseção entre segmentos — tomados aqui como retas — se dá de forma unívoca quando esses são tomados dois a dois, mas o objetivo, no entanto, é encontrar o ponto que representa a interseção de todas as retas do conjunto simultaneamente. Para tanto, muitos trabalhos calculam um ponto equidistante a todas as retas como sendo o ponto de fuga, como é feito em [Alvarez, 2002]. No entanto, [Hartley and Zisserman, 2003] afirma que o melhor método para calcular os pontos de fuga a partir de múltiplos segmentos de

reta é utilizar uma parametrização que envolve retas que se interceptam em um único ponto (Seção 3.1.4).

2.1.2

Calibração a partir de EXIF tags

Câmeras digitais disponibilizam meta-dados que compõem o padrão EXIF — EXchangeable Image File — e são armazenadas em um cabeçalho próprio dentro de alguns arquivos de imagens digitais. Esses dados podem ser encontrados em arquivos codificados em formatos como JPEG¹ e TIFF². Dentre as informações providas pelas tags EXIF se encontram algumas que são de maior interesse para o problema de reconstrução de câmera, pois permitem que alguns parâmetros internos da câmera sejam calculados diretamente, com pouco esforço computacional.

A partir da leitura de dados específicos dentro do cabeçalho de dados EXIF, é possível obter diretamente dados que permitem a reconstrução da geometria interna de uma câmera, tais como sua distância focal e altura do plano de projeção. Em [Snavely et al., 2006], as distâncias focais de câmeras de um conjunto denso de fotos são recuperadas diretamente dessas tags para serem usadas no processo de calibração e de reconstrução de uma cena 3D. Tal trabalho deu origem ao sistema PhotoSynth³ criado pela Microsoft® que tem como objetivo organizar de forma tridimensional grandes conjuntos de fotos.

Contudo, algumas câmeras antigas ou muito simples não dão suporte ao padrão EXIF ou não o fazem de forma completa. Tais cameras não permitem que se recupere as informações necessárias para se obter parâmetros intrínsecos. No entanto, muitas câmeras atuais suportam essas tags, o que permite o cálculo da geometria da câmera de forma automatizada e a eliminação da necessidade de se utilizar outro método mais complexo, como o de calibração por pontos de fuga. O trabalho desenvolvido por esta pesquisa utiliza preferencialmente o método de reconstrução de câmera utilizando tags EXIF. Somente quando estas não estão disponíveis na foto de entrada é que o método de calibração por pontos de fuga é utilizado.

¹O padrão JPEG (Joint Photographic Experts Group) para codificação de imagens pode ser encontrado em <http://www.faqs.org/rfcs/rfc2035.html>.

²Padrão TIFF (Tagged Image File Format) para codificação de imagens pode ser encontrado em <http://www.faqs.org/rfcs/rfc2302.html>.

³O *site* oficial da aplicação PhotoSynth é <http://labs.live.com/photosynth/>.

2.2

Posicionamento de Câmera

Como mencionado, o processo de reconstrução de câmera pode ser dividido em duas etapas: a reconstrução dos parâmetros internos e a dos parâmetros externos. De fato, a função de projeção P pode ser decomposta em $P = KRT$, onde K é a matriz de parâmetros intrínsecos e RT é uma transformação de corpo rígido composta por uma translação seguida de uma rotação [Hartley and Zisserman, 2003]. Assim, a função P pode claramente ser dividida em duas transformações distintas, uma que depende apenas dos parâmetros internos da câmera, representada pela matriz K e outra que é função de seus parâmetros externos, representadas pela transformação RT . O processo de obtenção dos parâmetros internos já foi descrito na seção anterior e esta seção dedica-se a analisar métodos para calcular a posição e a orientação da câmera em relação à origem de algum objeto conhecido da cena.

Para calcular rotação e posição, é comum utilizar câmeras já calibradas, ou seja, câmeras cujos parâmetros internos já foram reconstruídos. Isso porque, como o mapeamento de pontos do mundo em pontos da imagem pode ser decomposto em uma transformação linear homogênea e uma transformação projetiva, torna-se mais simples recuperar apenas os parâmetros da primeira transformação quando a segunda é constante. Quando se possui um conjunto suficientemente grande de associações entre pontos da imagem e pontos do mundo e a câmera calibrada, pode-se estimar diretamente a rotação e a translação da câmera em relação à cena através de um sistema linear. A translação obtida é, no entanto, função de um fator de escala, que só pode ser calculado através de alguma relação métrica conhecida entre o modelo computacional de referência e o modelo real.

Quando se calibra através de pontos de fuga, é possível calcular, além dos parâmetros internos, a orientação da câmera em relação ao objeto de referência, porque os pontos de fuga medem direções que não dependem da translação da câmera (vide Seção 3.1.4). Essa rotação pode ser usada juntamente com a matriz de calibração K para se recuperar uma translação em relação a algum ponto da imagem adotado como sendo a projeção do ponto-origem da cena [Cipolla et al., 1999b; Cipolla et al., 1999a]. Tal translação, no entanto, também é uma função de um fator de escala desconhecido. A Seção 3.1.5 apresenta um método desenvolvido por esta dissertação para recuperar a translação completa de uma câmera calibrada — tendo em mãos sua rotação — utilizando apenas dois pontos conhecidos na imagem.

Embora a calibração por pontos de fuga forneça uma boa estimativa da orientação da câmera, essa estimativa não é usada neste trabalho. Portanto,

apenas os parâmetros internos da câmera são calculados utilizando-se essa técnica. Para encontrar tanto a rotação quanto a translação de câmera, utilizou-se outro método proposto inicialmente em [Taylor and Kriegman, 1995]. Esse método sofreu adaptações para funcionar com uma única imagem por vez, já que ele é construído originalmente para funcionar com múltiplas imagens simultaneamente.

2.3 Sistemas Fotogramétricos

Muitas das idéias propostas nesta dissertação foram inspiradas basicamente em dois sistemas fotogramétricos: Photosynth da Microsoft, que teve origem no trabalho [Snavely et al., 2006] de Noah Snavely, e no Façade, criado e proposto em [Debevec, 1996] por Paul E. Debevec.

Em [Snavely et al., 2006], Snavely utiliza um conjunto denso de fotos para recuperar nuvens de pontos tridimensionais e modelos de câmeras dessas fotos utilizando apenas correspondências entre elas, calculadas com os algoritmos RANSAC ([Fischer and Bolles, 1981]) e SIFT [Lowe, 2004]. A partir das correspondências encontradas, Snavely et al utiliza geometria epipolar e computa uma matriz fundamental para recuperar pontos do mundo, que é refinada de acordo com o número de correspondências encontradas. Embora seja uma técnica bem-sucedida e cujo resultado é bastante eficaz, principalmente em relação à recuperação de câmeras, ela possui uma condição de funcionamento que não é desejada neste trabalho: por ser totalmente automatizada e utilizar algoritmos computacionalmente complexos, demanda longo tempo de processamento. De fato, o sistema Photosynth pode levar dias processando as imagens de entrada. Além disso, tal sistema necessita fundamentalmente de um conjunto denso de fotos para obter sucesso, devido à própria natureza de seu funcionamento.

Tais fatos vão de encontro com alguns requisitos deste trabalho, como a capacidade de se trabalhar com um conjunto esparsa de imagens e ter eficiência no tempo de processamento. Além disso, esta dissertação assume que a geometria do modelo é conhecida, o que caracteriza uma diferença substancial em relação ao objetivo do referido trabalho de Snavely. Tal assunção permite simplificar o processo de reconstrução das câmeras. No entanto, apesar das diferentes abordagens, certos conceitos da aplicação final de Snavely foram utilizados aqui para dar origem à aplicação de exemplo, proposta no Capítulo 4. Algumas das principais funcionalidades tomadas como inspiração foram o sistema de navegação direcional entre fotos, o sistema de calibração utilizando EXIF tags e o conceito geral de álbum de fotos tridimensional.

Já o trabalho realizado por Paul Debevec, adota uma abordagem semi-automática em que um conjunto esparsa de fotos é utilizado e a geometria do modelo é parcialmente conhecida. Tal trabalho também trata do problema da reconstrução de câmara, mas o faz com o objetivo de recuperar proporções de um modelo paramétrico previamente criado por um usuário. O sistema Façade criado por Debevec, deu origem a outros sistemas comerciais e é utilizado com sucesso para recuperar proporções de modelos e recuperar câmeras (previamente calibradas) com eficácia.

A técnica proposta por Debevec foi implementada inicialmente por esta dissertação, mas se concluiu que ela demandava intensa interação do usuário, o que diverge dos objetivos aqui traçados. Em tal abordagem, o usuário é responsável por criar um modelo paramétrico e associar a ele diversas marcações em diversas imagens, o que caracteriza um trabalho potencialmente longo e complexo. Além disso, o resultado final desta técnica é tão bom quanto a qualidade do modelo paramétrico criado e das marcações feitas nas imagens. No entanto, Debevec utiliza tal técnica com o objetivo principal de reconstruir proporções de um modelo rústico que é usado posteriormente como entrada para outras técnicas integradas, tal como mapeamento de textura dependente de ponto de vista (*View-Dependent Texture Mapping*), geração de mapas de profundidade usando o modelo como restrição e outras técnicas que aumentam a qualidade da renderização final, mas fogem ao escopo desta dissertação.

Em sua tese, Paul Debevec baseia-se principalmente em [Taylor and Kriegman, 1995] para reconstruir a estrutura tridimensional de uma cena a partir de múltiplas fotos. Ele propõe também um novo método para estimar uma rotação inicial de câmara usando direções conhecidas na cena. Foi a partir das idéias para reconstrução de câmara de Debevec e de sua principal referência, [Taylor and Kriegman, 1995], que teve origem a técnica de ajuste de câmara e de casamento entre modelos e imagens proposta nesta dissertação.

Em suma, a aplicação de exemplo (Capítulo 4) e as diversas técnicas que compõe o método proposto por esta dissertação se baseiam principalmente nos trabalhos de Debevec, Kriegman e Snavely. Contudo, a abordagem aqui utilizada trata apenas uma única foto por vez e assume dispor integralmente da geometria do modelo.