

Análise comparativa de detectores e descritores de características locais em imagens no âmbito do problema de autocalibração de câmeras

Darlan Nunes de Brito¹
Flávio Luis Cardeal Pádua²
Aldo Peres Campos e Lopes³
Daniel Hasan Dalip²

Resumo: Este trabalho apresenta uma análise comparativa de diferentes métodos do estado da arte para detecção e descrição de características locais em imagens, com o objetivo de solucionar de forma robusta e eficiente o problema de autocalibração de câmeras. Para atingir esse objetivo, é essencial a utilização de métodos detectores e descritores eficazes, uma vez que a correspondência robusta de características em um conjunto de imagens sucessivas sujeitas a uma ampla variedade de distorções afins e mudanças no ponto de vista 3D da cena, é crucial para a exatidão dos cálculos dos parâmetros da câmera. Muito embora diversos detectores e descritores têm sido propostos na literatura, seus impactos no processo de autocalibração de câmeras não foram ainda devidamente estudados. Nesse trabalho de análise comparativa, utilizam-se como critérios de qualidade da autocalibração os erros: epipolar, de reprojeção e reconstrução, bem como os tempos de execução dos métodos. Os resultados experimentais demonstram que detectores e descritores binários de características (ORB, BRISK e FREAK) e de ponto flutuante (SIFT e SURF) apresentam erros de reprojeção e reconstrução equivalentes. Considerando-se, porém, o menor custo computacional dos métodos binários, recomenda-se, fortemente, o uso destes em soluções de problemas de autocalibração de câmeras.

Palavras-chave: Autocalibração de Câmeras, Detectores e Descritores Locais, Correspondência de Pontos em Imagens.

Abstract: *This work presents a comparative analysis of different state-of-art local feature detectors and descriptors aiming to accurately and efficiently solve the camera self-calibration problem. To achieve this goal, it is essential the usage of effective image feature detectors and descriptors, since the robust matching of features in a set of successive images subjected to a substantial range of affine distortion and changes in 3D viewpoint is crucial to the accurate computation of the camera parameters. Even though a lot of detectors and descriptors have been proposed in the literature, their impacts in the camera self-calibration task have not yet been carefully studied. Our comparative analysis uses as criteria the estimated epipolar, reprojection and reconstruction errors, as well as the running times of the methods. Our experiments show that binary (ORB, BRISK and FREAK) and floating-point (SIFT and SURF) feature detectors and descriptors present equivalent reprojection and reconstruction errors. Therefore, considering the smaller computational cost of binary methods, their usage is strongly recommended in camera self-calibration problems.*

Keywords: *Camera Self-Calibration, Local Detectors and Descriptors, Image Matching*

¹Departamento de Computação e Sistema de informação, Universidade Federal de Ouro Preto, 35931-026, João Monlevade, MG, Brasil.
{darlan@decsi.ufop.br}

²Departamento de Computação, Centro Federal de Educação Tecnológica de Minas Gerais, 30421-169, Belo Horizonte, MG, Brasil.
{cardeal,hasan@decom.cefetmg.br}

³Departamento de Matemática, Universidade Federal de Itajubá, 35900-373, Itabira, MG, Brasil
{aldolopes@unifei.edu.br}

1 Introdução

A obtenção dos parâmetros geométricos de câmeras é uma etapa fundamental em muitas aplicações recentes, em visão computacional, tais como robótica [1], reconstrução de cena 3D [2] e formação de imagens panorâmicas [3]. Nas últimas décadas, novos hardwares de captura de imagens têm surgido, tais como câmeras de alta resolução, as quais têm impulsionado o desenvolvimento de aplicações de alto desempenho, fornecendo grande fidelidade geométrica nas tarefas de modelagem tridimensional do ambiente. Para que seja alcançada fidelidade, uma enorme demanda por baixo custo computacional e exatidão é acrescentada aos procedimentos de recuperação dos parâmetros geométricos, forçando, assim, que novas técnicas de recuperação sejam desenvolvidas. Estas técnicas, de acordo com a forma na qual é obtido o conjunto de parâmetros intrínsecos, podem ser divididas em dois grupos diferentes: calibração e autocalibração. Quando é possível obter dados geométricos da cena, como por exemplo distâncias relativas entre pontos, o processo é chamado de calibração; se não há nenhuma informação geométrica prévia sobre a cena 3D, é chamado de autocalibração [4]. Autocalibração oferece a grande vantagem de exigir pouca ou nenhuma interação do usuário. No entanto, essas técnicas são fortemente dependentes de um processo de correspondência de pontos de interesse, na cena[5]. Uma técnica de correspondência ineficaz pode fazer com que os parâmetros recuperados pela autocalibração não tenham nenhum sentido [6].

Este trabalho aborda o problema de avaliação experimental de diferentes métodos do estado da arte para detecção e descrição de características locais em imagens, no que diz respeito a seus impactos na robustez e na eficiência de processos de autocalibração de câmeras. Especificamente, considera-se cenários de autocalibração em que os parâmetros intrínsecos da câmera sejam fixos nas diferentes imagens obtidas a partir de diferentes vistas, tais como aqueles estudados por Loukaris e Deriche [7].

Muito embora diversos métodos para detecção, descrição e correspondência de características em imagens tenham sido propostos nos últimos anos [2], observa-se a ausência de estudos que avaliem os efeitos de suas aplicações no desempenho de estratégias para autocalibração de câmeras. Neste cenário, a principal contribuição deste trabalho é fornecer dados criteriosos de avaliações experimentais quanto aos impactos de alguns dos principais detectores e descritores binários de características (ORB, BRISK e FREAK) e de ponto flutuante (SIFT e SURF) na literatura em processos de autocalibração de câmeras.

Nesse contexto, o método de autocalibração de câmeras utilizado neste trabalho é essencialmente baseado na técnica de Faugeras [8], com modificações propostas por Loukaris e Deriche [7] e Viéville e Lingrand [9]. Portanto este método faz uso de um conjunto simplificado de Equações de Kruppa, consistindo de três passos básicos [10]:

1. Cálculo da matriz fundamental para cada par de vistas da cena, por meio de um processo de correspondência de pontos;
2. Cálculo do dual da imagem da cônica absoluta (DICA) por meio das matrizes fundamentais e das equações de Kruppa;
3. Cálculo da matriz de parâmetros intrínsecos a partir do DICA (processo de autocalibração).

Este artigo está organizado da seguinte forma. Na Seção 2, é apresentada a fundamentação teórica. Os trabalhos relacionados são apresentados na Seção 3. A metodologia utilizada no presente trabalho é apresentada na Seção 4 e os resultados na Seção 5. Finalmente, a conclusão é apresentada na Seção 6.

2 Fundamentação teórica

Nesta Seção são apresentados os conceitos fundamentais para o entendimento do presente trabalho. Especificamente, são brevemente descritos os conceitos referentes à geometria projetiva, ao modelo de câmera *pinhole*, ao plano infinito, à autocalibração e à correspondência de pontos.

2.1 Geometria projetiva

A geometria projetiva é a que modela as transformações perspectivas presentes, quando se captura uma imagem bidimensional de uma cena que é tridimensional. O espaço projetivo de dimensão n , \mathbb{P}^n , é o espaço quociente $\mathbb{R}^{n+1} \setminus \{\mathbf{0}_{n+1}\}$ dado pela relação de equivalência:

$$[x_1, \dots, x_n]^\top \sim [x'_1, \dots, x'_n]^\top \Leftrightarrow \exists \alpha \neq 0, [x_1, \dots, x_n]^\top = \alpha [x'_1, \dots, x'_n]^\top.$$

No espaço projetivo bidimensional, pontos e retas são representados por vetores homogêneos em \mathbb{R}^{2+1} , definidos como sendo pertencentes ao \mathbb{P}^2 . Por exemplo, um ponto x é representado por: $(x_1, x_2, x_3)^\top$ e uma reta l , por (a, b, c) . O mesmo pode ser estendido para pontos e planos no espaço \mathbb{P}^3 , que são representados por vetores homogêneos no \mathbb{R}^{3+1} . Pode-se ver, então, que pontos e hiperplanos são representados por vetores de $n+1$ dimensões, cuja consequência direta é o princípio da dualidade que, em linhas gerais, diz que teoremas construídos para pontos também se aplicam a hiperplanos. Também é característica importante do espaço projetivo a definição do hiperespaço infinito que consiste em pontos que possuem coordenadas $(x_1, \dots, x_n, 0)$.

Outra entidade geométrica importante para autocalibração de câmeras é a cônica absoluta, Ω_∞ , também chamada de cônica de um ponto. Todos os pontos dessa cônica estão localizados no plano infinito, π_∞ , o que faz com que ela pertença a esse plano. Uma cônica absoluta é dada por pontos que satisfazem a Equação (1) e é composta de pontos puramente imaginários:

$$\left. \begin{aligned} X_1^2 + X_2^2 + X_3^2 \\ X_4^2 \end{aligned} \right\} = 0, \quad (1)$$

onde X_1, \dots, X_4 são as coordenadas do ponto \mathbf{X} no sistema de coordenadas cartesianas global (SCCG). Para direções no plano infinito (pontos com $X_4 = 0$), a equação (1) pode ser reescrita conforme a seguir:

$$(X_1, X_2, X_3)I(X_1, X_2, X_3)^\top = 0, \quad (2)$$

onde I é uma matriz identidade 3×3 , que corresponde à cônica absoluta.

Em duas dimensões, o dual da cônica absoluta é dado pela coleção de todas as retas tangentes à cônica; enquanto, no espaço tridimensional, são planos tangentes. A representação na forma canônica do dual da cônica absoluta é dada por:

$$Q_\infty^* = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}.$$

A cônica Ω_∞ , que é uma representação geométrica dos 5 graus de liberdade da matriz de parâmetros intrínsecos da calibração, pode ser utilizada, na recuperação de tais parâmetros, em diferentes modelos de câmeras, entre eles o *pinhole*.

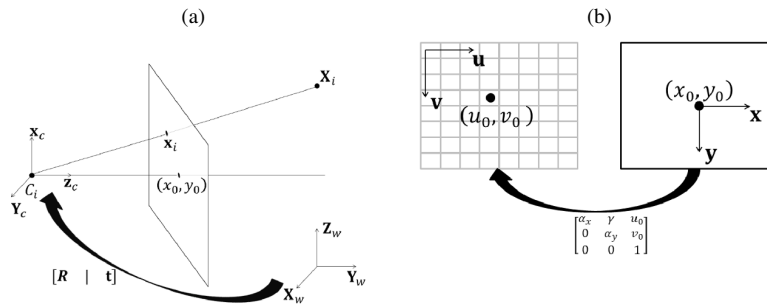
2.2 Modelo de câmera *pinhole*

O modelo mais simples e eficiente de câmera é o *pinhole* que desconsidera parâmetros de distorção introduzidos pela lente da câmera. A representação geométrica desse modelo é mostrada na Figura 1. Nele, o mapeamento de um ponto \mathbf{X} , representado no sistema de coordenadas cartesianas globais SCCG, para o plano de imagem, é dado pela interseção do raio óptico com o plano de imagem, aqui representada por \mathbf{x} . O raio óptico é o segmento de reta que inicia no centro da câmera \mathbf{C} e vai até o ponto da cena \mathbf{X} , conforme ilustrado na Figura 1 (a). Nesse caso, as relações geométricas, na cena, são dadas pelo teorema de Tales:

$$x_i^c = f \frac{X_i^c}{Z_i^c}, \quad (3)$$

$$y_i^c = f \frac{Y_i^c}{Z_i^c}. \quad (4)$$

Figura 1: Transformações em uma cena 3D (a) do sistema de coordenadas cartesianas globais para o sistema de coordenadas da câmera, (b) do plano de imagem para coordenadas em pixels.



Em coordenadas homogêneas a equação de projeção é dada por:

$$\mathbf{x}_i = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_i^c \\ Y_i^c \\ Z_i^c \\ 1 \end{pmatrix}.$$

Há ainda a transformação entre o plano de imagem e a câmera, na qual as coordenadas da origem do plano de imagem (x_0, y_0) devem ser corrigidas para coincidirem com o centro óptico (u_0, v_0) da câmera. Deve-se considerar ainda que, na câmera, as coordenadas são em pixels, cujas dimensões nas direções x e y , bem como o ângulo entre o plano de imagem e o plano da câmera precisam ser observados. Em um modelo completo de câmera, considerando-se P a matriz de projeção, \mathbf{X}_i a representação em coordenadas homogêneas do i -ésimo ponto de um conjunto, no sistema de coordenadas cartesianas globais (SCCG), e \mathbf{x}_i sua projeção no plano de imagem, tem-se que:

$$\mathbf{x}_i = P\mathbf{X}_i. \quad (5)$$

Em que a matriz de projeção P é dada por:

$$P = A[R | \mathbf{t}], \quad (6)$$

sendo A , a matriz de parâmetros intrínsecos, representada por:

$$A = \begin{bmatrix} \alpha_x & \gamma & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix},$$

R é a matriz de rotação e \mathbf{t} o vetor de translação, ambos entre o sistema de coordenadas de referência da câmera e o SCCG. A concatenação de R e \mathbf{t} conforme exposto na Equação 6, define a chamada matriz de parâmetros extrínsecos da câmera. A matriz de parâmetros extrínsecos é dada pela concatenação da matriz de rotação R e o vetor de translação. A matriz de parâmetros intrínsecos é uma matriz triangular superior, cujos termos $\alpha_x = fm_x$ e $\alpha_y = fm_y$ são a distância focal em termos da dimensão, em pixels, nas direções x e y , respectivamente, e m_x e m_y representam os números de pixels por unidade métrica de distância. Similarmente x_0 e y_0 representam as coordenadas da origem do plano de imagem em pixels nas direções x e y , sendo que $x_0 = m_x p_x$ e $y_0 = m_y p_y$ nos quais p_x e p_y representam as coordenadas do centro óptico da câmera. No conjunto de parâmetros intrínsecos há ainda o ângulo, chamado de oblíquo, que representa o ângulo entre o eixo óptico e o plano de imagem. Esse parâmetro, para aplicações de uso geral, utilizando-se câmeras, pode ser considerado zero [11]. Como descrito anteriormente, a recuperação automática desse conjunto de parâmetros é chamada de autocalibração.

2.3 Plano infinito

Alguns dos principais métodos de autocalibração [12], [8], [13], de câmeras utilizam as equações de Kruppa baseadas no cálculo do dual da imagem da cônica absoluta (DICA), uma cônica no plano infinito. Este plano pode ser definido pelo vetor $\pi_\infty = [0, 0, 0, 1]$; sendo que os pontos que pertencem a esse plano possuem coordenadas homogêneas $X_\infty = (\mathbf{d}^\top, 0)$ e são projetados de acordo com a Equação:

$$\mathbf{x} = P\mathbf{X}_\infty = A[R | \mathbf{t}] \begin{pmatrix} \mathbf{d}^\top \\ 0 \end{pmatrix} = AR\mathbf{d}.$$

Essa equação mostra que o mapeamento, entre π_∞ e o plano de imagem, é dado por uma homografia H como a seguir:

$$H = AR.$$

Verifica-se, então, que este mapeamento é independente da posição da câmera, dependendo apenas dos parâmetros intrínsecos e da orientação com respeito ao SCCG. Considerando-se que pontos são mapeados por H ($\mathbf{x} \mapsto H\mathbf{x}$), uma cônica C é mapeada de acordo com $C \mapsto H^{-\top}CH^{-1}$. Fazendo-se $C = \Omega_\infty = I$ no plano infinito, a imagem ω da cônica absoluta pode ser obtida por meio de álgebra simples, fazendo-se algumas manipulações:

$$\omega = (AA^\top)^{-1}.$$

Já que ω não possui pontos reais, pode-se, embasado no princípio da dualidade, usar a imagem do dual da cônica absoluta Q_∞ que é ω^* definida como:

$$\omega^* = \omega^{-1} = AA^\top.$$

Esta é a cônica que será utilizada no processo de autocalibração.

2.4 Autocalibração

No primeiro passo para a autocalibração de uma câmera, usualmente, determina-se um conjunto de parâmetros que compõem a chamada matriz fundamental [11]. Por meio de pontos correspondentes em imagens de diferentes vistas de uma cena, é possível recuperar esta matriz que é 3×3 com posto igual a 2 e espaço nulo, representado pelos epípolos [11].

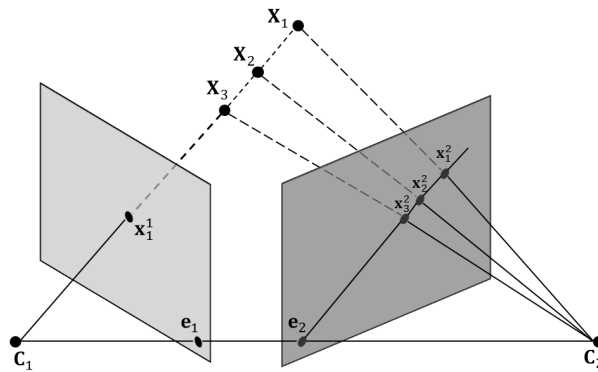
Dadas duas imagens I e I' , representativas de uma cena de interesse, a matriz fundamental F é responsável por projetar pontos da imagem I , em retas na imagem I' e vice-versa. Isso restringe os pontos correspondentes nesta reta. Essa relação é descrita por meio da Equação:

$$\mathbf{x}'_i F \mathbf{x}_i = 0.$$

A reta, acima mencionada, é chamada reta epipolar e passa por um ponto específico no plano de imagem, o epípolo; que é a projeção dos chamados centros das câmeras [11], conforme ilustrado na Figura 2. A geometria epipolar é a geometria intrínseca da visualização dupla. Por meio dessa restrição, é possível utilizando-se técnicas estatísticas como RANSAC, a determinação da matriz fundamental por meio da correspondência de pontos de interesse em imagens. A geometria epipolar, para visualização dupla, pode ser estendida para visualização tripla ou múltipla, facilmente, considerando a matriz fundamental entre cada par de visualização.

As equações de Kruppa, usadas no cálculo da matriz de parâmetros intrínsecos, foram introduzidas por Faugeras e Maybank [8]. Estas equações são uma representação algébrica da correspondência das retas epipolares tangentes ao dual da imagem da cônica absoluta. Considerando-se a imagem do dual de uma cônica absoluta no

Figura 2: Geometria epipolar.



plano infinito ω^* , a matriz fundamental F e o epipolo e' , é possível representar as relações entre as retas epipolares tangentes ao dual da imagem da cônica absoluta pela Equação (7):

$$\begin{bmatrix} e' \\ \times \end{bmatrix} \omega^* \begin{bmatrix} e' \\ \times \end{bmatrix} = F\omega^*F^T, \quad (7)$$

na qual $\begin{bmatrix} e' \\ \times \end{bmatrix}$ é a matriz *skew-symmetric* obtida a partir do epipolo da imagem I' . O dual da imagem da cônica absoluta ω^* é derivada da matriz de parâmetros intrínsecos de acordo com a fórmula $\omega^* = AA^T$; sendo ω^* a imagem da cônica associada à imagem I e ω^* , à imagem I' . Considerando os parâmetros intrínsecos constantes, tem-se $\omega^* = \omega^*$, determinando-se, assim, forma original da equação apresentada em [9], $\begin{bmatrix} e' \\ \times \end{bmatrix} \omega^* \begin{bmatrix} e' \\ \times \end{bmatrix} = F\omega^*F^T$. Uma forma sucinta e mais utilizável das equações de Kruppa [6], considerando o espaço nulo de $\begin{bmatrix} e' \\ \times \end{bmatrix}$, comum em ambos os lados na Equação (7), é dada pela Equação (8):

$$\begin{pmatrix} \mathbf{u}_2^T \omega^* \mathbf{u}_2 \\ -\mathbf{u}_1^T \omega^* \mathbf{u}_2 \\ \mathbf{u}_1^T \omega^* \mathbf{u}_1 \end{pmatrix} \times \begin{pmatrix} \sigma_1^2 \mathbf{v}_1^T \omega^* \mathbf{v}_1 \\ \sigma_1 \sigma_2 \mathbf{v}_1^T \omega^* \mathbf{v}_2 \\ \sigma_2^2 \mathbf{v}_2^T \omega^* \mathbf{v}_2 \end{pmatrix} = 0, \quad (8)$$

em que \mathbf{u}_i , \mathbf{v}_i e σ_i são as colunas e valores singulares da decomposição em valores singulares (SVD) da matriz fundamental F . A matriz ω^* deverá ser simétrica, definida positiva. Assim, por meio da decomposição de Cholesky, é possível recuperar a matriz de parâmetros intrínsecos A . Com duas vistas diferentes de uma mesma cena, pode-se determinar uma equação, como a Equação (8), que produz quatro equações quadráticas nos termos de ω . Dessas, duas são independentes, levando a duas restrições nos parâmetros de ω . A matriz ω é simétrica, definida positiva até uma escala, possuindo então, 5 termos independentes precisando de, no mínimo, três vistas diferentes para se obter 6 restrições sobre os 5 parâmetros. Com essas restrições é possível determinar os termos de ω unicamente.

Com o conhecimento das matrizes fundamental e de parâmetros intrínsecos, é possível, até uma certa escala, determinar as posições relativas dos objetos da cena, em relação a um SCCG, coincidindo com o sistema de coordenadas de uma vista qualquer. Os parâmetros responsáveis por esta transformação são agrupados na matriz de parâmetros extrínsecos. Para recuperação dessa matriz, pode ser utilizada uma técnica embasada na matriz Essencial, que é a especialização da matriz fundamental para o caso de coordenadas normalizadas. Por coordenadas normalizadas entende-se o caso em que as matrizes de projeção são dadas por $P = [I \mid 0]$ e $P' = [R \mid \mathbf{t}]$. A matriz essencial é dada pela relação entre a matriz de rotação e o vetor de translação entre as imagens como mostrado na Equação 9.

$$E = [\mathbf{t}]_{\times} R. \quad (9)$$

Se as matrizes fundamental e de parâmetros intrínsecos são conhecidas, é possível determinar a matriz

essencial, E , por meio da Equação 10:

$$E = A' F * A. \quad (10)$$

A partir da decomposição em valores singulares da matriz essencial, pode-se determinar a matriz de parâmetros extrínsecos para cada uma das vistas, caso em que as matrizes de rotação e translação são consideradas fixando-se o SCCG em uma vista qualquer. Adicionalmente ao processo de autocalibração pode-se realizar, então, a reconstrução tridimensional que neste trabalho foi usada como meio de verificação da qualidade dos parâmetros intrínsecos calculados. Para finalizar o processo de reconstrução, pode-se criar um mapa denso de pontos tridimensionais da cena de interesse, usando a técnica de triangulação descrita em [11].

2.5 Correspondência de pontos

O passo inicial crítico nas técnicas de autocalibração é a determinação da matriz fundamental. Essa matriz é obtida por meio da correspondência de pontos de interesse nas diferentes vistas de uma cena tridimensional. O processo de corresponder pontos em imagens pode ser visto, como a determinação da correlação entre pontos de imagens diferentes, que correspondam ao mesmo ponto da cena tridimensional. Os métodos de correspondência de pontos podem ser divididos em quatro etapas: detecção, descrição, correspondência e refinamento. As etapas de correspondência e refinamento são realizadas com técnicas simples, mas eficientes, como distância de Hamming ou norma L2 com busca exaustiva para correspondência e RANSAC para refinamento. Diversas técnicas para detecção e descrição têm sido propostas nos últimos anos. Entre tais técnicas, destacam-se: *Scale Invariant Feature Transform* (SIFT) [14], *Speeded Up Robust Features* (SURF) [15], *Oriented Fast and Rotated Brief* (ORB) [16] e *Binary Robust Invariant Scalable Keypoints* (BRISK) [17], como detectores e descritores; e *Fast Retina Keypoint* (FREAK) [18] apenas como descritor.

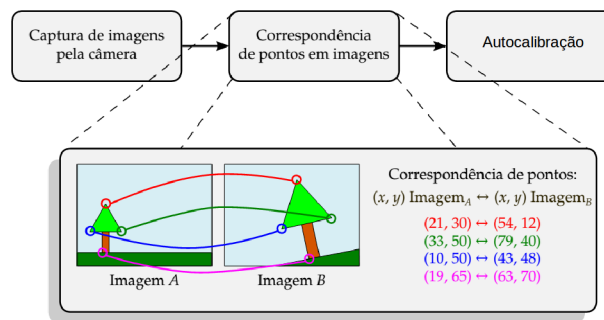
Os algoritmos de correspondência de pontos podem ser divididos em dois grandes grupos: algoritmos baseados em área e algoritmos baseados em características [19]. Os algoritmos baseados em área formam um mapa denso de pontos correspondidos, enquanto os algoritmos baseados em características formam um mapa esparsos de características. Os algoritmos baseados em características possuem uma série de vantagens: são mais rápidos, porque consideram menos pontos; a correspondência é mais exata já que vértices podem ser localizados com precisão de subpixel e, por fim, são menos sensíveis a variações fotométricas já que representam propriedades geométricas da cena [20]. Essas vantagens justificam uma utilização muito maior deste tipo de método para correspondência de pontos em imagens e determinaram, portanto, a escolha do estudo de tais métodos neste trabalho.

Os algoritmos podem ser agrupados também, de acordo com o tipo de descritores: descritores de ponto flutuante (que utilizam números de ponto flutuante) e descritores binários (que utilizam bits). Atualmente, os principais algoritmos de correspondência de pontos baseados em características e em descritores de ponto flutuante são SIFT e SURF. Entre os descritores binários baseados em características, destacam-se os métodos ORB e BRISK. Os algoritmos SIFT e SURF fazem uma abordagem estatística por meio de funções Gaussianas para determinar uma pirâmide de escalas. Assim, seus descritores são invariantes à escala e à rotação. Os algoritmos ORB, BRISK e FREAK, por outro lado, fazem parte de uma série de técnicas com descritores binários, que têm como principal objetivo, a melhoria no desempenho computacional dos algoritmos de correspondência de pontos. Eles possuem, na etapa de descrição, variações do detector de vértices FAST [17], que consiste em um método de detecção com um desempenho superior aos métodos que se baseiam em funções Gaussianas.

Frequentemente, os algoritmos de correspondência de pontos são avaliados quanto ao seu desempenho. Apesar de as técnicas de autocalibração e, conseqüentemente, de reconstrução tridimensional possuírem muita importância para um grande número de aplicações em visão computacional [5], avaliações são comumente focadas em aplicações como reconhecimento de objetos 3D [21, 22], recuperação de imagens [23] e classificação de texturas [24].

Nesse contexto, este trabalho propõe realizar a avaliação de métodos detectores e descritores de pontos correspondentes em imagens. Na Figura 3, é mostrado um diagrama contendo as etapas básicas de um processo de autocalibração. Neste processo, a etapa inicial consiste em obter um conjunto de imagens a partir da câmera com diferentes ângulos de visualização do ambiente. Na etapa seguinte é realizada a correspondência de pontos de interesse nestas imagens. A partir dos pontos correspondidos é possível determinar a matriz fundamental e,

Figura 3: Etapas básicas de um processo de autocalibração.



Fonte: do próprio autor.

com essa, o conjunto de parâmetros intrínsecos. A última etapa consiste no processo de autocalibração em si. Nesse diagrama está destacada a etapa de correspondência de pontos, objeto principal de estudo e avaliação neste trabalho.

3 Trabalhos relacionados

Avaliações de desempenho têm cada vez mais importância em visão computacional [25], haja vista a quantidade de trabalhos nesta área. Dentre esses, muitos trabalhos surgiram com objetivo de avaliar algoritmos de correspondência de pontos em imagens, por meio de diferentes estratégias e métricas avaliativas [25].

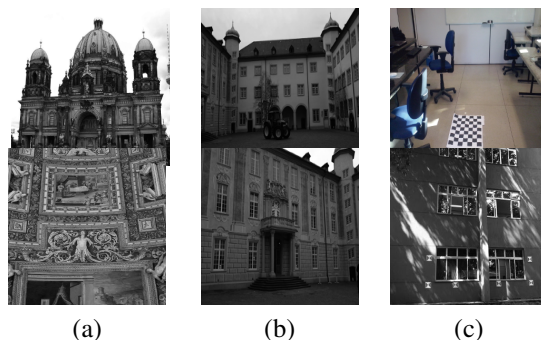
Jiang, Xu e Liu [26] realizam uma avaliação bastante criteriosa de algoritmos detectores e descritores, dos quais apenas um é binário, o ORB; os autores, usando imagens de bases de dados da internet e por eles obtidas, fornecem dados sobre o desempenho dos algoritmos utilizados no trabalho e uma série de métricas para avaliação de desempenho, consistentes com o estado da arte de avaliações em algoritmos de correspondência de pontos em imagens. Naquele trabalho foi analisado o desempenho em relação à odometria baseada em visão estéreo.

Quatro são os métodos de correspondência testados pelos autores Ahmadabadian et al. em [2] para determinação de um mapa de pontos denso com alta exatidão: duas versões do método *matching semi-global MicMac* [26] e duas versões do método *Patch-based Multi-View Stereo (PMVS)* [26]. Os pontos correspondentes são utilizados como entrada para um programa responsável pela obtenção da nuvem densa de pontos, um software chamado Bundler, desenvolvido por Snavely, Seitz e Szeliski [27]. Os resultados mostraram que a combinação de técnicas de *bundle adjustment* podem produzir resultados equivalentes ao scanner a laser, caso câmeras com uma alta qualidade (alta resolução) sejam utilizadas.

Dwarakanath et al. [25] fazem uma avaliação dos algoritmos SIFT, SURF e ORB para sistemas de autocalibração, baseados em múltiplas visualizações de uma cena 3D. Neste artigo, são considerados o erro epipolar em pixel e o tempo de processamento de cada algoritmo como métricas avaliativas. Os experimentos são realizados utilizando-se um conjunto de imagens reais de uma base de dados disponibilizada na internet, com modificações nas imagens simuladas por um módulo, por eles denominado Módulo de Degradação. Nesse caso, considerou-se que o erro da geometria epipolar é suficiente para determinar o desempenho do sistema de reconstrução tridimensional. De fato, o erro epipolar é um importante parâmetro de qualidade para reconstrução tridimensional da cena por meio de câmeras. Mais adiante ver-se-á que, nos resultados aqui apresentados, avaliar tais métodos considerando-se apenas o erro epipolar como métrica, pode levar ao negligenciamento de informações relevantes tais como o custo computacional e o erro de projeção, pois essas informações são importantes para a autocalibração e consequentemente reconstrução tridimensional eficiente.

Algumas avaliações são feitas pelos autores Li e Wang em [28] para que se tenha uma visão geral sobre técnicas de correspondência de pontos em imagens para vários sistemas de navegação baseados em visão. Naquele trabalho foram avaliados os pontos fracos e fortes e apresentada uma modificação no algoritmo SIFT, chamada ASIFT, para lidar com grandes mudanças no ponto de vista da câmera.

Figura 4: Exemplos de imagens das bases de dados obtidas: (a) em Heinly, Dunn e Frahm [29], (b) em Strecha et al. [30] e (c) neste trabalho.



A principal contribuição deste trabalho consiste em apresentar resultados de análise comparativas de métodos de detecção e descrição de características locais em imagens, no que diz respeito aos seus impactos no processo de autocalibração de câmeras, fornecendo assim uma valiosa base de informação capaz de nortear as escolhas dos melhores métodos a serem utilizados em uma dada aplicação. A seguir é descrita a metodologia utilizada por este trabalho.

4 Metodologia

Este trabalho avalia os impactos dos algoritmos de detecção e descrição de características locais utilizados pelos métodos SIFT [14], SURF [15], ORB [16] e BRISK [17], bem como do algoritmo de descrição utilizado pelo método FREAK [18]. Os algoritmos de tais métodos foram escolhidos pelo fato de representarem soluções do estado da arte no que se refere ao problema de correspondência de pontos e serem, amplamente conhecidos e aplicados em diversos problemas no campo da visão computacional.

Neste estudo foram utilizadas diversas sequências de imagens de duas bases de dados, disponibilizadas publicamente na internet e usadas nos trabalhos anteriores de Heinly, Dunn e Frahm [29] e Strecha et al. [30]. No que se refere à base de dados disponibilizada pelos autores em [29] tem-se duas sequências, uma contendo onze imagens, outra contendo oito imagens, em ambos os casos, imagens com resolução de 1024×683 . Já na base de dados disponibilizada em [30], tem-se duas sequências, uma contendo dez imagens e outra trinta imagens, em ambos os casos com resolução de 3072×2048 .

Adicionalmente, foi criada uma base de dados especialmente para este trabalho, caracterizada por cenários em ambientes internos e externos, capturados com uma câmera CANON, modelo EOS-70D. Nessa base de dados, tem-se oito sequências, contendo oito imagens cada uma, com resolução 2736×1824 . Todos os experimentos foram realizados com a utilização de um computador, com processador Intel Core i5-2430M, 4GB de memória RAM e sistema operacional Linux, versão 14.04 LTS para execução dos programas. A Figura 4 mostra exemplos de imagens utilizadas para cada uma das sequências de experimentos. Em todos experimentos deste trabalho foram usadas matrizes de projeção de referência (*ground truth*), como meio de avaliar a exatidão do algoritmo de autocalibração aplicado. Heinly, Dunn e Frahm [29] usam um software desenvolvido por Snavely, Steven e Szeliski [27] para determinar as matrizes de projeção de referência. Já Strecha et al. [30] utilizam um sensor do tipo LIDAR para obter um conjunto de medidas de referência (*ground truth*). Nesse caso, marcos são espalhados pela cena com o objetivo de realizar a calibração e obtenção das matrizes de projeção.

O primeiro conjunto de experimentos contendo imagens de cenas planares e não planares foi realizado com a base de dados especialmente criada para este estudo. Para isso, um alvo de calibração do tipo tabuleiro de xadrez serviu para apoiar a calibração das câmeras, viabilizando-se a obtenção das matrizes de projeção de referência e os pontos correspondentes nas imagens, bem como a determinação dos erros de reprojeção, epipolar, e de reconstrução. Para cada cena planar e não planar visualizada, com e sem a presença do alvo de calibração, cinco imagens de pontos de vista diferentes foram especialmente obtidas. Em especial, aplicou-se o método de Zhang [31] sobre as imagens, contendo os alvos de calibração, para determinar as matrizes de projeções das câmeras.

Para a sequência de experimentos realizada com a base de dados criada em ambientes externos contendo cenas não planares, foram realizadas as calibrações dos parâmetros intrínsecos da câmera, utilizando-se novamente alvos de calibração do tipo tabuleiro de xadrez e a técnica proposta em [31]. Na cena, foram adicionadas marcações para determinação dos parâmetros extrínsecos. Nessa etapa dos experimentos foram capturadas oito imagens de pontos de vista diferentes, com pequeno intervalo de tempo para minimizar possíveis variações de luminosidade da cena. Para as bases de dados de Heinly, Dunn e Frahm [29] e Strecha et al. [30], as matrizes de projeção de referência foram fornecidas pelos autores e os pontos correspondentes de referência nos pares de imagens foram obtidos visualmente, utilizando uma técnica descrita por Mikolajczyk e Schmid [32].

Extratores de características foram usados para estimar as posições nas imagens de pontos correspondentes. Pontos estes que podem estar fora da reta epipolar causando um erro chamado de erro epipolar. O erro epipolar E_p é dado pela Equação (11) conforme descrito por [25]:

$$E_p = \sum_{i=1}^{N_p} \frac{(\hat{\mathbf{x}}'_i F \hat{\mathbf{x}}^i)}{(F \hat{\mathbf{x}}_i)_1^2 + (F \hat{\mathbf{x}}_i)_2^2 + (F^T \hat{\mathbf{x}}'_i)_1^2 + (F^T \hat{\mathbf{x}}'_i)_2^2} \quad (11)$$

Em que $\hat{\mathbf{x}}_i$ e $\hat{\mathbf{x}}'_i$ são pontos correspondidos, estimados pelo método de correspondência de pontos que está sendo avaliado, nas imagens I e I' , respectivamente; e F é a matriz fundamental de referência. Os pontos nas imagens I e I' , obtidos dos alvos de calibração com uma precisão de subpixel, bem como a técnica de decomposição em valores singulares SVD [11], foram usados para se estimar a matriz fundamental de referência nos casos de cenas planares. Para os casos de cenas não planares, incluindo as imagens obtidas de bases de dados da internet, foi utilizada a mesma forma de cálculo do erro em pixel. No entanto, as correspondências corretas foram obtidas visualmente, mas com um erro de subpixel, por meio da técnica descrita por Mikolajczyk e Schmid [32].

Além do erro epipolar, foi determinado o erro de reprojeção dos pontos E_r . Empregando-se os pontos correspondidos e a matriz de projeção foi feita a triangulação, denotada por τ e descrita esquematicamente pela Equação (12), conforme a técnica descrita por Hartley e Zisserman [11]:

$$\mathbf{X} = \tau(\mathbf{x}, \mathbf{x}', P, P'). \quad (12)$$

Em que \mathbf{X} é o ponto no SCCG em coordenadas homogêneas; \mathbf{x} e \mathbf{x}' , pontos correspondidos nas imagens I e I' , também em coordenadas homogêneas, P e P' , matrizes de projeção das imagens. Os pontos no SCCG foram novamente projetados nos planos de imagem I e I' utilizando-se as matrizes de projeção de referência e a Equação (5). A diferença entre os pontos projetados e os pontos estimados é o erro de reprojeção E_r , descrito pela Equação (13):

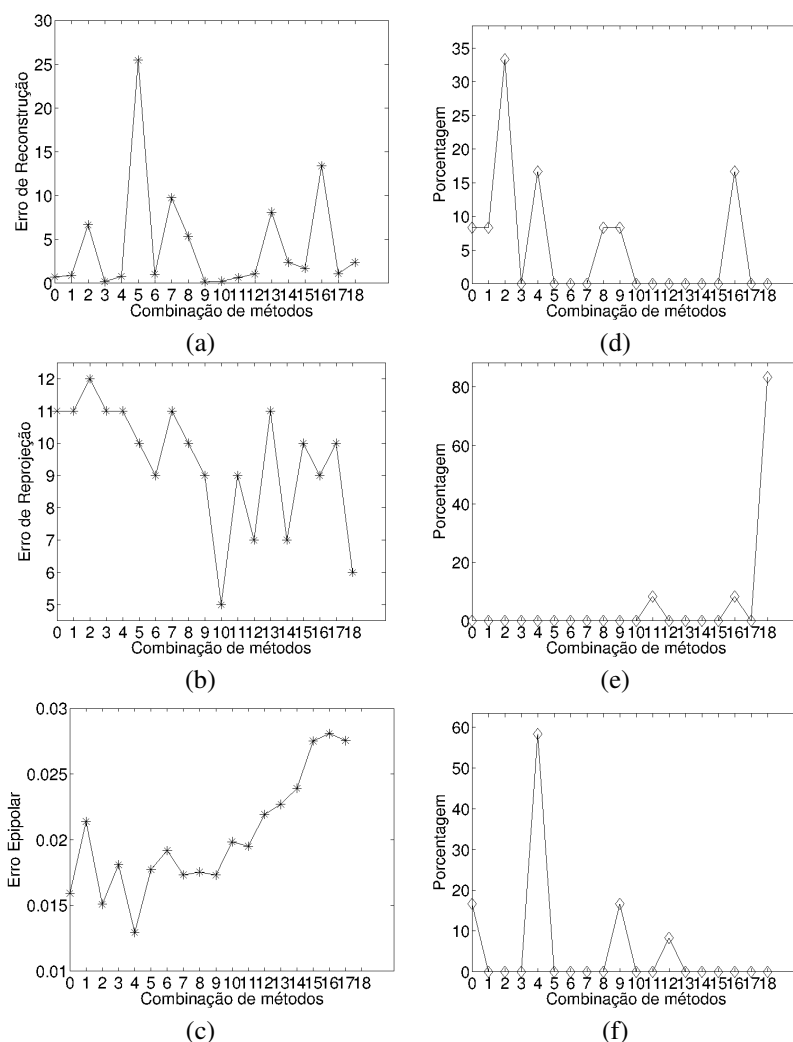
$$E_r = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\mathbf{x}}_i) + (\mathbf{x}'_i - \hat{\mathbf{x}}'_i). \quad (13)$$

Outra métrica utilizada foi o erro de reconstrução E_{SCCG} , representado pela diferença entre a estrutura reconstruída e a estrutura de referência. Esse erro foi calculado utilizando as diferenças das distâncias euclidianas entre os pontos de referência \mathbf{X}_i e os pontos estimados $\hat{\mathbf{X}}_i$ conforme a Equação (14):

$$E_{SCCG} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^i \left(d(\mathbf{X}_i, \mathbf{X}_j) - d(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j) \right)^2 \quad (14)$$

Nas Equações (11), (12), (13) e (14), utilizou-se diferentes conjuntos de dados com o objetivo de definir quais são os melhores algoritmos detectores e descritores de pontos para um sistema computacionalmente eficiente e robusto de autocalibração. Os dados utilizados foram matrizes de projeção estimadas, baseadas nas matrizes fundamentais; e matrizes de projeção de referência obtidas pela calibração, conforme o método de [31] ou fornecidas pelos autores nos casos das bases de dados da internet. A triangulação foi feita tanto com as matrizes de projeção de referência, quanto com as matrizes de projeção estimadas.

Figura 5: Erro quadrático médio normalizado: (a) erro de reconstrução, (b) erro de reprojeção e (c) erro epipolar, para cada combinação de detector e descritor. Porcentagem de vezes na qual o algoritmo detector-descritor aparece como o método que retornou o menor erro quadrático médio normalizado: (d) reconstrução, (e) de reprojeção e (f) epipolar.



5 Resultados experimentais

Nesta Seção são apresentados os resultados, utilizando os três conjuntos de imagens para os quais se conhecia as matrizes de projeção correspondentes. São eles: a base criada no presente trabalho e as bases extraídas do trabalho de Heninly et al. (2012) e Strecha et al. (2008) [29, 30]. Nos gráficos das Figuras 5 (a) e 5 (b), no eixo das abscissas, são vistos apenas os índices dos pares de métodos detector-descritor combinados e avaliados. A Tabela 1 identifica o índice de acordo com os métodos detectores e descritores utilizados.

Nas Figuras 5 (a) e 5 (b) são apresentados os resultados para os erros quadráticos médios de reconstrução e reprojeção, respectivamente, normalizados pela norma dos dados. Nessas Figuras a combinação que apresentou os menores erros em ambos os casos, utiliza o ORB como detector, sendo no primeiro, o SIFT como descritor e no segundo o SURF (combinações 9 e 10, respectivamente), para os erros de reconstrução e reprojeção respectivamente. Já para o erro epipolar, apresentado no gráfico da Figura 5 (c), a combinação que retornou o menor erro médio é aquela que utiliza como detector o SURF e como descritor o SIFT (combinação 4).

Considerando-se o gráfico apresentado na Figura 6 (a), em 58% dos casos, combinações que utilizavam o algoritmo SURF como detector apresentou o menor erro epipolar. Ainda na mesma Figura, pode-se ver que a

Tabela 1: Combinações de algoritmos de detecção e descrição de características avaliados nos experimentos.

Índice	Detector	Descritor	Índice	Detector	Descritor	Índice	Detector	Descritor
0	SIFT	SIFT	6	SURF	ORB	12	ORB	BRISK
1	SIFT	SURF	7	SURF	BRISK	13	ORB	FREAK
2	SIFT	BRISK	8	SURF	FREAK	14	BRISK	SIFT
3	SIFT	FREAK	9	ORB	SIFT	15	BRISK	SURF
4	SURF	SIFT	10	ORB	SURF	16	BRISK	ORB
5	SURF	SURF	11	ORB	ORB	17	BRISK	BRISK
						18	BRISK	FREAK

combinação com descritor BRISK foi a que apresentou o menor erro epipolar na maior parte dos casos, 41,67%. É possível observar também na Figura 6 (a) que a combinação que retornou menos vezes o menor erro epipolar é aquela que o detector utilizado foi o SIFT. Já para o descritor, as combinações que utilizam os algoritmos SURF e ORB apresentaram o mesmo percentual de erro. Já, considerando o erro de reprojeção, mostrado no gráfico da Figura 6 (c), o BRISK é o detector que apresenta o menor erro – 91,67% dos casos, com o SIFT visto como o melhor descritor com o menor erro em 41,67% dos casos. Considerando-se o erro de reconstrução, Figura 6 (b), o algoritmo que apresenta o menor erro do detector é o algoritmo SIFT, seguido pelo SURF, com 50% e 25% dos casos, respectivamente. Os descritores com o número maior de vezes que retornaram o menor erro de reconstrução foram o SIFT e o SURF que apresentaram os percentuais de 42,67% e 16,67% respectivamente, sendo que os descritores binários BRISK e FREAK apresentaram o mesmo resultado. É importante destacar que, para os gráficos das Figuras 6 (a), 6 (b) e 6 (c), o método FREAK não tem reportado resultados como detector, visto que é um método exclusivamente desenvolvido para descrição de características.

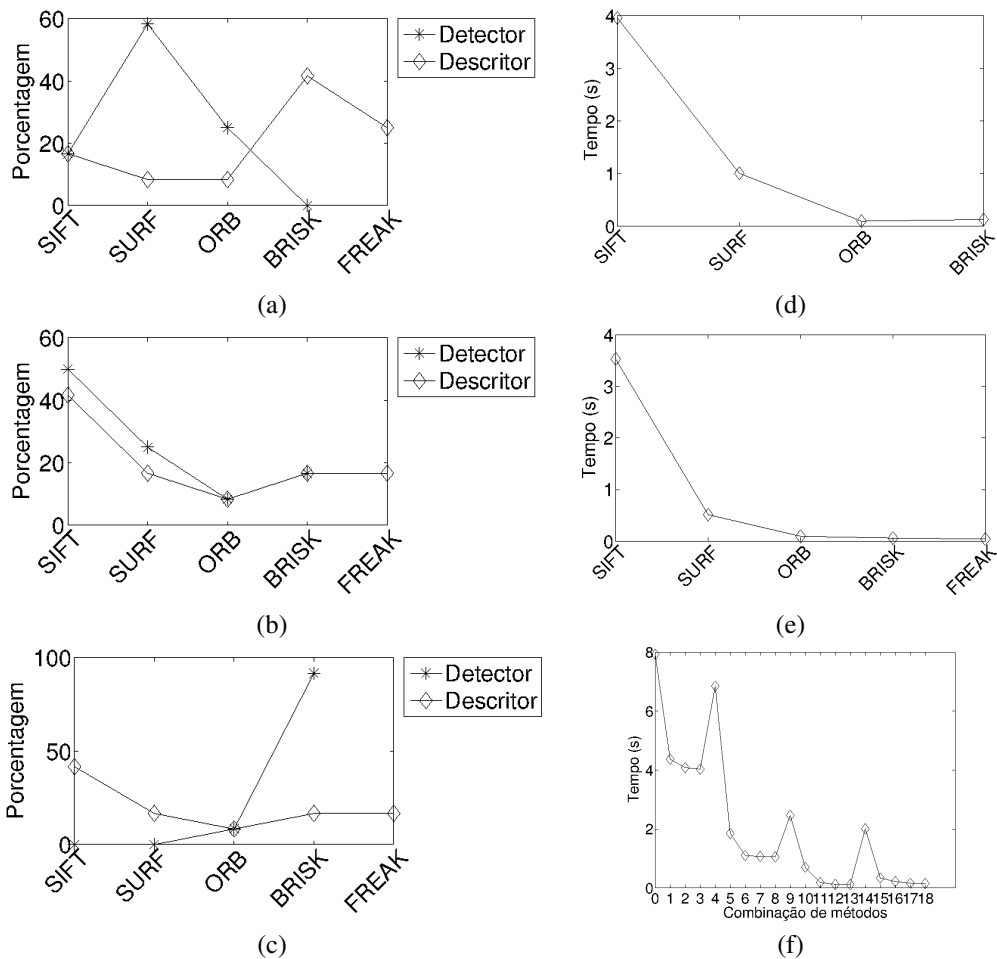
Tendo em vista não mais separadamente, mas o conjunto detector-descritor, verifica-se que a combinação SURF-SIFT (combinação 5) retornou o menor erro epipolar em 58,33% dos casos, como observado na Figura 5 (f). Na mesma Figura, observa-se que o par SIFT-SIFT apresenta o segundo melhor desempenho com o menor erro em 16,67% – mesmo desempenho do ORB-SIFT.

Na Figura 5 (e), pode-se notar que o BRISK-FREAK apresentou o menor erro de reprojeção, em 83,33%, o que, portanto, representa uma grande diferença em relação aos outros métodos. Já para o erro de reconstrução, que é o foco principal deste trabalho, há uma distribuição mais uniforme dos métodos com descritores e detectores binários como BRISK-ORB (combinação 16), apresentando o mesmo resultado que métodos de ponto flutuante como SURF-SIFT (combinação 5) como mostrado na Figura 5 (d). No mesmo gráfico, nota-se que o par detector-descritor que apresentou o melhor resultado foi SIFT-BRISK (combinação 2). Ele retornou o menor erro em 33,33% dos casos.

Como esperado, os tempos de execução para os algoritmos que usam descritores de ponto flutuante foram muito maiores do que o dos algoritmos com descritores binários. Os algoritmos detectores binários apresentaram em média 2,76% do tempo de execução do algoritmo SIFT e 10,90% do algoritmo SURF como pode ser observado na Figura 6 (d). Já para os algoritmos descritores, as diferenças em relação aos tempos de execução foram 1,84% e 10,41% para os algoritmos de SIFT e SURF respectivamente como se observa na Figura 6 (e). Outro importante gráfico sobre o tempo de execução pode ser visto na Figura 6 (f). Nesse, cada ponto representa o tempo total de detecção e descrição de cada combinação e por ele pode-se ver que as combinações com detectores binários são mais rápidas do que combinações com detectores de ponto flutuante, ainda que seja utilizado um descritor binário. Observa-se também, neste mesmo gráfico, que métodos com descritor e detector binários são mais rápidos do que aqueles que utilizam números de ponto flutuante SIFT e SURF. Por meio do gráfico da Figura 6 (f), observa-se ainda que combinações, que utilizam o descritor FREAK são sempre mais rápidas como se podemos observar a partir dos resultados associados às combinações 3, 8, 13 e 18.

Observa-se que os erros de reconstrução dos algoritmos binários ORB, BRISK e FREAK, sejam como detectores, descritores ou detectores e descritores ao mesmo tempo, foram equivalentes aos erros de reconstrução dos algoritmos SIFT e SURF independentemente de serem detectores ou descritores. Adicionalmente, apresentaram tempos de processamento consideravelmente menores. Sugere-se, então, em sistemas onde há restrições ao tempo de execução, que os algoritmos ORB, BRISK e FREAK sejam escolhidos como potenciais métodos a serem utilizados em problemas de autocalibração de câmeras e conseqüentemente, reconstrução tridimensional.

Figura 6: Porcentagem de vezes que os algoritmos de detecção e descrição retornam o menor erro quadrático médio normalizado: (a) erro epipolar, (b) erro de reconstrução e (c) erro de reprojeção. Tempos de execução dos algoritmos detectores (d), descritores (e) e da soma dos tempos do detector e descritor para cada método, (f).



6 Conclusões

O presente trabalho avaliou diferentes métodos do estado da arte para correspondência de pontos de interesse em imagens, no que diz respeito a seus impactos na robustez e eficiência de um problema de autocalibração de câmeras. Trabalhos recentes, como [25] e [26], apresentam avaliações de desempenho, cuja preocupação única, é determinação do algoritmo que retorne o menor erro na localização dos pontos sem ponderar o desempenho dos algoritmos e não fazem uma avaliação, em um sistema implementado de reconstrução tridimensional, como neste trabalho. Os resultados obtidos demonstram que a avaliação da robustez e eficiência de algoritmos detectores e descritores de características no contexto de autocalibração de câmeras deve-se basear não somente no erro da geometria epipolar mas também dos erros de reconstrução e reprojeção.

A análise dos dados mostrou que os erros no SCCG são moderados e linearmente correlacionados com os erros de reprojeção, apresentando um coeficiente de correlação de Pearson igual a 0,20. Isso indica que o algoritmo detector-descritor pode funcionar bem para reconhecimento de estruturas em diferentes imagens, mas produzir um resultado ruim quando o objetivo é a reconstrução de objetos da cena. Além disso, avaliar algoritmos de correspondência de pontos por meio de métricas, baseadas apenas nos pontos das imagens, não é uma estratégia adequada. É fundamental a implementação do algoritmo de reconstrução tridimensional.

Pode-se concluir, a partir do presente trabalho, que algoritmos de correspondência binários, além de possuírem custo computacional muito menor do que algoritmos de ponto flutuante, produzem resultados muito satisfatórios, algumas vezes até equivalentes para autocalibração de câmeras, produzindo, assim, uma reconstrução

tridimensional de objetos da cena com um menor erro.

Como trabalhos futuros, pretende-se desenvolver uma nova metodologia de autocalibração, que seja mais eficiente e robusta, no sentido de produzir erros algumas ordens de grandeza menores do que os dados e que possa ser executado em um sistema de autocalibração de tempo real.

Agradecimentos

Os autores agradecem o suporte financeiro do CNPq (Procs. 468042/2014-8 e 313163/2014-6); FAPEMIG (Procs. APQ-01180-10 e APQ-02269-11); CEFET-MG (Procs. PROPESQ-088/12 e PROPESQ-076/09); CAPES e à FUFOP – Fundação Universitária Federal de Ouro Preto.

Referências

- [1] YU, H. et al. Visual Localization for Mobile Robots Based on Composite Map. *Journal of Robotics and Mechatronics*, v. 25, n. 1, p. 25–37, 2013. Disponível em: <http://www.fujipress.jp/finder/xslt.php?mode=present&inputfile=ROBOT002500010003.xml>.
- [2] AHMADABADIAN, A. H. et al. A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 78, p. 157–167, 2013.
- [3] BROWN, M.; LOWE, D. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*, Springer, v. 74, n. 1, p. 59–73, 2007.
- [4] FAUGERAS, O. D.; LUONG, Q.-T.; MAYBANK, S. J. Camera self-calibration: Theory and experiments. In: SANDINI, G. (Ed.). *Computer Vision — ECCV 92*. [S.l.]: Springer Berlin Heidelberg, 1992, (Lecture Notes in Computer Science, v. 588). p. 321–334.
- [5] BRÜCKNER, M.; BAJRAMOVIC, F.; DENZLER, J. Intrinsic and extrinsic active self-calibration of multi-camera systems. *Machine Vision and Applications*, v. 25, n. 2, p. 389–403, 2013. ISSN 0932-8092.
- [6] HARTLEY, R. Projective reconstruction. In: IKEUCHI, K. (Ed.). *Computer Vision*. [S.l.]: Springer US, 2014. p. 640–651.
- [7] LOURAKIS, M. I.; DERICHE, R. *Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3D measurements*. [S.l.], 1999.
- [8] MAYBANK, S. S.; FAUGERAS, O. O. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, Kluwer Academic Publishers, v. 151, n. 1992, p. 123–151, 1992.
- [9] VIÉVILLE, T.; LINGRAND, D. Using singular displacements for uncalibrated monocular visual systems. In: BUXTON, B.; CIPOLLA, R. (Ed.). *Computer Vision ECCV 96*. [S.l.]: Springer Berlin Heidelberg, 1996, (Lecture Notes in Computer Science, v. 1065). p. 207–216.
- [10] FAUGERAS, O.; LUONG, Q.-T.; PAPADOPOULOU, T. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. Cambridge, MA, USA: MIT Press, 2001. 659 p.
- [11] HARTLEY, R.; ZISSERMAN, A. *Multiple view geometry in computer vision*. [S.l.]: Cambridge Univ Press, 2000. v. 2.
- [12] LOURAKIS, M.; DERICHE, R. *Camera self-calibration using the Kruppa equations and the SVD of the fundamental matrix: The case of varying intrinsic parameters*. [S.l.], 2000.
- [13] GENG, L.-C. et al. A new camera self-calibration method based on csa. In: *Visual Communications and Image Processing (VCIP), 2013*. [S.l.: s.n.], 2013. p. 1–6.

- [14] LOWE, D. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Springer, v. 60, n. 2, p. 91–110, 2004.
- [15] BAY, H.; TUYTELAARS, T.; GOOL, L. SURF: Speeded Up Robust Features. In: LEONARDIS, A.; BISOCHOF, H.; PINZ, A. (Ed.). *Computer Vision – ECCV 2006*. [S.l.]: Springer, 2006, (Lecture Notes in Computer Science, v. 3951). p. 404–417.
- [16] RUBLEE, E. et al. ORB: An efficient alternative to SIFT or SURF. In: IEEE. *Computer Vision (ICCV), 2011 IEEE International Conference on*. [S.l.], 2011. p. 2564–2571.
- [17] LEUTENEGGER, S.; CHLI, M.; SIEGWART, R. Y. BRISK: Binary Robust invariant scalable keypoints. *2011 International Conference on Computer Vision*, Ieee, p. 2548–2555, 2011.
- [18] ALAHI, A.; ORTIZ, R.; VANDERGHEYNST, P. FREAK: Fast Retina Keypoint. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.: s.n.], 2012. p. 510–517.
- [19] MEDIONI, G.; NEVATIA, R. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, Elsevier, v. 31, n. 1, p. 2–18, 1985.
- [20] TRUCCO, E.; VERRI, A. *Introductory Techniques for 3-D Computer Vision*. [S.l.]: Prentice Hall, 1998. 145–146 p.
- [21] MIAN, A.; BENNAMOUN, M.; OWENS, R. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, Springer, v. 89, n. 2-3, p. 348–361, 2010.
- [22] BRITO, D. N. et al. Evaluation of interest point matching methods for projective reconstruction of 3d scenes. *IEEE Latin America Transactions*, IEEE, v. 14, n. 3, p. 1393–1400, 2016.
- [23] FREITAS, M. H.; PÁDUA, F. L.; ASSIS, G. T. d. Object-based image retrieval using local feature extraction and relevance feedback. *International Journal of Computer Applications*, Foundation of Computer Science, v. 78, n. 7, 2013.
- [24] RODRIGUES, M. T. et al. Evaluating cluster detection algorithms and feature extraction techniques in automatic classification of fish species. *Pattern Analysis and Applications*, Springer, v. 18, n. 4, p. 783–797, 2015.
- [25] DWARAKANATH, D. et al. Evaluating Performance of Feature Extraction Methods for Practical 3D Imaging Systems. In: *Proc. of the 27th Conf. on Image and Vision Computing New Zealand*. [S.l.: s.n.], 2012.
- [26] JIANG, Y.; XU, Y.; LIU, Y. Performance evaluation of feature detection and matching in stereo visual odometry. *Neurocomputing*, Elsevier, v. 120, n. 0, p. 380–390, 2013.
- [27] SNAVELY, N.; SEITZ, S. M.; SZELISKI, R. Photo tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics*, v. 25, n. 212, p. 835–846, 2006.
- [28] LI, X.; WANG, J. Image matching techniques for vision-based indoor navigation systems: performance analysis for 3d map based approach. In: *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. [S.l.: s.n.], 2012. p. 1–8.
- [29] HEINLY, J.; DUNN, E.; FRAHM, J. M. Comparative evaluation of binary features. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [S.l.: s.n.], 2012. v. 7573 LNCS, p. 759–773.
- [30] STRECHA, C. et al. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2008. p. 1–8.
- [31] ZHANG, Z. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 22, n. 11, p. 1330–1334, 2000.
- [32] MIKOLAJCZYK, K.; SCHMID, C. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 27, n. 10, p. 1615–1630, 2005.