

2 Múltipla calibração de câmeras

Neste capítulo descreveremos alguns conceitos relacionados ao processo de múltipla calibração de câmeras. Para organizar esses conceitos, propomos um framework conceitual que ajudará a esquematizar alguns aspectos e etapas comumente encontrados em diversos métodos de calibração, tanto para o caso de uma câmera como para o de múltiplas câmeras.

Para modelar o funcionamento de uma câmera, usaremos o clássico modelo de câmera Pinhole. No apêndice A, encontra-se a descrição detalhada dos diversos componentes e parâmetros usados para modelar teoricamente o processo de captura óptica, que é realizado internamente numa câmera real.

Define-se como calibração de câmera o processo realizado para calcular o valor dos parâmetros intrínsecos, extrínsecos e coeficientes de distorção da lente, que são comumente usados para modelar o funcionamento de uma câmera real. Logo, considera-se como múltipla calibração de câmeras o processo que permite a calibração de várias câmeras simultaneamente.

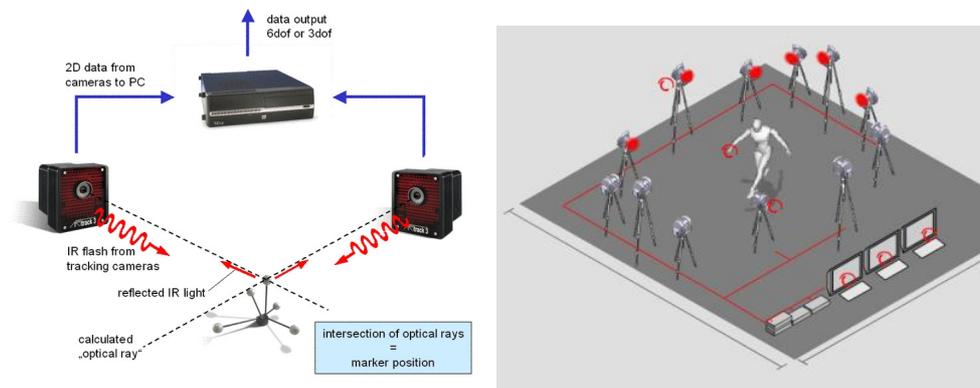


Figura 3: Sistema estéreo e de múltiplas câmeras.

(Fontes: <http://www.ar-tracking.de/General.144.0.html> e <http://www.digital-humans.org/Report2004/Documents/06-VirtualRealityForSantos.htm>)

O menor número de câmeras necessário para um sistema ser considerado de múltipla calibração é dois. Esse tipo de sistema é mais conhecido como sistema de câmeras em estéreo. Já o sistema composto por mais de duas câmeras ganha o nome de sistema de múltiplas câmeras (Figura 3).

Diferentes métodos têm sido definidos para afrontar esse tipo de problema, mas analisando a maioria deles, podemos estabelecer um framework conceitual que ajudará a descrever um conjunto de conceitos usado para resolver o problema específico da múltipla calibração de câmeras. Esse framework conceitual é composto de três etapas, as quais também têm sido definidas num estudo realizado sobre diferentes métodos de calibração de câmera individual, descrito no trabalho de Marques [19]. Nós adaptaremos e estenderemos a definição das três etapas, para um cenário onde analisaremos métodos de múltipla calibração de câmeras. Finalmente as etapas que compõem nosso framework conceitual são as seguintes:

1. Aquisição de dados para a calibração do sistema de múltiplas câmeras.
2. Calibração inicial, correspondente ao cálculo inicial dos parâmetros que usaremos para modelar as câmeras reais.
3. Processo de otimização dos valores calculados inicialmente para os parâmetros de calibração das câmeras.

A seguir, faremos uma descrição mais detalhada dos conceitos e processos de cada etapa do nosso framework conceitual.

2.1. Primeira etapa: aquisição de dados para calibração

A primeira etapa do processo de múltipla calibração envolve diferentes métodos e técnicas utilizados na captura de dados de entrada que serão usados no posterior cálculo dos parâmetros que modelam teoricamente o processo de calibração das câmeras.

Esses métodos e técnicas são comumente compostos por diferentes algoritmos da área de processamento de imagem e reconhecimento de padrões, que, em conjunto com o uso e a definição de certos tipos de padrões sintéticos, ajudam a realizar a captura e o reconhecimento preciso de diferentes pontos de referência, conhecidos como marcadores, que são definidos pela estrutura de cada padrão.

Os algoritmos de processamento de imagem são utilizados com o objetivo de ressaltar características próprias do padrão, o que permite isolá-lo do resto de objetos presentes na mesma imagem. Depois, os algoritmos de reconhecimento de

padrão se encarregam de extrair e identificar individualmente os marcadores que um padrão contém.

O objetivo da utilização de um padrão é a captura de uma amostra relativamente grande de pontos de referência corretamente reconhecidos e distribuídos na maior parte do espaço de visão das câmeras, definido como espaço de rastreamento.

O formato dos diferentes tipos de padrões propostos na bibliografia sobre calibração de câmeras tenta cumprir alguns requisitos:

- O formato proposto deve estar definido por uma estrutura que permite sua nítida diferenciação em relação a outros objetos semelhantes que possam estar presentes no mesmo espaço de rastreamento. Isso ajuda a evitar possíveis erros, como falsos positivos no reconhecimento do padrão.
- O formato também deve permitir a sua fácil manipulação e movimentação no espaço de rastreamento. Isso permitirá que o campo visual enxergado pelas câmeras esteja corretamente preenchido pela captura de várias vistas válidas do padrão, o que representará um maior número de pontos de referência corretamente identificados e espalhados tanto no campo visual das câmeras como no espaço de rastreamento físico.
- A quantidade e a distribuição dos pontos de referência contidos no padrão devem permitir a recuperação de múltiplas informações em relação aos mesmos, como a identificação individual de cada ponto de referência em relação ao resto dos pontos, a identificação grupal de todos os pontos de referência que determinam o formato único do padrão, além de distâncias, ângulos e outras medidas que possam ser recuperadas a partir do conhecimento prévio da distribuição tridimensional desses pontos de referência que formam o padrão.

À medida que o formato de um padrão passa a conter muitos pontos de referência e detalhes na sua estrutura, também se encarece diretamente o processo de reconhecimento do padrão e, algumas vezes, a sua manipulação e movimentação em frente às câmeras torna-se limitada.

A falta de liberdade para manipular e movimentar o padrão diante das diferentes câmeras diminui a quantidade de vistas válidas do padrão capturadas

simultaneamente no campo visual compartilhado pelas várias câmeras que compõem o sistema. A observação acima será melhor exemplificada e entendida na descrição sobre diferentes tipos de padrão na seção seguinte.

2.1.1. Padrões usados na calibração de câmeras

Os padrões de calibração têm como objetivo permitir a captura de pontos de referência a partir da detecção e do reconhecimento de algumas características implícitas do seu formato.

A amostra de pontos de referência extraída a partir da análise e do processamento de vistas capturadas do padrão é usada no processo de calibração tanto na etapa de calibração inicial como numa etapa final de otimização do valor dos parâmetros obtidos pela calibração inicial.

A captura de uma amostra de pontos de referência tem sido a principal ferramenta usada no processo de calibração de câmeras. O uso de padrões contendo esses pontos de referência tem sido estudado desde os métodos desenvolvidos e apresentados por [11][35][36] e [40], onde o padrão de calibração usado tem um formato planar e contém na sua superfície pontos de referência definidos por marcadores explícitos ou implícitos desenhados e/ou inseridos na superfície do padrão.

Outros padrões propostos para casos de estéreo e múltipla calibração de câmeras surgiram na última década, dentro de um grupo considerado como padrões 1D e muitas vezes usados em abordagens de auto-calibração de câmera descrita em trabalhos apresentados por [4] e [32]. Esses recentes trabalhos mostram a nova tendência de auto-calibração de câmeras, e têm como característica em comum a proposta de padrões mais simples, compostos por um ou dois marcadores, que trazem como principal vantagem a fácil manipulação, movimentação e reconhecimento nas diversas vistas capturadas do padrão dentro do espaço de rastreamento.

Também podemos citar outro tipo de padrão conhecido como padrão tridimensional composto por um conjunto de dois ou três planos colocados perpendicularmente um em relação ao outro, este tipo de padrão é pouco utilizado para a calibração de múltiplas câmeras principalmente pelos problemas de oclusão que pode gerar os planos que o conformam quando mais de uma câmera tenta

enxergar simultaneamente os marcadores inseridos na sua estrutura, por esse motivo não é considerado na nossa análise sobre múltipla calibração de câmera.

Tanto os padrões planares como os novos padrões 1D serão descritos a seguir, quando detalharemos um pouco mais as vantagens e desvantagens dos mesmos, tendo sempre como foco a calibração feita em duas ou mais câmeras simultaneamente.

2.1.2. Padrão planar – 2D

Dentro dessa tipificação de padrão, iremos considerar como padrão planar aquele cujos pontos de referência se encontram distribuídos sobre uma superfície plana 2D. Trabalhos apresentados por [11][35][36] e [40] mostram variações desse tipo de padrão, como os apresentados na Figura 4.

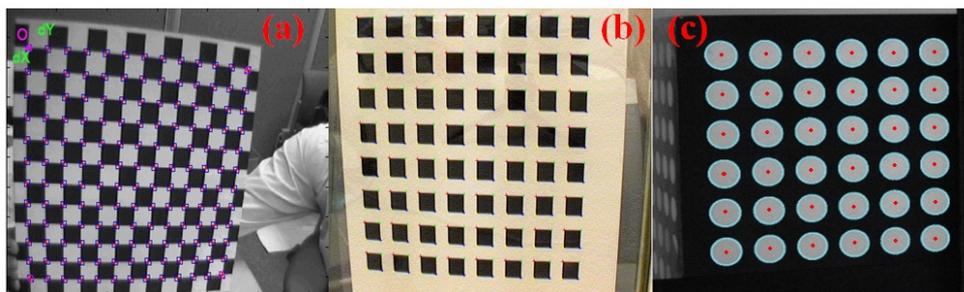


Figura 4: Padrões planares usados para calibração de câmera em trabalhos como, (a) Bouquet [5], (b) Zhang [40] e Tsai [35][36] e (c) Heikkila [11].

Como podemos observar, o aspecto dos padrões que define os pontos de referência varia em cada um dos diferentes trabalhos. Por exemplo, no trabalho de [35][36] e [40], o padrão é definido como um conjunto de quadrados pretos alinhados em um determinado número de filas e colunas, e os pontos de referência são determinados pelos cantos dos quadrados pretos impressos sobre um fundo branco. Essa arquitetura permite a captura de 4 pontos de referência por cada quadrado corretamente reconhecido. Já no trabalho de [11], a forma dos desenhos no padrão muda de um quadrado preto para círculos brancos num fundo preto, e os pontos de referência são definidos pelos centróides dos círculos brancos. Isso permite a captura de um ponto de referência por cada área circular corretamente reconhecida.

Ambos modelos de padrão planar compartilham a característica do formato em grade ou grelha, que permite a captura de um número de pontos de referências abundante por vista de padrão corretamente reconhecido e correlacionado.

Esse tipo de padrão planar também já foi utilizado na calibração de sistemas estéreo e de múltipla calibração de câmeras, como descrito em trabalhos realizados por [5] (estéreo) e [25] (múltiplas câmeras). Na figura 5, mostramos o tipo de padrão usado em cada trabalho.

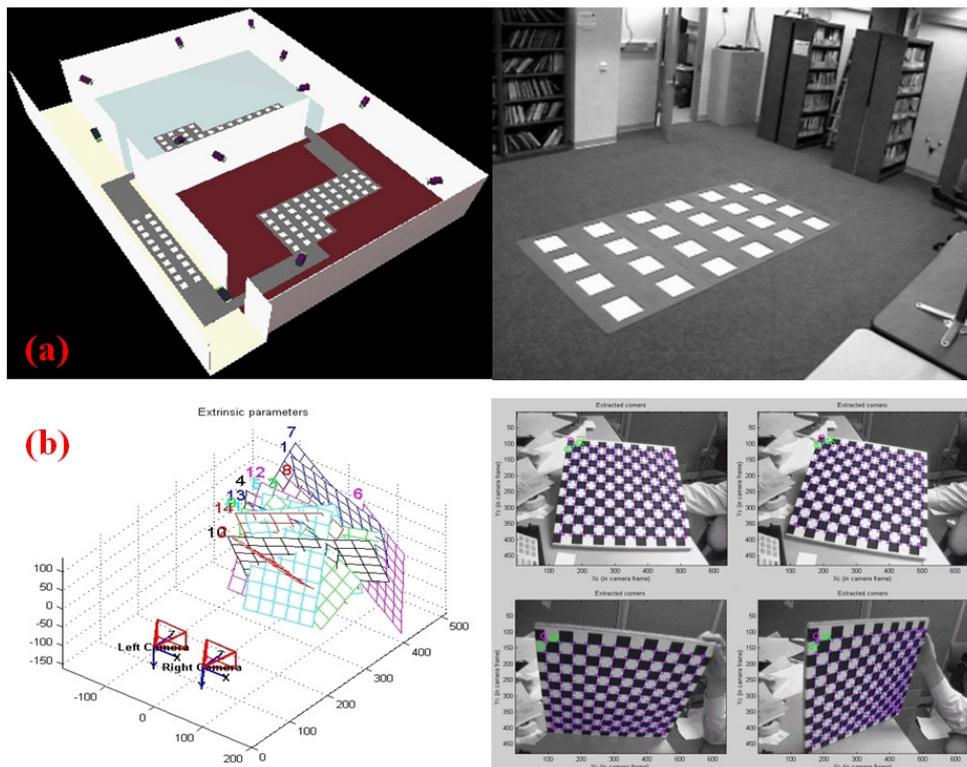


Figura 5: Padrões planares usados em (a) Olsen e Hoover [25] e (b) Bouquet [5].

Apesar de esse tipo de padrão ser muito usado e conter muita informação em relação aos pontos de referência que podem ser recuperados por cada vista do padrão corretamente reconhecido, ele tem como principal desvantagem a complexidade do processamento da imagem e do reconhecimento do padrão requerido para poder ser identificado corretamente. Esse problema surge por causa de limitações em relação à distância e diferentes orientações em que uma vista do padrão pode ser capturada, limitando sua livre movimentação diante de uma ou várias câmeras do sistema.

Algumas restrições sobre a movimentação desse tipo de padrão já foram reconhecidas na implementação dos métodos de calibração individual. Tratam-se

principalmente de restrições na orientação do padrão planar que, em ângulos menores que 45° e distâncias muito longas da câmera, já apresentava perda no reconhecimento acertado. Essas restrições ficaram ainda mais notórias e restritivas quando esse tipo de padrão foi estendido para uso em sistemas com mais de uma câmera. Nesse novo ambiente de calibração, a interseção dos campos de visão das câmeras limitou muito mais as orientações mínimas nas quais o padrão pode ficar quando ele está sendo capturado pelas câmeras.

O trabalho de [5] mostra uma extensão no uso desse tipo de padrão para a calibração de um sistema estéreo. Nesse ambiente com duas câmeras, as limitações na movimentação e na orientação máxima de 45° do padrão ficam ainda mais notórias quando o ângulo entre as câmeras fica menor que 180° (Figura 5b). Já no trabalho de [25], o uso do padrão planar tenta contornar os problemas de movimentação com a geração de um padrão fracionado e de grandes dimensões que lhe permitem ser visível por várias câmeras bem distantes e com pouca interseção entre os campos de visão das câmeras (Figura 5a).

Em relação à abordagem apresentada por [5], essa última tem como principal diferença o número de pontos de referência que podem ser capturados. O método definido por [5] permite a captura de uma amostra muito maior de pontos de referência que o de [25]; em contrapartida, o método de [25] tenta quebrar a restrição do posicionamento e do uso de só duas câmeras imposta no trabalho de [5], e ampliá-la para suportar múltiplas câmeras distribuídas em um ambiente maior de rastreamento.

Por fim, as abordagens dos métodos acima têm como desvantagem comum a complexidade na aquisição dos pontos de referência. Devido a essa limitação foram criados novos tipos de padrões. A seguir serão descritas estas novas propostas de padrão com formato 1D e adimensional.

2.1.3. Padrão unidimensional – 1D

O padrão considerado unidimensional (1D) foi proposto e descrito inicialmente no trabalho publicado por [4]. Ele é composto por dois marcadores esféricos colocados nos extremos de uma vara metálica (Figura 6). O trabalho de [4] propôs esse padrão como parte de um novo método de calibração de um

sistema de câmeras em estéreo. Esse método pode ser considerado já um método de auto-calibração, cujo conceito é descrito na seção 2.2.2.4.

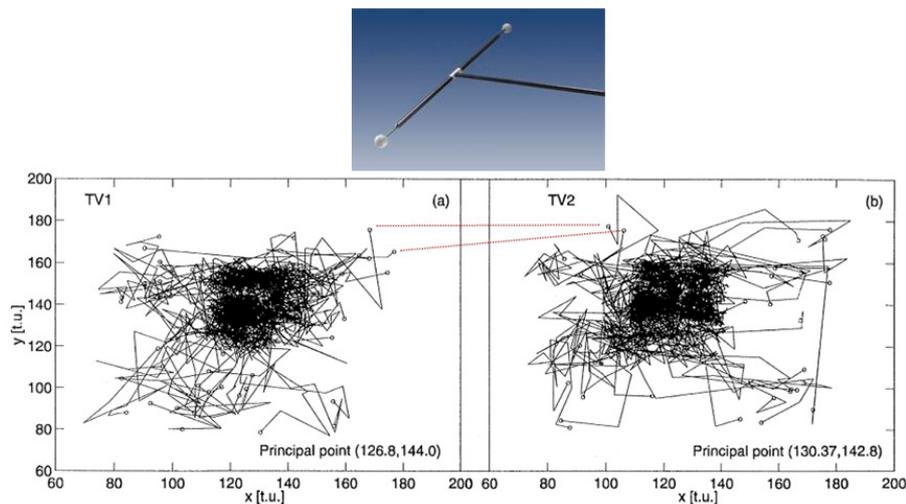


Figura 6: Padrão de calibração e captura proposto no trabalho de Borghese[4].

Uns anos mais tarde esse método foi estendido e usado nos trabalhos de [14] [22] e [37] para a calibração de um sistema formado por múltiplas câmeras, no qual se manteve o mesmo conceito do padrão 1D e só se mudou o procedimento de captura e reconhecimento.

Esse tipo de padrão 1D tem sido também muito utilizado como padrão de calibração em sistemas de rastreamento óptico comerciais como [1] e [39]. As principais vantagens desse tipo de padrão traz podem ser resumidas em:

- Liberdade na movimentação e manipulação do padrão.
- Fácil captura e visualização pelas diferentes câmeras dos sistemas
- Nenhuma restrição enquanto ao posicionamento das câmeras em relação a certas posições ou orientação do padrão.
- O processamento requerido para a captura e extração do padrão a partir das vistas capturadas pelas múltiplas câmeras é mais simples se comparado ao dos padrões planares.

Em relação à liberdade de movimentação e manipulação, esse novo formato permite que um usuário comum não tenha muitas restrições quando ao realizar o processo de captura de uma amostra de vistas do padrão, que é feito normalmente movimentando-o por toda a área de rastreamento. Essa simplicidade permite que o padrão seja colocado em áreas do espaço de rastreamento bem complicadas de capturar nos distintos campos de visão das câmeras.

Finalmente, o processamento da imagem e a extração das posições 2D dos marcadores que compõem esse padrão são bem flexíveis e menos complexos que o requerido para os padrões planares. Normalmente o processamento da imagem fica bem caracterizado pelo objetivo de encontrar alguma estratégia que ressalte os marcadores da vara em relação a outros objetos ou ao fundo do espaço de rastreamento. Um exemplo de duas adaptações são os trabalhos de [22] e [37] (Figura 7). No primeiro, a solução foi montar a vara com marcadores coloridos. As cores dos marcadores são facilmente distinguíveis em relação ao fundo do espaço de rastreamento, e só a aplicação de um algoritmo de *threshold* que use o valor da cor RGB dos marcadores é a única característica necessária para extraí-los de uma imagem capturada pelas câmeras. Já no trabalho de [37] segue uma estratégia mais próxima à usada em sistemas de rastreamento comerciais: usar iluminação infravermelha no espaço de rastreamento, junto a câmeras sensíveis a esse tipo de luz e marcadores revestidos com tecido retrorreflexivo. Essa arquitetura permite que os marcadores apareçam nas imagens das câmeras como os objetos mais brilhantes. Novamente, aplica-se um algoritmo de *threshold*, nesse caso, com um certo valor de brilho alto, perto da cor branca numa escala de tons de cinza, para extrair facilmente os marcadores nas imagens capturadas. Na figura 7 podemos observar padrões do tipo 1D utilizados tanto nos trabalhos de [37][22].

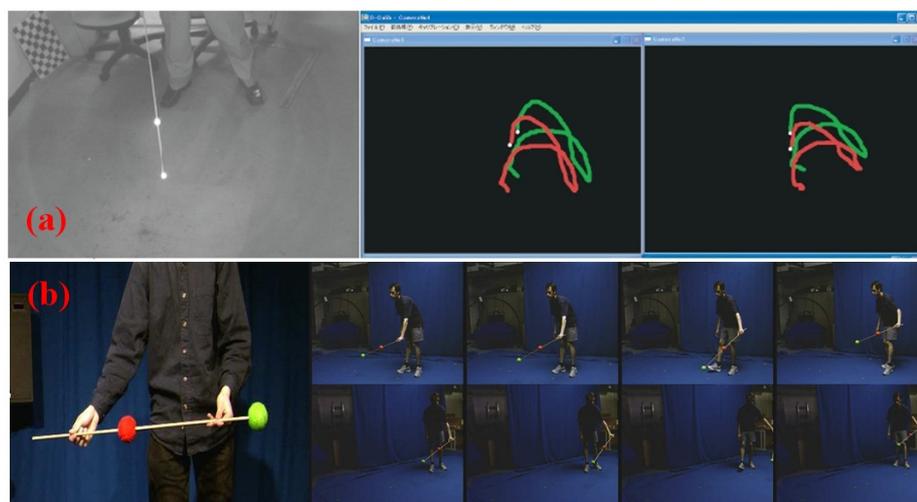


Figura 7: (a) Ambiente com uso de marcadores retrorreflexivos em Uematsu et al.[37], (b) marcadores coloridos usados como padrão em Mitchelson et al.[22].

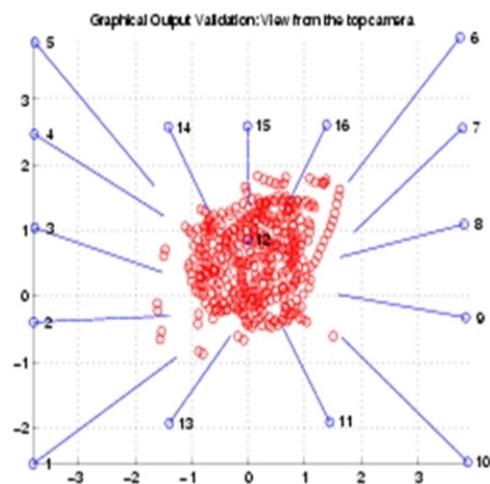
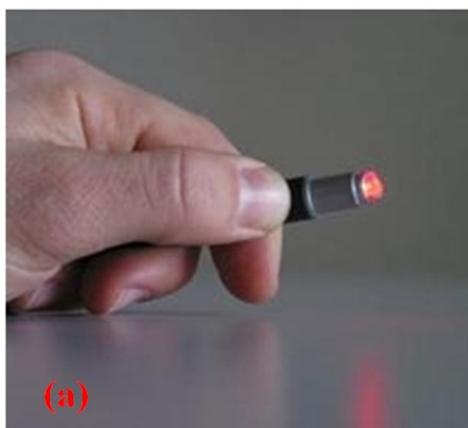
2.1.4. Padrão adimensional

O padrão considerado adimensional (0D) foi proposto e descrito recentemente no trabalho publicado por [32], que propôs um novo método de calibração de um sistema de múltiplas câmeras.

Esse padrão adimensional traz como vantagens as mesmas do padrão 1D: liberdade na movimentação do padrão, fácil visualização e processamento muito leve na captura e extração do padrão a partir das vistas capturadas pelas múltiplas câmeras. A principal diferença e a vantagem adicional é estar composto por um único marcador, o que o faz mais simples quanto ao seu desenho e às restrições para seu correto reconhecimento.

Uma versão modificada e mais restritiva do método apresentado por [32] tem sido usado por [28] na implementação de uma versão acadêmica de um sistema de rastreamento óptico. Esse sistema evoluiu para uma versão comercial chamado IOTracker [13].

Na figura 8, são mostradas algumas imagens que exemplificam a captura deste padrão e a posterior reconstrução 3D. No caso do trabalho publicado em [32] mostra-se como esse tipo de padrão foi implementado a partir de um ponteiro laser modificado (Figura 8a). Já nas propostas de [28] e [13] o padrão foi modelado como um marcador esférico coberto por tecido retrorreflexivo (Figura 8b).



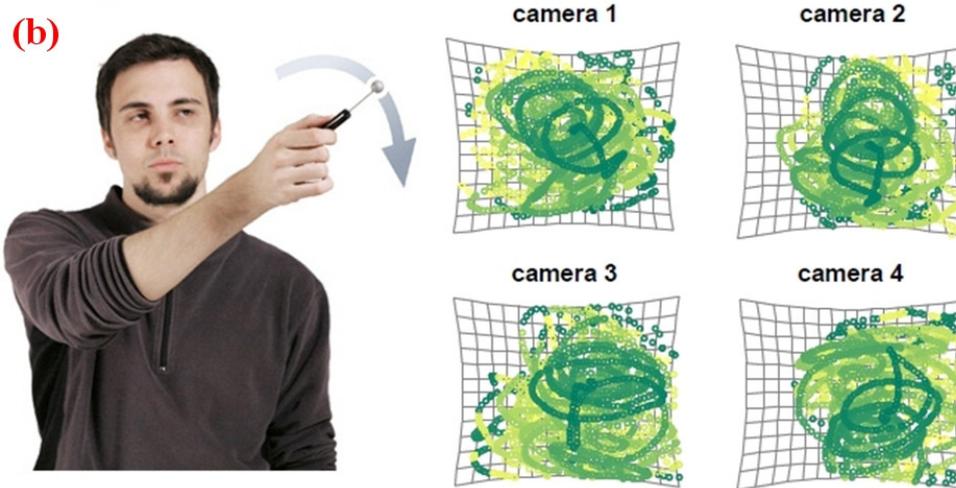


Figura 8: Marcadores adimensionais propostos por (a) Svoboda et al. [32], (b) Pintaric et al. [28] e [13].

Para poder achar os centróides desse tipo de marcador, o primeiro passo na captura do padrão é estabelecer uma estratégia de contraste que permita distinguir os marcadores dos outros objetos presentes no espaço de rastreamento. Por exemplo, para a calibração das câmeras no caso do padrão definido por [37] e os software comerciais [1] e [39] foi necessário montar uma infra-estrutura que usa:

- Iluminação de fontes de luz infravermelha invisível na faixa dos 850nm aproximadamente.
- Marcadores cobertos com fita ou tecido retroreflexivo.
- Câmeras com alta sensibilidade à luz infravermelha.

Essa infraestrutura permite que os marcadores sejam facilmente distinguidos nas imagens capturadas pelas câmeras por causa do brilho que eles apresentam em relação aos outros objetos no espaço de captura da amostra. No caso do trabalho apresentado por [32], esse contraste é conseguido pelo alto brilho que o ponteiro laser gera. Para isso tenta-se deixar esse espaço o mais escuro possível.

Uma vez criado o contraste, os próximos passos comumente definidos para extrair e identificar os marcadores nas imagens são:

- Conversão da imagem capturada para um formato em tons de cinzas.
- Aplicação de um filtro de *threshold* sobre a imagem em tons de cinza com a finalidade de binarizar e ressaltar mais ainda os marcadores que aparecem na imagem.

- Uma vez binarizada a imagem, os marcadores devem aparecer como áreas circulares dentro da imagem. É sobre essas áreas que se tenta calcular o centróide dos pixels.
- Após ter achado o centróide dessas áreas brancas, a coordenada 2D do centróide é considerada nosso ponto de referência relacionado à imagem do marcador correspondente.

O processamento da imagem descrito acima é feito para cada quadro capturado em cada câmera durante a movimentação do padrão 1D ou adimensional.

Tanto no caso do padrão unidimensional como no do adimensional, a simplicidade do processamento das imagens e a facilidade de movimentação tornam possível a captura de uma amostra de pontos de referência muito bem espalhada dentro da área de interseção dos campos de visão das câmeras, assim como a captura de um número grande de pontos de referência. Em contraposição, essa simplicidade do formato pode acarretar alguns problemas posteriores que podem desencadear em erros no processo de calibração das câmeras.

O principal problema desse tipo de padrão é a fácil contaminação da amostra de pontos de referência por causa de ruído ou de outros elementos erroneamente reconhecidos como possíveis marcadores válidos. Esses problemas são descritos nos trabalhos de [4] e [32], e algumas soluções e estratégias são propostas para descartar esses ruídos que podem gerar a identificação de falsos marcadores dentro das imagens processadas. As ferramentas mais usadas para descartar esses falsos marcadores variam desde o uso de algoritmos de predição do movimento até rotinas de correspondência só utilizadas quando é computada a matriz fundamental, que é definida na seção 2.2.2.4.

Nas soluções antes citadas, também se observam os diferentes momentos no processo de calibração nos quais elas podem ser aplicadas. Por exemplo, o uso de um algoritmo de predição de movimento seria considerado uma medida preventiva que pode ser utilizada logo ao início do processo de calibração, isto é, na etapa de captura do padrão, sem se estender para uma etapa posterior. Já a solução do uso da teoria por trás do cálculo da matriz fundamental para encontrar falsos marcadores é aplicada após a amostragem ter sido capturada. Essa solução é um pouco mais corretiva que preventiva e traz como ponto negativo a necessidade

de esperar até esse estágio avançado no processo de calibração para detectar os possíveis erros vindos da etapa de captura da amostra.

2.2. Segunda etapa: calibração inicial

A segunda etapa tem como objetivo agrupar e descrever os diferentes métodos utilizados para calcular o valor inicial dos parâmetros intrínsecos, extrínsecos e da distorção das lentes das múltiplas câmeras que compõem nosso sistema. O processo do cálculo dos parâmetros das câmeras é definido como calibração de câmera.

Em geral, as diferentes técnicas de calibração de câmera podem ser agrupadas em duas categorias [7][8]: calibração fotogramétrica e auto-calibração.

Na implementação dos diversos métodos desenvolvidos para múltipla calibração de câmeras é muito comum dividir a tarefa de calibração de certos parâmetros usando métodos das duas categorias.

A seguir, definiremos essas duas categorias e as complementaremos com a descrição de dois algoritmos que pertencem a cada uma delas. Esses algoritmos são utilizados na implementação do método de múltipla calibração de câmeras proposto nesta tese.

2.2.1. Calibração fotogramétrica

Neste tipo de método calculam-se os parâmetros de calibração de uma câmera usando a informação fornecida por um padrão cuja geometria é conhecida com precisão no espaço tridimensional onde ele é posicionado [7][8]. Um exemplo clássico desse tipo de método é o proposto por Tsai [35][36], que usa um padrão planar montado num sistema de trilhos que movimenta o padrão com precisão em frente à câmera. Outro trabalho semelhante, porém mais flexível, é proposto por Zhang [40], que também utiliza um padrão planar movimentado livremente diante da câmera.

Com base nas informações fornecidas pelos pontos referenciais colocados sobre o padrão planar, encontra-se uma correspondência direta entre a posição 3D dos marcadores nesse padrão e os pontos 2D correspondentes extraídos das imagens capturadas do padrão.

Este tipo de método gera resultados muito precisos, mas na prática, é pouco utilizado por causa do alto custo requerido para montar um padrão planar que ocupe o maior campo de visão da câmera, além de movimentá-lo, capturá-lo e reconhecê-lo com precisão em cada vista capturada do padrão. Outro inconveniente é o tempo despendido na captura de uma amostra de vistas válidas do padrão.

Este método está mais relacionado ao processo de calibração individual de câmera, e é muito utilizado especialmente para calcular os parâmetros intrínsecos e os coeficientes de distorção da lente. Os parâmetros extrínsecos também podem ser recuperados, porém, não são únicos e um valor diferente é calculado para cada uma das vistas do padrão planar capturadas e posteriormente processadas. Isso acontece porque os parâmetros extrínsecos calculam a posição e a orientação da câmera em relação a cada posição em que o padrão planar é capturado. Será descrito a seguir o método de calibração de câmera definido por Zhang [40].

2.2.1.1. Calibração baseada no cálculo de homografias

Para descrever esse tipo específico de calibração de câmera, incluído na categoria de métodos fotogramétricos, usaremos o modelo descrito no trabalho de Zhang [40], definido como: um método flexível para calibração de câmera que requer como dados de entrada vistas capturadas de um padrão planar que será movimentado em frente à câmera.

A calibração nesse método estima uma homógrafa H entre cada vista capturada de um padrão planar e a imagem projetada dele no plano da imagem da câmera. É assim que múltiplas homografias calculadas fornecem restrições suficientes para extrair os diferentes parâmetros da câmera.

O primeiro passo será calcular de forma robusta as homografias para cada vista do padrão capturada. Em seguida continuaremos com a extração dos parâmetros intrínsecos (únicos para a câmera) e extrínsecos contidos em cada homografia e finalmente faremos o cálculo dos coeficientes de distorção da lente.

2.2.1.2. Cálculo da homografia

Como foi apontado na seção 2.1.2, um padrão planar fornece um conjunto de pontos de referência distribuídos num formato de grade sobre sua superfície.

Sem perda de generalidade, podemos assumir que esse plano determina nosso sistema de coordenadas do mundo (SCM) e que cada ponto sobre este plano tem sua coordenada $Z = 0$. Com base nessa restrição, a matriz de projeção é definida na equação 1:

$$\begin{aligned} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &\cong K \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong P \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &\cong [p_1 \quad p_2 \quad p_3 \quad p_4] \begin{bmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{bmatrix} \cong [p_1 \quad p_2 \quad p_4] \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \end{aligned} \quad (1)$$

Como podemos observar uma das colunas da matriz de projeção da câmera P é anulada devido à coordenada $Z = 0$. Se nós definíssemos a equação 1 mais explicitamente, determinando quais componentes dos parâmetros das câmeras estariam definidos pelas colunas p_1, p_2, p_4 na equação 1, teríamos a seguinte equação:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad (2)$$

Na equação 2, r_1 e r_2 são as duas primeiras colunas da matriz de rotação R , t é o vetor de translação que definem os parâmetros extrínsecos da câmera, e a matriz K determina os parâmetros intrínsecos da câmera.

Então, a homografia H achada para cada vista do nosso padrão planar é determinada pelo produto $K [r_1, r_2, t]$ mais um fator de escala. A homografia permite relacionar um ponto de referência m detectado numa imagem capturada e processada do padrão planar, com seu ponto correspondente M , definido em coordenadas do mundo $(X_w, Y_w, 1)$ definido no modelo real do padrão planar. Essa correspondência é exemplificada na equação 3.

$$\begin{aligned} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &\cong K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong H \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \Rightarrow m \cong H M \\ H &= \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \end{aligned} \quad (3)$$

Baseados na relação descrita na equação 3, nós podemos extrair duas equações para cada ponto correspondente $m_i \leftrightarrow M_i$:

$$\begin{aligned} u(h_{31}X_w + h_{32}Y_w + h_{33}) &= h_{11}X_w + h_{12}Y_w + h_{13} \\ v(h_{31}X_w + h_{32}Y_w + h_{33}) &= h_{21}X_w + h_{22}Y_w + h_{23} \end{aligned} \quad (4)$$

Com essas duas equações relacionando os pontos m_i e M_i em cada vista do padrão planar capturado, podemos montar um sistema de equações que nos permitirá calcular os componentes h_{ij} da nossa matriz de homografia H . O sistema é definido na equação 5.

$$\begin{bmatrix} X_w & Y_w & 1 & 0 & 0 & 0 & -uX_w & -uY_w & -u \\ 0 & 0 & 0 & X_w & Y_w & 1 & -vX_w & -vY_w & -v \end{bmatrix} h = 0 \quad (5)$$

$$h = [h_{11} \quad h_{12} \quad h_{13} \quad h_{21} \quad h_{22} \quad h_{23} \quad h_{31} \quad h_{32} \quad h_{33}]^T$$

Se tivermos em nosso padrão n pontos correspondentes $m_i \leftrightarrow M_i$, poderemos formar um sistema de equações definido pela multiplicação da matriz L , de dimensão $2n \times 9$, e o vetor h , que é formado pelos componentes da nossa matriz de homografia H . Esse sistema de equações é definido por $Lh = 0$ e tem como solução o autovetor associado ao menor autovalor da matriz $L^T L$. Para calcular a solução para o sistema $Lh = 0$ aplicaremos o método chamado de “decomposição em valores singulares” (SVD - *singular value decomposition*). Devido às diferentes escalas dos componentes da matriz L , esta fica numericamente mal condicionada, e é recomendável fazer uma normalização dos dados m_i e M_i antes de montar nosso sistema de equações.

Por fim, uma vez aplicado o método de SVD, determinaremos todos os componentes h_{ij} que definem nossa matriz de homografia H , lembrando que esse cálculo da homografia H é feito para cada vista do padrão planar capturada, o que produzirá ao final, n matrizes de homografia.

2.2.1.3. Determinação dos parâmetros intrínsecos e extrínsecos

Uma vez calculadas as n homografias, correspondentes às n vistas válidas capturadas de nosso padrão planar, continuaremos com a extração da matriz de parâmetros intrínsecos K baseada na análise das múltiplas homografias calculadas.

Para realizar a extração dessa matriz K , utilizaremos o conceito sobre cônica absoluta, cuja descrição pode ser encontrada no livro de Hartley e Zisserman [10], nele a cônica absoluta é definida como:

$$K^{-T}K^{-1} = \begin{bmatrix} \frac{1}{f_x^2} & -\frac{s}{f_x^2 f_y} & \frac{sv_0 - u_0 f_y}{f_x^2 f_y} \\ -\frac{s}{f_x^2 f_y} & \frac{s^2}{f_x^2 f_y^2} + \frac{1}{f_y^2} & -\frac{s(sv_0 - u_0 f_y)}{f_x^2 f_y^2} - \frac{v_0}{f_y^2} \\ \frac{sv_0 - u_0 f_y}{f_x^2 f_y} & -\frac{s(sv_0 - u_0 f_y)}{f_x^2 f_y^2} - \frac{v_0}{f_y^2} & \frac{(sv_0 - u_0 f_y)^2}{f_x^2 f_y^2} + \frac{v_0^2}{f_y^2} + 1 \end{bmatrix} \quad (6)$$

A matriz K contém seis incógnitas definidas pelos parâmetros intrínsecos (f_x, f_y, s, u_0, v_0) mais um fator de escala λ . Já a cônica absoluta é definida como a matriz simétrica $K^{-T}K^{-1}$ contendo seis entradas. Como pode ser observado na equação 6, as incógnitas da matriz K determinam as entradas da matriz $K^{-T}K^{-1}$, que por sua vez é a definição da cônica absoluta. Essa relação permite calcular os componentes da matriz $K^{-T}K^{-1}$ e extrair os parâmetros intrínsecos a partir desses componentes.

A imagem de uma cônica absoluta é obtida a partir das seguintes propriedades baseadas na definição da homografia H :

$$H = \lambda K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \quad (7)$$

As propriedades são determinadas pela multiplicação de HH^T , no qual λ é o fator de escala, assim temos:

$$\begin{aligned} H^T H &= \lambda^2 (K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix})^T K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \\ H^T H &= \lambda^2 \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}^T K^T K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \\ H^T K^{-T} K^{-1} H &= \lambda^2 \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}^T \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \end{aligned} \quad (8)$$

Considerando o lado direito da equação 8, determinamos duas propriedades a partir da ortonormalidade da matriz de rotação definidas pelas suas colunas r_1 e r_2 .

1. A coluna r_1 da matriz de rotação é perpendicular a r_2 , o que faz do seu produto zero, sendo esse resultado inalterado pela escala λ .

$$r_1^T r_2 = 0 \quad (9)$$

2. Outra restrição que vem da ortonormalidade é que cada coluna da matriz de rotação é um vetor unitário, o que pode ser expresso como $r_1^T r_1 = 1$.

$$r_1^T r_1 = r_2^T r_2 \quad (10)$$

Se a matriz de homografia H é formada pelos vetores das colunas h_1, h_2 e h_3 , então podemos escrever as restrições dadas nas equações 9 e 10 dentro da equação 8, assim teremos:

$$h_1^T K^{-T} K^{-1} h_2 = 0 \quad (11)$$

$$h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \quad (12)$$

Agora se definimos uma matriz B como:

$$B = K^{-T} K^{-1} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \quad (13)$$

Notamos que esta matriz B representa a imagem da cônica absoluta, descrita como uma matriz simétrica, o que permite defini-la com seis incógnitas. Reescreveremos as equações em 11 e 12 usando essa matriz.

$$h_1^T B h_2 = 0 \quad (14)$$

$$h_1^T B^{-1} h_1 - h_2^T B h_2 = 0 \quad (15)$$

Expandindo as equações 14 e 15, podemos montar um sistema de equações que permita calcular os valores dos componentes b_{ij} da matriz simétrica B . Essa expansão é definida na equação embaixo.

$$\begin{bmatrix} h_1 h_{12} & h_2 h_{21} + h_1 h_{22} & h_2 h_{22} & h_2 h_{31} + h_1 h_{32} & h_2 h_{31} + h_2 h_{32} & h_3 h_{32} \\ h_{11}^2 - h_{12}^2 & 2(h_1 h_{21} - h_{12} h_{22}) & h_{21}^2 - h_{22}^2 & 2(h_1 h_{31} - h_{12} h_{32}) & 2(h_2 h_{31} - h_{22} h_{32}) & (h_{31}^2 - h_{32}^2) \end{bmatrix} b = 0 \quad (16)$$

Na qual b é definida pelos componentes de matriz simétrica B :

$$b = [b_{11} \quad b_{12} \quad b_{22} \quad b_{13} \quad b_{23} \quad b_{33}]^T$$

Se nós temos n imagens de nosso padrão planar, podemos determinar n equações semelhantes à equação 16 e formar uma matriz V de dimensão $2n \times 6$ que defina o sistema:

$$Vb = 0 \quad (17)$$

Esse novo sistema de equações precisa de pelo menos três vistas de nosso padrão planar. Se $n \geq 3$ podemos achar uma solução utilizando novamente o método de decomposição de valores singulares (SVD).

Com os componentes de nossa matriz B já calculados, podemos extrair os valores dos parâmetros intrínsecos definidos na matriz K , utilizando as seguintes equações apontadas no trabalho de Zhang [40].

$$v_0 = \frac{b_{12} b_{13} - b_{11} b_{23}}{b_{11} b_{22} - b_{12}^2} \quad \lambda = b_{33} - \frac{b_{13}^2 + v_0 (b_{12} b_{13} - b_{11} b_{23})}{b_{11}}$$

$$f_x = \sqrt{\frac{\lambda}{b_{11}}} \quad f_y = \sqrt{\frac{\lambda b_{11}}{b_{11} b_{22} - b_{12}^2}}$$

$$s = \frac{-b_{12}f_x^2 f_y}{\lambda} \quad u_0 = \frac{sv_0}{f_y} - \frac{b_{13}f_x^2}{\lambda}$$

Uma vez calculado o valor da matriz K , podemos prosseguir estimando os parâmetros extrínsecos, isto é, a matriz de rotação e o vetor de translação. Para isso usaremos a equação 15, que define a nossa matriz da homografia H :

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \lambda K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix}$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = [h_1 \quad h_2 \quad h_3]$$

$$r_1 = \lambda K^{-1} h_1$$

$$r_2 = \lambda K^{-1} h_2$$

$$r_3 = r_1 \times r_2$$

$$t = \lambda K^{-1} h_3$$

Na qual r_1, r_2, r_3 são as colunas da matriz de rotação R e t é o vetor de translação.

Finalmente, tendo completado a extração dos parâmetros intrínsecos e extrínsecos para nossa câmera, faremos a extração dos coeficientes de distorção da lente.

2.2.1.4. Determinação dos coeficientes de distorção da lente

Uma vez calculados os parâmetros intrínsecos e extrínsecos para a câmera, podemos achar uma solução para definir os valores dos coeficientes de distorção radial (k_1, k_2) considerados no trabalho de Zhang [40]. Esses coeficientes modelam bem o problema gerado pela distorção da lente.

Para calcular o valor dos coeficientes utilizaremos a informação sobre a posição dos pontos referenciais $m_i = (u_d, v_d)$ processados e extraídos a partir das vistas capturadas de nosso padrão planar. Consideraremos essas posições como as posições reais dos pontos na imagem. Porém, também podemos obter uma aproximação dessas posições reprojando os pontos 3D $M_i = (X_w, Y_w, I)$ determinados por nosso padrão real usando a homografia H previamente calculada. Isso permite obter pontos reprojados sobre o plano da imagem $m'_i =$

(u_u, v_u) considerados pontos ideais sem distorção. O resíduo da diferença entre os valores dos pontos reais (u_d, v_d) e os pontos ideais (u_u, v_u) será usado para encontrar o valor dos nossos coeficientes de distorção. Essa relação entre pontos distorcidos ou reais e os sem distorção ou ideais é definida nas equações abaixo:

$$u_d = u_u + (u_u - u_0)(k_1 r^2 + k_2 r^4) \quad (18)$$

$$v_d = v_u + (v_u - v_0)(k_1 r^2 + k_2 r^4) \quad (19)$$

O valor de r é o raio determinado pelos valores normalizados do ponto sem distorção (u_u, v_u) em relação ao centro da imagem (u_0, v_0) . Esse novo ponto é definido como (x, y) e raio é calculado na equação embaixo.

$$r^2 = x^2 + y^2 \quad (20)$$

Com as equações 18, 19 e 20 podemos determinar o seguinte sistema de equações:

$$\begin{bmatrix} (u_u - u_0)r^2 & (u_u - u_0)r^4 \\ (v_u - v_0)r^2 & (v_u - v_0)r^4 \end{bmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} u_d - u_u \\ v_d - v_u \end{pmatrix} \quad (21)$$

Como temos m pontos referenciais (u_d, v_d) capturados por cada vista do padrão planar e n vistas do padrão, podemos montar um sistema de equações de dimensão $2mn$, definido da seguinte forma:

$$Dk = d$$

$$D = \begin{bmatrix} (u_u^1 - u_0)r^2 & (u_u^1 - u_0)r^4 \\ (v_u^1 - v_0)r^2 & (v_u^1 - v_0)r^4 \\ \vdots & \vdots \\ (u_u^{mn} - u_0)r^2 & (u_u^{mn} - u_0)r^4 \\ (v_u^{mn} - v_0)r^2 & (v_u^{mn} - v_0)r^4 \end{bmatrix}, k = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}, d = \begin{pmatrix} u_d^1 - u_u^1 \\ v_d^1 - v_u^1 \\ \vdots \\ u_d^{mn} - u_u^{mn} \\ v_d^{mn} - v_u^{mn} \end{pmatrix} \quad (22)$$

Usando o método de mínimos quadrados, podemos achar uma solução para esse sistema., que pode ser definida pela pseudo-inversa da matriz D :

$$k = (D^T D)^{-1} D^T d \quad (23)$$

Finalmente, termina-se o processo de calibração de câmera, usando como únicas informações as vistas de um padrão planar.

Um resumo do método apresentado nesta seção é descrito a seguir:

Método de calibração baseado no cálculo de homografias

- **Definição:** Calibração de câmera usando n vistas de um padrão planar

- **Dados de entrada:** Pontos 2D extraídos de cada vista do padrão linear
- **Fluxo do processo:**
 1. Calcular as homografias H_i para cada uma das n vistas capturadas do padrão planar, são necessárias $n \geq 3$ vistas.
 2. Calcular a matriz V de dimensão $2n \times 6$ determinada pelas restrições de ortonormalidade definidas nas equações 14, 15 e 16.
 3. Calcular uma solução para o sistema $Vb = 0$, no qual b contém os componentes da matriz simétrica B , que representa a imagem de um cônica absoluta definida a partir da matriz de parâmetros intrínsecos K (equações 6 e 16).
 4. Extrair os parâmetros intrínsecos, e depois os parâmetros extrínsecos usando as equações apontadas na seção 2.2.1.3.
 5. Estimar o valor dos coeficientes de distorção das lentes, conforme seção 2.2.1.4.
- **Dados de saída:** Solução inicial para os valores dos parâmetros intrínsecos, extrínsecos e coeficientes de distorção da lente para uma câmera.

Com a descrição desse método fotogramétrico tentamos explicar o processo que permite calibrar todos os parâmetros de uma câmera. Na seção seguinte descreveremos outro tipo de método de calibração, chamado de auto-calibração. Este tipo de método foi proposto para resolver a calibração de mais de uma câmera, o que já o diferencia dos métodos fotogramétricos.

2.2.2. Auto-calibração

Neste tipo de método, a calibração dos parâmetros intrínsecos e extrínsecos é feita a partir das informações fornecidas por pontos correspondentes corretamente identificados em imagens capturadas pela(s) câmera(s). Nesse sentido, este tipo de método não precisa de padrões complexos com uma estrutura bem conhecida, como os métodos fotogramétricos. Pelo contrário, os pontos correspondentes podem ser capturados usando padrões unidimensionais ou adimensionais, como os descritos na seção 2.1.3 e 2.1.4.

No caso de usar diferentes câmeras é preciso que os pontos referenciais estejam devidamente correlacionados e identificados entre as diversas câmeras

que visualizam um mesmo cenário. Pode-se afirmar que esses métodos assumem que a cena visualizada é estática, e é só baseado na informação 2D dos pontos referenciais capturados e corretamente correlacionados, que podemos extrair e calcular os parâmetros de calibração das câmeras. Esse tipo de método é também muito utilizado em aplicações que realizam reconstrução 3D.

A área de pesquisa que explora essa nova estratégia de calibração de câmera e define esse tipo de método é conhecida como “estrutura a partir do movimento” (*structure from motion*). A base da implementação deste tipo de método se encontra na teoria da geometria epipolar, representada principalmente pelo cálculo da matriz fundamental.

Nesta seção, descreveremos alguns conceitos referentes à geometria epipolar, ao cálculo da matriz fundamental e à extração dos parâmetros extrínsecos das câmeras baseada na decomposição da matriz fundamental. Esse processo de extração dos parâmetros extrínsecos é utilizado na implementação do método proposto neste trabalho.

2.2.2.1. Geometria epipolar

Geometria epipolar é a geometria intrínseca projetiva existente entre duas vistas de um mesmo cenário, capturadas a partir de diferentes perspectivas. As vistas podem ser capturadas por uma mesma câmera, sendo necessário apenas mudar a posição da câmera no momento da captura, ou as vistas podem vir da captura de várias câmeras enxergando um mesmo cenário simultaneamente.

A geometria epipolar é independente da estrutura da cena e só depende dos parâmetros internos das câmeras e da posição relativa entre elas [10]. Para definir a posição relativa entre as câmeras utiliza-se uma transformação composta de uma matriz de rotação e um vetor de translação, que permite levar o sistema de coordenadas da câmera (SCC) da primeira câmera para o da segunda. A matriz de rotação e o vetor de translação são considerados os parâmetros extrínsecos que relacionam os sistemas das duas câmeras, mais conhecido como sistema estéreo de câmeras.

Essa geometria é geralmente requerida devido à necessidade de encontrar uma correspondência entre pontos que aparecem nas diferentes imagens das vistas de um mesmo cenário. Assim, suponhamos que temos um ponto M no mundo, que

é projetado sobre o plano da imagem de duas câmeras como sendo os pontos m e m' respectivamente (Figura 9). Para mostrar a correspondência que existe entre os pontos m e m' , primeiro definiremos o plano π formado pelos vetores \overrightarrow{CM} , $\overrightarrow{C'M}$ e $\overrightarrow{C'C}$, nos quais estão contidos nossos pontos m e m' . Este plano é definido como o plano epipolar.

Agora, se projetamos o vetor \overrightarrow{CM} , que passa pelo ponto m no plano da imagem da câmera da esquerda sobre o plano da imagem da direita, descendo pelo plano epipolar, nós vemos que esse vetor é projetado na forma de uma linha l' no plano da imagem da direita. Essa linha deve conter o ponto m' que é a projeção do mesmo ponto M sobre o plano da imagem da direita. Esse processo permite demonstrar a correspondência entre os pontos m e m' , que são as projeções de um ponto no espaço M sobre os planos da imagem das câmeras.

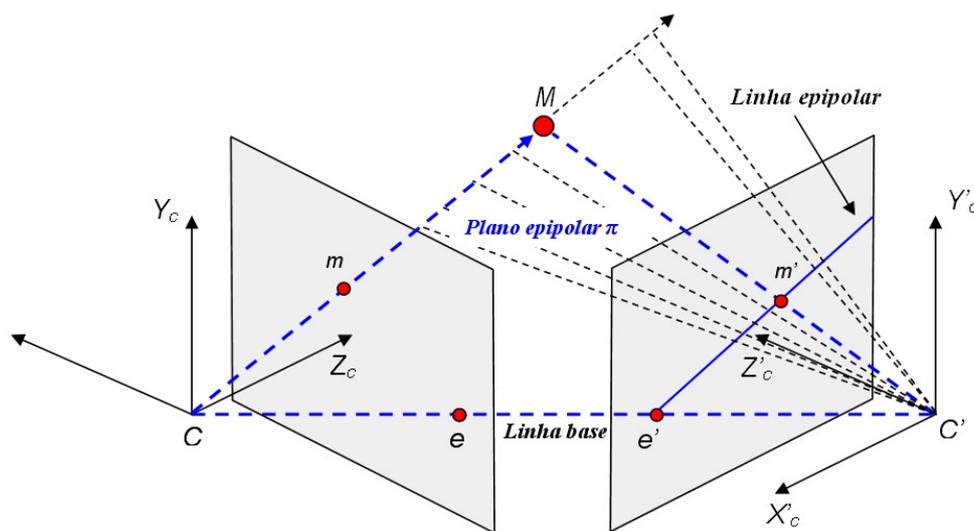


Figura 9: Elementos da geometria epipolar - linha epipolar, epipolos, linha base, plano epipolar.

Os pontos de interseção da linha base, determinada pelo vetor $\overrightarrow{C'C}$, com cada plano da imagem são definidos como os epipolos e e e' . Já as linhas l e l' formadas pelas projeções dos vetores definidos por \overrightarrow{CM} , $\overrightarrow{C'M}$ sobre os planos da imagem das câmeras, são descritas pelas linhas epipolares (Figura 10).

Como foi observado, a geometria epipolar ajuda a entender a geometria projetiva implícita existente entre duas vistas de um mesmo cenário, mas um elemento chave que formula essa correspondência formalmente é a matriz

fundamental. Na seção seguinte, descreveremos como essa matriz também relaciona os parâmetros das câmeras.

2.2.2.2. Matriz fundamental

Como foi definido, a matriz fundamental é um componente que permite expressar matematicamente as propriedades da geometria epipolar usando a álgebra linear.

Essa matriz encapsula a geometria intrínseca que descreve a geometria epipolar. Internamente essa matriz pode ser determinada pelos parâmetros intrínsecos e extrínsecos que correlacionam nossas câmeras.

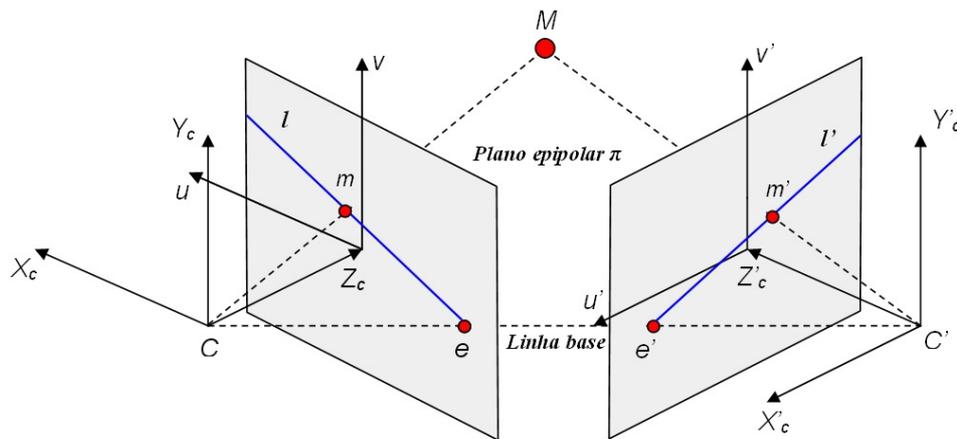


Figura 10: Correspondência entre duas vistas de um sistema estéreo de câmeras.

Tendo como referência a Figura 10, observarmos que um ponto M ($M' = M$, M' expresso no SCC da direita) junto à origem das câmeras C e C' define três vetores coplanares, \overrightarrow{CM} , $\overrightarrow{C'C}$ e $\overrightarrow{C'M'}$. Essa relação pode ser expressa como:

$$\overrightarrow{C'M'}^T \left(\overrightarrow{C'C} \times \left(\overrightarrow{CM} \right)' \right) = 0 \quad (24)$$

Na equação 24, $\left(\overrightarrow{CM} \right)'$ é o vetor \overrightarrow{CM} expresso no SCC da direita. Considerando m a imagem do ponto no espaço M expresso no SCC da esquerda e m' a imagem do ponto no espaço M' expresso no SCC da direita, no qual $M' = M$, nós podemos expressar essa relação com as seguintes equações:

$$m = K \begin{bmatrix} I_3 & | 0 \end{bmatrix} M = KM = K \overrightarrow{CM} \quad (25)$$

$$m' = K' \begin{bmatrix} I_3 & | 0 \end{bmatrix} M' = K'M' = K' \overrightarrow{C'M'} \quad (26)$$

O vetor \overrightarrow{CM} está localizado no SCC da esquerda cuja origem é C . Então podemos rodar o SCC da esquerda para ficar alinhado com o SCC da direita usando uma matriz de rotação R .

$$\left(\overrightarrow{CM}\right)' = R \overrightarrow{CM} \quad (27)$$

Substituindo a equação 27 na equação 24, temos:

$$\overrightarrow{C'M'}^T \left(\overrightarrow{C'C} \times R \overrightarrow{CM}\right) = 0 \quad (28)$$

Se resolvermos a equação 25 e 26 em relação aos vetores 3D \overrightarrow{CM} e $\overrightarrow{C'M'}$, definimos $\overrightarrow{CM} = K^{-1}m$ e $\overrightarrow{C'M'} = K'^{-1}m'$. Inserindo essas expressões na equação 28, podemos redefini-la como:

$$\left(K'^{-1}m'\right)^T \left(\overrightarrow{C'C} \times R K^{-1}m\right) = 0 \quad (29)$$

Na equação 29, podemos determinar que nossa linha base $\overrightarrow{C'C}$ é o vetor de translação t ($\overrightarrow{C'C} = t$), que levará a origem da nossa câmera à esquerda para a posição da origem da câmera à direita. Nós podemos usar a versão “skew symmetric matrix” do vetor t definida como \tilde{t} para representar o produto vetorial do vetor $\overrightarrow{C'C}$ e a matriz de rotação R na equação 29.

$$m'^T K'^{-T} \tilde{t} R K^{-1}m = 0 \quad (30)$$

Na qual o vetor \tilde{t} é definido como:

$$\tilde{t} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \Rightarrow t = \overrightarrow{C'C} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

É na transformação algébrica descrita na equação 30 que podemos expressar a correspondência entre os pontos m e m' nas duas vistas de nosso sistema estéreo de câmeras. Então a matriz fundamental é definida assim:

$$\begin{aligned} m'^T K'^{-T} \tilde{t} R K^{-1}m &= 0 \\ m'^T F m &= 0 \\ F &= K'^{-T} \tilde{t} R K^{-1} \end{aligned} \quad (31)$$

Como podemos observar, é a matriz fundamental que representa a álgebra definida pela geometria epipolar, internamente relacionando os parâmetros

intrínsecos e extrínsecos de nossas câmeras. A projeção do ponto m no plano da esquerda sobre o plano da imagem da câmera da direita fica definida como a linha epipolar l' e também pode ser expressa em função da matriz fundamental:

$$l' = Fm \quad (32)$$

$$l = F^T m' \quad (33)$$

Na equação 32, define-se a projeção do ponto m' no plano da direita sobre o plano da imagem da câmera da esquerda que determina a linha epipolar l . Essa projeção também é expressa em função da matriz fundamental (Figura 10).

Na seção a seguir será descrito o processo de cálculo da matriz fundamental baseado na informação fornecida por um conjunto de pontos correspondentes identificados corretamente entre as imagens capturadas por um sistema estéreo de câmeras.

2.2.2.3. Cálculo da matriz fundamental

Existem diferentes técnicas para o cálculo da matriz fundamental, como os apresentados no livro de Hartley e Zisserman [10], mas um método muito utilizado por sua simplicidade é o algoritmo dos 8 pontos, que precisa de 8 pontos correspondentes $m \leftrightarrow m'$ entre duas imagens para realizar esse cálculo. As informações sobre o conjunto de pontos correspondentes e a correspondência baseada na definição da matriz fundamental $m^T F m = 0$, permitirão construir um sistema de equações cuja solução achará o valor dos componentes de nossa matriz fundamental.

Para iniciar, nós teremos como dados de entrada um número, $n \geq 8$, de pontos correspondentes $m \leftrightarrow m'$ que são relacionados pela matriz fundamental baseados na equação $m^T F m = 0$. Nesta equação representaremos em coordenadas homogêneas nossos pontos $m = (u, v, 1)^T$ e $m' = (u', v', 1)^T$. Usando as coordenadas homogêneas de m e m' podemos expandir a equação 31 e redefini-la numa única equação:

$$m'^T F m = 0$$

$$\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix}^T \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0$$

$$uu'f_{11} + uv'f_{21} + uf_{31} + vu'f_{12} + vv'f_{22} + vf_{32} + u'f_{13} + v'f_{23} + f_{33} = 0 \quad (34)$$

É assim que cada ponto correspondente gerará uma equação como a apresentada acima. A equação 34 também pode ser modelada como o produto de dois vetores:

$$\begin{pmatrix} uu' & vu' & u' & uv' & vv' & v' & u & v & 1 \end{pmatrix} f = 0 \quad (35)$$

Na qual o vetor f contém os componentes que determinam a matriz fundamental F . Se nós consideramos um conjunto de n pontos correspondentes, poderemos modelar um sistema de equações da seguinte forma:

$$Af = \begin{bmatrix} u_1u'_1 & v_1u'_1 & u'_1 & u_1v'_1 & v_1v'_1 & v'_1 & u_1 & v_1 & 1 \\ \vdots & \vdots \\ u_nu'_n & v_nu'_n & u'_n & u_nv'_n & v_nv'_n & v'_n & u_n & v_n & 1 \end{bmatrix} f = 0 \quad (36)$$

Para poder achar uma solução para f baseada na matriz A , ela deve ter posto 8. A fim de evitar uma solução trivial $f = 0$, adicionamos algumas restrições, como obrigar a norma da solução para $|f| = 1$ e definir o último componente de f igual $f_{33} = 1$. Com essas duas restrições, podemos resolver o sistema de equações $Af = 0$ usando mínimos quadrados. Para calcular uma solução para $Af = 0$ podemos usar o método SVD, que determina a solução para f pelo autovetor relacionado ao menor autovalor definido para a matriz A . Com o vetor f calculado, nós determinamos os valores dos componentes da matriz fundamental F .

É importante garantir que a matriz fundamental seja singular e tenha posto 2, mas comumente a solução da matriz fundamental F achada com o método SVD não garante que a solução achada para os valores da matriz F tenha essas propriedades. Porém, existe um método simples que permite achar uma nova matriz fundamental F' que tenha posto 2 e esteja baseada na matriz F previamente calculada. Com base na minimização da norma de Frobenius $|F - F'|$ sujeita à restrição $\det(F') = 0$, o método consiste em realizar a SVD de nossa matriz F , que é definida como $F = UDV^T$, na qual $D = (r,s,t)$ é a matriz que contém na diagonal os dados dos autovalores para nossa matriz F . Os autovalores estão em ordem decrescente $r \geq s \geq t$. Então nossa nova matriz F' será definida como $F' = U \text{diag}(r,s,0) V^T$, onde o menor autovalor é zerado. Essa nova matriz F' minimiza a norma definida por Frobenius e estabelecida como nossa nova matriz fundamental.

Finalmente, com a matriz fundamental já calculada, o passo seguinte será entender como extrair os parâmetros extrínsecos para o sistema estéreo de câmeras. O cálculo dos parâmetros intrínsecos e da distorção da lente de cada câmera fica por conta do método fotogramétrico descrito na seção 2.2.1.3.

2.2.2.4. Determinação dos parâmetros extrínsecos

Como se observa na equação 31, a matriz fundamental pode ser descrita usando os parâmetros intrínsecos e extrínsecos das câmeras de um sistema estéreo. Até este momento temos já calculados os parâmetros intrínsecos (K , K') e a matriz fundamental F para esse sistema, mas ainda falta calcular os parâmetros extrínsecos definidos por R e t na equação 31.

Os parâmetros extrínsecos determinam um novo tipo de matriz conhecida como a matriz essencial E . Essa matriz contém todas as informações que permitem posicionar uma câmera em relação à câmera vizinha e é definida na equação 37:

$$\begin{aligned} F &= K'^{-T} \tilde{t} R K^{-1} \\ F &= K'^{-T} E K^{-1} \\ E &= \tilde{t} R \end{aligned} \tag{37}$$

Uma forma de calcular os componentes R e t dessa matriz essencial será apresentada a seguir:

Primeiro, com base na definição da equação 31 e sabendo os valores das matrizes de parâmetros intrínsecos das câmeras, podemos extrair a matriz essencial de nossa matriz fundamental, previamente calculada, da seguinte forma:

$$E = K'^T F K \tag{38}$$

Segundo, vamos assumir que o SCC da esquerda é nosso sistema de coordenadas base. Isso significa que nossa câmera à direita estará posicionada em relação ao SCC da esquerda, o que nos permite criar duas novas matrizes de projeção P e P' simplificadas. Estas são definidas unicamente pela matriz de rotação R e pelo vetor de translação t . A definição de P e P' exemplifica o processo que permite levar o SCC da direita para o SCC da esquerda:

$$\begin{aligned} P &= [I_3 \quad | \quad 0] \\ P' &= [R \quad | \quad t] \end{aligned}$$

Agora suponhamos que a SVD da nossa matriz essencial E é definida da seguinte forma:

$$E = U \text{diag} (1 \quad 1 \quad 0) V^T \quad (39)$$

Baseados na SVD da matriz essencial E , nós podemos determinar a matriz R e o vetor t como segue:

$$R = U W V^T \quad \text{ou} \quad R = U W^T V^T \quad (40)$$

$$t = u_3 \quad \text{ou} \quad t = -u_3 \quad (41)$$

Onde u_3 é a última coluna de nossa matriz U , e a matriz W é definida como:

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Essa estratégia para achar possíveis soluções para a matriz R e o vetor t baseada na fatorização das matrizes componentes da SVD da matriz E , gera quatro possíveis soluções produto das combinações das duas possíveis soluções para R e t determinadas nas equações 40 e 41. Essas soluções ficam definidas como:

$$\begin{aligned} P' &= [U W V^T \mid u_3] & P' &= [U W V^T \mid -u_3] \\ P' &= [U W^T V^T \mid u_3] & P' &= [U W^T V^T \mid -u_3] \end{aligned}$$

A ambiguidade criada pela definição dessas quatro possíveis soluções é ilustrada na figura 11.

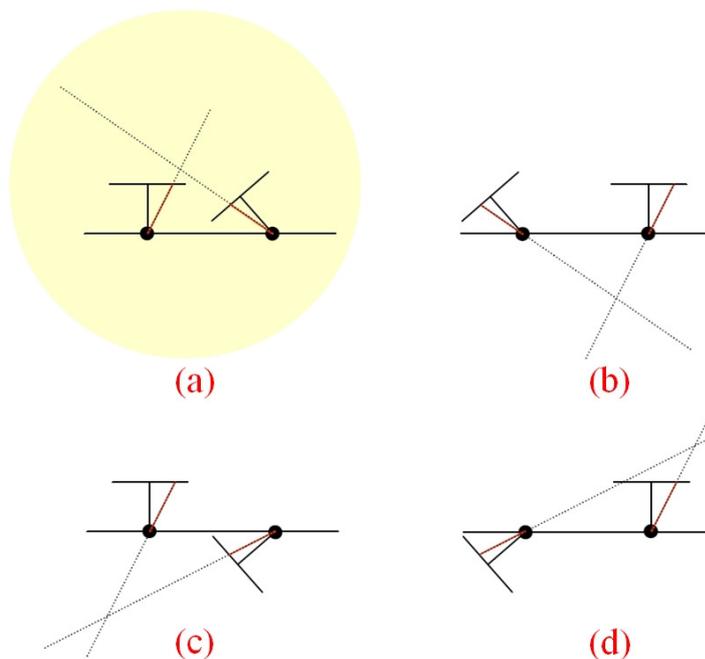


Figura 11: Ambiguidade da definição das quatro possíveis soluções para R e t [10].

Como podemos observar na Figura 11, só uma das soluções gera um sistema de coordenadas válido, onde o espaço de visão das câmeras ficará definido em frente às duas câmeras (Figura 11a). Para poder encontrar qual solução das quatro soluções propostas tem a característica visualizada na Figura 11a é realizado um teste simples, que consiste em:

1. Escolher um ponto da amostra de pontos $m \leftrightarrow m'$.
2. Usando os valores para R e t determinados em cada solução, reconstruir esse ponto no espaço.
3. Determinar a profundidade desse ponto em relação a cada câmera.
4. Escolher a solução na qual a profundidade encontrada para o ponto reconstruído é positiva em relação às duas câmeras.

Finalmente, terminamos o processo de cálculo de todos os parâmetros das câmeras que formam um sistema estéreo. Os métodos descritos nesta seção têm como objetivo achar uma solução inicial para esses valores. No entanto, quando há um sobredimensionamento de informações de entrada – como um número maior de vistas do padrão planar, no caso do método de calibração fotogramétrico ou mais do que oitos pontos referenciais correspondentes no cálculo da matriz fundamental – talvez seja necessário algum método de otimização que permita melhorar essa estimativa inicial. Na seção a seguir será descrito esse tipo de método, que define a última etapa do nosso framework conceitual definido para o processo de múltipla calibração de câmeras.

2.3. Terceira etapa: otimização da calibração

Esta terceira etapa tem como objetivo agrupar os diferentes métodos usados para otimizar o valor inicial dos parâmetros de calibração, obtidos com os métodos anteriormente descritos. As soluções obtidas para os parâmetros das câmeras, tanto com o método fotogramétrico como com o de auto-calibração, foram calculadas com base na minimização algébrica de um sistema de equações que modela a correspondência entre pontos 3D-2D (fotogramétrica) e pontos 2D-2D (auto-calibração). Essas soluções satisfazem uma aproximação algébrica linear para um problema que não é linear, como o processo de calibração de câmera. A minimização algébrica pode também não levar em conta outras propriedades físicas do modelo de câmera real definidas pelas características dos pontos de

referência usados no processo de calibração e por alguns componentes internos das câmeras, como a lente. Por esse motivo e pela definição de não-linearidade do processo de calibração, é requerida a aplicação de algoritmos de otimização que minimizem o erro produzido na avaliação das características físicas implícitas, que vem junto as informações fornecidas pelos pontos referenciais capturados a partir dos nossos padrões de calibração.

Os algoritmos de minimização não-linear tentam melhorar a estimativa inicial obtida para os parâmetros intrínsecos, extrínsecos e coeficientes de distorção da lente das múltiplas câmeras incluídas no sistema. Para poder realizar esse aprimoramento dos valores dos parâmetros, os algoritmos precisam da definição de uma função global de erro que avalie, a cada iteração, se a nova solução estimada a partir de alguma variação da solução inicial gera melhores resultados. Muitas vezes essa função de erro global não é única e é definida por um conjunto de funções de controle especificamente criadas e relacionadas a propriedades e características físicas presentes tanto no processo de calibração como no formato definido entre os pontos referenciais capturados a partir dos padrões de calibração. Algumas das funções de controle comumente utilizadas serão descritas na próxima seção.

2.3.1. Funções de controle

Entende-se como função de controle aquela função que avalia o resultado definido por alguma característica ou propriedade física obtida a partir dos componentes incluídos no processo de calibração de câmera. Descreveremos a seguir algumas das funções mais utilizadas.

2.3.1.1. Reprojção 3D – 2D

Este tipo de função é utilizada quando temos informação tanto da posição 2D na imagem como da posição 3D no espaço dos pontos referenciais capturados em nossa amostra para calibração. Com essa informação mais a estimativa dos parâmetros de calibração, nós podemos reprojeter sobre o plano da imagem os pontos 3D, que devem ficar muito próximos aos pontos 2D correspondentes aos mesmos pontos referenciais, mas que foram extraídos diretamente da imagem

baseados nos algoritmos de processamento de imagem como os descritos na seção sobre padrões de calibração.

O somatório da diferença entre o ponto 2D extraído a partir da imagem e o calculado com base no processo de reprojeção é o valor de erro determinado para esse tipo de função de controle. O objetivo da minimização é aproximar este de zero. Essa função é utilizada tanto no método fotogramétrico quanto no método de auto-calibração e é definida como:

$$\mathcal{E}_{Reproj. 3D-2D} = \sum_{i=1}^n \sum_{j=1}^m \left\| m_{ij} - m_{Reproj}(K, R, t, M_j) \right\|^2 \quad (42)$$

Na equação 42, o erro de reprojeção é estabelecido pelo somatório da norma das distâncias entre o ponto de referência extraído a partir da imagem m_{ij} e o ponto correspondente m_{Reproj} – que é a reprojeção do ponto M_j no espaço usando os parâmetros da câmera K, R, t . Nós podemos ter $i=1\dots n$ número de vistas do padrão de calibração capturadas que internamente fornecem $j = 1..m$ pontos de referência por vista.

2.3.1.2. Reprojeção 2D – 2D

Este tipo de função é definida como a distância da linha epipolar ao ponto correspondente num sistema estéreo de câmeras, como exposto na seção 2.2.2.4. Para poder modelar essa função precisamos das informações sobre os pontos correspondentes 2D $m_i \leftrightarrow m_i'$, extraídos a partir de duas imagens de um mesmo cenário. Para ambas, já calculamos o valor da matriz fundamental que as relaciona.

A relação estabelecida pela matriz fundamental permite a projeção de um ponto m numa câmera A sobre o plano da imagem de uma câmera B . A matriz fundamental F determina a relação entre as câmeras A e B . A projeção do ponto m então cria uma linha l' no plano da imagem da câmera B (Figura 12) e sobre essa linha é que deve estar posicionado o ponto correspondente m' . No caso ideal, a distância do ponto m à linha l' deve ser zero quando o ponto m cai exatamente sobre a linha l' . No entanto, por causa do ruído e do valor inicial dos parâmetros da câmera, essa distância fica perto de zero.

Então essa função é definida como uma projeção 2D-2D, na qual o erro estabelecido é medido pela distância de um ponto m' à sua linha correspondente l' e pela distância de um ponto m à sua linha correspondente l .

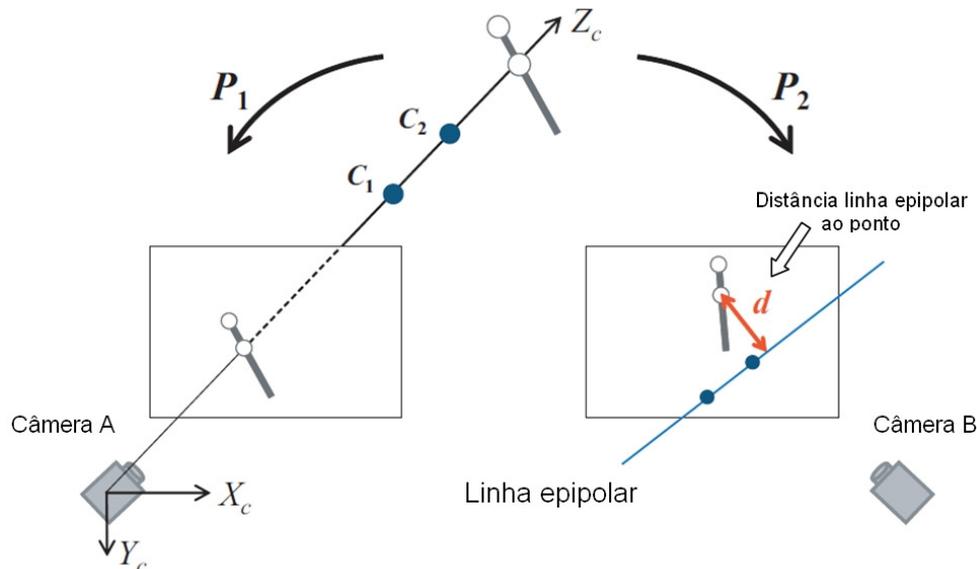


Figura 12: Exemplificação do erro definido pela distância da linha epipolar ao ponto correspondente.

O valor do erro estabelecido pelo somatório das diferenças entre as linhas l e l' a partir da projeção dos pontos correspondentes $m_i \leftrightarrow m_i'$ é o que o algoritmo de otimização minimizará. Esse tipo de função é utilizado nos métodos de auto-calibração, especialmente nos que calculam o valor da matriz fundamental. Essa função é definida como:

$$\mathcal{E}_{\text{Reproj. 2D-2D}} = \sum_{i=1}^n \text{dist}(m_i, F^T m_i') + \text{dist}(m_i', F m_i) \quad (43)$$

A equação 43 expressa o erro de reprojeção 2D- 2D, estabelecido pelo somatório das distâncias entre o ponto m ou m' e a linha correspondente $l = F^T m'$ e $l' = F m$. Nós podemos ter $i=1 \dots n$ número de pontos correspondentes $m_i \leftrightarrow m_i'$.

2.3.1.3. Distância 3D entre pontos de referência reconstruídos

Este tipo de função é definida como a distância 3D que pode ser recuperada quando temos nossos pontos correspondentes $m_i \leftrightarrow m_i'$ reconstruídos tridimensionalmente no espaço determinado pelo campo de visão das câmeras. Essa reconstrução 3D deve ser feita em escala semelhante a alguma unidade de medida métrica (metros, centímetros, milímetros, etc). Um exemplo de aplicação

dessa função de controle é quando usamos padrões unidimensionais, como os descritos na seção 2.1.3.

O processo de reconstrução 3D dos pontos correspondentes $mi \leftrightarrow mi'$ pode ser feito usando um algoritmo de triangulação como o apontado em [29][30], que usa todos os parâmetros da câmera para calcular a posição 3D de um marcador M com informações dadas por mi e mi' . Com esses dados poderemos comparar as distâncias medidas a partir dos marcadores reconstruídos com o valor real medido manualmente diretamente entre os marcadores físicos do padrão. Este tipo de função é utilizada nos métodos fotogramétricos e no de auto-calibração e é definida como:

$$\mathcal{E}_{dist.3D} = \sum_{i=1}^m \left(\left| d_0 - \|X_0^i - X_1^i\| + \dots + \left| d_n - \|X_n^i - X_{n+1}^i\| \right| \right| \right) \quad (44)$$

A equação 44 expressa o somatório do erro gerado pela diferença das distâncias $\|X_n - X_{n+1}\|$ entre marcadores reconstruídos $X_{i=1...n+1}$ e a medida real da distância definida como $d_{i=1...n}$, tudo isso para m vistas reconstruídas do padrão.

Assim estabelecemos quais são as funções de controle mais utilizadas para determinar a função de erro global que nosso algoritmo de otimização usará para aprimorar nossa solução inicial. A função de erro global pode ser expressa da seguinte maneira:

$$\mathcal{E}_{\text{Erro Global}} = \mathcal{E}_{\text{Reproj. 3D-2D}} + \mathcal{E}_{\text{Reproj. 2D-2D}} + \mathcal{E}_{\text{dist. 3D}} \quad (45)$$

Essa função de erro pode mudar dependendo do método aplicado e do padrão a ser usado na aquisição de dados, onde deve se garantir que haja informação de entrada suficiente para modelar as funções de controle antes descritas. Um exemplo simples que representa a observação anterior é o uso de um padrão adimensional, que só gera informações suficientes para implementar o erro de reprojeção 3D-2D e reprojeção 2D-2D. Já os padrões planar e unidimensional podem garantir informações para definir uma função de erro da forma explicitada na equação 45.

Por fim, uma vez determinados os dois componentes básicos para rodar um algoritmo de otimização – a amostra de entrada correspondente aos pontos referenciais capturados com diversos formatos de padrões e a definição das funções de controle e função de erro global – poderemos executar o algoritmo de

otimização. No caso do processo de calibração o comportamento dos parâmetros que definem o processo não é linear. Um dos algoritmos muito utilizados para esse processo é o descrito por Levenberg Marquart [2] e uma boa implementação pode ser encontrada em [6]. Outro método também utilizado para a otimização de parâmetros no processo de calibração é o Downhill Simplex [24], que foi usado nos trabalhos de [37]. Na implementação do método proposto neste trabalho foram aplicados os dois métodos para otimizar nossos parâmetros e realizar algumas medidas de comparação no capítulo de resultados.

A seção a seguir descreverá mais uma característica importante para estabelecer como podem ser montados os sistemas de múltiplas câmeras: a distribuição e posicionamento das câmeras em relação ao espaço de rastreamento. Esta característica gera restrições adicionais no momento de escolher qual o método de calibração a ser utilizado para um determinado tipo de distribuição.

2.4. Distribuição das câmeras

Entre os formatos existentes para distribuir e posicionar múltiplas câmeras que formarão parte de um sistema único – como o definido quando resolvemos o problema de múltipla calibração de câmeras – nós encontramos dois tipos bem diferentes: distribuição centralizada e descentralizada de câmeras, que serão descritos nas seções seguintes.

2.4.1. Distribuição centralizada de câmeras

Neste tipo de distribuição e posicionamento de múltiplas câmeras, encontram-se classificados os sistemas que se caracterizam pelo fato de todas as câmeras conseguirem enxergar um mesmo ponto central de referência, definido como centro do sistemas de coordenadas do mundo (SCM). Trabalhos apresentados em [4],[32] e [37] propõem e testam algoritmos de estéreo e múltipla calibração de câmeras em sistemas com esse tipo de distribuição. O posicionamento das câmeras costuma ter uma estrutura circular, como o apresentado na figura 13.

Para determinar onde fica o centro do sistema de coordenadas do mundo é usado algum tipo de padrão geralmente planar, como os descritos na seção 2.1.2. Usando o padrão como referência calculam-se a posição e a orientação das

câmeras em relação ao centro do padrão. Esse tipo de cálculo pode ser feito usando um método de calibração fotogramétrico que permite calcular os parâmetros extrínsecos de cada câmera em relação ao padrão .

A principal vantagem desse tipo de distribuição é que um mesmo padrão pode ser visto simultaneamente por todas as câmeras, especialmente na captura de uma amostra para calibração. A desvantagem fica por conta da área restrita que as câmeras podem cobrir ou enxergar como área real de calibração. Um exemplo desse tipo de distribuição é mostrado na Figura 13.



Figura 13: Distribuição centralizada de câmeras.

(Fonte: <http://blogs.ign.com/Blogs/BlogPage.aspx?blog=EA-FIFA-Street&year=2008&month=01>)

Porém, existe outro grupo de distribuição de câmeras cujo objetivo é cobrir áreas amplas de captura e rastreamento que não podem ser enxergadas simultaneamente por todas as câmeras do sistema. Essa outra forma de distribuição é conhecida como descentralizada e será descrita na seção a seguir.

2.4.2. Distribuição descentralizada das câmeras

Neste outro tipo de classificação de distribuição das câmeras, encontram-se os sistemas que não têm uma área de visão comum entre todas as câmeras, porém, garante-se que para cada câmera há uma outra que tenha um campo de visão compartilhado. Esse tipo de distribuição é mais utilizado em sistemas que

precisam cobrir ou enxergar áreas amplas de rastreamento ou captura. Trabalhos como os apresentados em [22] e [14] descreveram esse tipo de distribuição de câmeras.

Neste tipo de distribuição, as câmeras são posicionadas uma em relação a outra de modo de criar uma estrutura semelhante a um grafo conexo, na qual algumas câmeras serão denominadas de câmeras referenciais ou pivôs. Uma câmera pivô é selecionada dependendo de certas características, como compartilhar um maior campo de visão com outras câmeras. Conforme descrito em [14] uma estratégia para posicionar as câmeras é montar um grafo cujos vértices são nossas câmeras e cujas arestas têm um determinado peso gerado por alguma função de erro como as apresentadas na seção 2.1.3.1. O valor do erro permite escolher entre quais arestas temos o menor caminho que relacione todas as câmeras. Esse caminho ajuda a conseguir um menor erro de propagação gerado pelo posicionamento concatenado das câmeras. Essa abordagem mostra mais flexibilidade e tem a vantagem de cobrir um maior campo útil de visão em relação à distribuição das câmeras. Um exemplo desses sistemas numa aplicação de captura de movimento para jogos é mostrada na Figura 14.



Figura 14: Distribuição descentralizada de câmeras.

(Fonte: <http://blogs.ign.com/Blogs/BlogPage.aspx?blog=EA-FIFA-Street&year=2008&month=01>)

As duas distribuições de câmeras descritas acima geram diferenças e problemáticas específicas que nos obrigam a diferenciá-las, especialmente quando é feita a escolha do método certo para a calibração das câmeras.

Terminadas a definição das três etapas que compõem nosso framework conceitual, proposto como parte deste trabalho, e da caracterização da distribuição e do posicionamento de câmeras, também termina a exposição dos conceitos preliminares. Neste capítulo, fizemos a descrição e a revisão de vários conceitos importantes que nos ajudaram a modelar o problema que surge quando necessitamos calibrar um sistema formado por múltiplas câmeras. Esses conceitos serão citados nos próximos capítulos, que descreverão a concepção e a implementação do método de múltipla calibração de câmeras proposto nesta tese.