UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO

Ariany França Cavalcante

Avaliação de Métodos de Calibração Automática de Câmeras Utilizando Pedestres como Referência

Recife

2025

Ariany França Cavalcante

Avaliação de Métodos de Calibração Automática de Câmeras Utilizando Pedestres como Referência

Trabalho de dissertação apresentado ao Programa de Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientação: Veronica Teichrieb

Coorientador: Rafael Alves Roberto

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Cavalcante, Ariany França.
   Avaliação de métodos de calibração automática de câmeras utilizando pedestres como referência / Ariany França Cavalcante. - Recife, 2025.
   67f.: il.

   Dissertação (Mestrado)- Universidade Federal de Pernambuco, Centro de Informática, Programa de Pós-Graduação em Ciência da Computação, 2025.
   Orientação: Veronica Teichrieb.

   1. Visão computacional; 2. Calibração de câmeras; 3. Detecção de pedestres. I. Teichrieb, Veronica. II. Título.

UFPE-Biblioteca Central

Ariany França Cavalcante

“I Avaliação de Métodos de Calibração Automática de Câmeras Utilizando Pedestres como Referência”

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Mídia e Interação.

Aprovado em: 31/03/2025.

_____

Orientadora: Profa. Dra. Veronica Teichrieb

BANCA EXAMINADORA

_____

Prof. Dr. Silvio de Barros Melo
Centro de Informática / UFPE

_____

Prof. Dr. Lucas Silva Figueiredo
Departamento de Computação / UFRPE

_____

Prof. Dr. Rafael Alves Roberto
University of Bath
(**coorientador**)

À minha família, que sempre esteve comigo, mesmo à distância, oferecendo apoio incondicional para que eu pudesse seguir esta jornada.

À minha orientadora, professora Verônica Teichrieb, pela dedicação e incentivo à pesquisa. Sua experiência e compromisso são fontes de inspiração.

Ao meu coorientador, Rafael Alves Roberto, por sua colaboração e comprometimento indispensáveis. Agradeço por cada ensinamento, palavra de incentivo e tempo dedicado.

Ao Voxar Labs, pelo ambiente estimulante e enriquecedor, que me proporcionou a oportunidade de crescer e aprender. A todos os membros do laboratório, agradeço pela troca de ideias, pelo apoio que tornou esta jornada um processo de aprendizado constante e, especialmente, pela amizade.

A todos os meus professores e pesquisadores mais experientes que, ao longo do caminho, com seus ensinamentos, me desafiaram e me impulsionaram para conhecimento.

Aos meus amigos de pesquisa, com quem compartilhei momentos de aprendizado e crescimento. A colaboração e o companheirismo de cada um foram essenciais para meu desenvolvimento, tanto profissional quanto pessoal.

RESUMO

A visão computacional desempenha um papel essencial em diversas aplicações, como vigilância inteligente e reconstrução 3D, permitindo o rastreamento de pessoas e objetos em sistemas multi-câmera. No entanto, para que esses sistemas operem corretamente, é fundamental que a calibração das câmeras seja precisa. A calibração automática surge como uma alternativa promissora à calibração manual tradicional, que apresenta desafios significativos, como a necessidade de um ambiente controlado, a exigência de intervenção humana e a dificuldade de recalibração em sistemas dinâmicos. Apesar do seu potencial, muitas técnicas do estado da arte ainda não foram amplamente testadas em cenários realistas, onde fatores como oclusões e rotas curtas podem impactar a precisão da calibração. Diante desse contexto, este trabalho investiga o desempenho de técnicas de calibração automática baseadas em pedestres, analisando sua eficácia e limitações em ambientes não controlados. Os experimentos demonstram que, embora a técnica avaliada apresente potencial, ainda há altos erros de calibração e grande variabilidade nas estimativas dos parâmetros extrínsecos. A qualidade dos dados de entrada mostrou-se um fator crítico, uma vez que, em condições reais, a detecção das poses humanas pode ser comprometida, afetando negativamente a calibração. Além disso, a rota dos pedestres influencia significativamente o desempenho do método. Os resultados indicam que a calibração automática de redes de câmeras ainda enfrenta desafios significativos para adaptação a cenários dinâmicos. Dessa forma, são necessárias abordagens mais robustas e generalizáveis, capazes de lidar com diferentes fontes de erro. A coleta de dados mais controlados pode ser uma estratégia para isolar e compreender melhor os fatores que afetam a calibração.

Palavras-chaves: Visão computacional. Calibração de câmeras. Calibração automática. Redes de câmeras. Detecção de pedestres. Parâmetros extrínsecos. Ambientes não controlados.

ABSTRACT

Computer vision performs a fundamental function in various applications, such as intelligent surveillance and 3D reconstruction, enabling the tracking of people and objects in multi-camera systems. However, for these systems to function correctly, precise camera calibration is essential. Automatic calibration emerges as a promising alternative to traditional manual calibration, which presents significant challenges, including the need for a controlled environment, human intervention, and difficulties in recalibrating dynamic systems. Despite its potential, many state-of-the-art techniques have not yet been extensively tested in realistic scenarios, where factors such as occlusions and short pedestrian trajectories may impact calibration accuracy. In this context, this study investigates the performance of pedestrian-based automatic calibration techniques, analyzing their effectiveness and limitations in uncontrolled environments. The experimental results show that, although the evaluated technique demonstrates potential, it still suffers from high calibration errors and significant variability in extrinsic parameter estimates. The quality of input data proved to be a critical factor, as, in real-world conditions, human pose detection may be compromised, negatively affecting calibration. Moreover, pedestrian motion patterns significantly influence the performance of the methods. The findings indicate that automatic camera network calibration still encounters considerable challenges in adapting to dynamic environments. Therefore, more robust and generalizable approaches are required to handle different sources of error. The collection of more controlled data may be a strategy to isolate and better understand the factors affecting calibration.

Keywords: Computer vision. Camera calibration. Automatic calibration. Camera networks. Pedestrian detection. Extrinsic parameters. Uncontrolled environments.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Organizações que atuam no planejamento de espaços urbanos frequentemente enfrentam o desafio de compreender como esses espaços são utilizados por seus usuários. Entre as informações mais relevantes para esse processo estão os trajetos mais comuns realizados por pedestres, os pontos de maior permanência e as atividades predominantes em determinadas áreas. Atualmente, esse tipo de dado é coletado de forma manual, por meio da observação presencial por profissionais em campo, durante algumas horas e em dias específicos, pela Massapê, uma ONG de urbanismo social de Recife - PE. Esse método, além de ser intensivo em tempo e mão de obra, oferece uma cobertura limitada e pontual da realidade.

Uma alternativa promissora para melhorar o atendimento dessa demanda consiste no uso de redes de câmeras aliadas a sistemas de visão computacional, capazes de monitorar continuamente os ambientes e fornecer dados espacializados, como mapas de calor de atividades e fluxos de movimentação de pedestres. Para que tais sistemas operem de forma precisa e confiável, é indispensável que as câmeras estejam corretamente calibradas, ou seja, que os parâmetros necessários para associar as imagens captadas com a geometria do ambiente real estejam devidamente estimados.

O processo de calibração de câmeras permite estimar dois conjuntos de parâmetros: os parâmetros intrínsecos, que descrevem características internas do dispositivo, como distância focal e distorção da lente; e os parâmetros extrínsecos, que representam a posição e a orientação da câmera no espaço em relação a um sistema de coordenadas global. Neste trabalho, o foco será direcionado à calibração extrínseca, que é a responsável por espacializar corretamente as informações capturadas pelas câmeras em um referencial comum.

Tradicionalmente, a calibração extrínseca é realizada de forma manual, com o auxílio de padrões artificiais, como tabuleiros de xadrez, posicionados em diferentes ângulos no campo de visão das câmeras. Esse procedimento, embora muito usado, demanda ambientes controlados e profissionais especializados, além de apresentar baixa flexibilidade em contextos em que as câmeras rotacionam e ampliam a imagem por meio de zoom. Sempre que uma câmera é reposicionada ou sofre deslocamentos, o processo precisa ser refeito, o que representa um entrave significativo para aplicações em larga escala ou sujeitas a mudanças frequentes.

Diante dessas limitações, a calibração automática de câmeras surge como uma abordagem mais eficiente e escalável. Essa técnica visa estimar os parâmetros extrínsecos sem a necessi-

dade de padrões artificiais ou intervenção manual, reduzindo o tempo e o custo do processo, além de possibilitar a recalibração contínua em ambientes sujeitos a mudanças. No contexto urbano, uma estratégia promissora consiste na utilização de pedestres como padrões naturais, uma vez que eles estão frequentemente presentes nesses ambientes, possuem morfologia relativamente estável e realizam trajetórias que podem ser exploradas para inferir a geometria da cena. Tais abordagens têm aplicações diretas em áreas como segurança pública, análise de fluxo de pessoas, planejamento urbano e outras tecnologias para cidades inteligentes.

Apesar de seu potencial, as técnicas atuais de calibração automática baseadas em pedestres ainda enfrentam importantes desafios, especialmente quando aplicadas a cenários urbanos reais. A literatura existente revela que os testes são frequentemente realizados em ambientes controlados, com condições simplificadas. Por exemplo, (TRUONG et al., 2019) validam sua abordagem utilizando no máximo sete pedestres simultaneamente, enquanto (LEE et al., 2022) avaliam sua técnica com apenas um indivíduo posicionado no centro de uma área monitorada por múltiplas câmeras. Essas condições não refletem as complexidades do ambiente urbano real, onde há intensa movimentação de pessoas.

Aplicações práticas em contextos reais trazem uma série de desafios adicionais, como oclusões causadas por pedestres e objetos do ambiente, variações nas condições de iluminação que geram sombras e dificultam a detecção, além da baixa resolução de muitas câmeras urbanas, o que prejudica a identificação e o rastreamento de indivíduos. Esses fatores impactam diretamente a robustez das técnicas de calibração automática, exigindo investigações mais aprofundadas sobre sua viabilidade e desempenho em condições reais de operação.

Diante desse contexto, o presente trabalho tem como objetivo investigar a eficácia de técnicas de calibração automática de câmeras baseadas em pedestres em cenários urbanos reais, explorando seus pontos fortes, limitações e a viabilidade de sua aplicação prática em ambientes dinâmicos. Importa destacar que não se trata de um estudo comparativo entre diferentes métodos, mas sim de uma análise exploratória voltada à compreensão do comportamento dessas abordagens quando aplicadas a situações reais. Os testes realizados não seguem uma padronização rígida, e sim assumem caráter experimental, com o intuito de levantar percepções qualitativas e quantitativas sobre o desempenho das técnicas estudadas.

- **Investigar técnicas de calibração automática de câmeras utilizando pedestres como referência** e compreender suas vantagens e limitações (Capítulos 2 e 3);
- **Testar e analisar o desempenho dessas técnicas** em diferentes conjuntos de dados

que simulam cenários reais, com foco na robustez frente a oclusões e trajetórias reais (Capítulos **??** e **??**);

- **Identificar os principais desafios e possíveis melhorias na calibração automática** para tornar essas abordagens mais viáveis para aplicações práticas (Capítulo 6).

Com essa investigação, espera-se contribuir para o desenvolvimento de sistemas mais eficientes e autônomos de calibração de redes de câmeras, reduzindo a necessidade de intervenção manual e tornando a visão computacional mais acessível e aplicável a cenários do mundo real.

2 REFERENCIAL TEÓRICO

A calibração automática de câmera é bastante relevante em diversos contextos. Alguns conceitos matemáticos são fundamentais encontrar os parâmetros que descrevem a câmera. Este capítulo tem como objetivo mostrar alguns desses principais conceitos. Eles incluem o modelo matemático de câmera mais utilizado, o pinhole, além de geometria epipolar. Também será apresentado duas técnicas recentes que ilustram como a calibração pode ser realizada usando pedestres.

2.1 MODELO DE CÂMERA PINHOLE

Também conhecido como modelo de orifício, o modelo Pinhole representa a relações geométricas entre a cena e a imagem (HARTLEY; ZISSERMAN, 2003). Ele assume que os raios de luz passam por um único ponto, conhecido como centro de projeção, antes de atingir o plano da imagem, 1. Isso simplifica a descrição da formação de imagens.

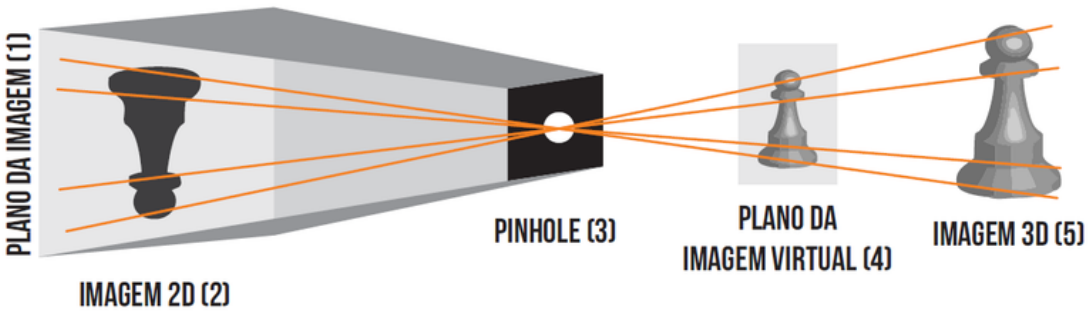

Figura 1 – Modelo de câmera pinhole. Fonte: (DIAS, 2015).

Matematicamente, o modelo pinhole é descrito por uma matriz de projeção $\mathbf{M}$ que mapeia as coordenadas de um ponto no espaço tridimensional $P_w = [x_w, y_w, z_w, 1]^T$ para suas coordenadas correspondentes no plano da imagem $p = [u, v, 1]^T$. A matriz de projeção é formada pela matriz de parâmetros intrínsecos $\mathbf{K}$ e a matriz de parâmetros extrinsecismos $[\mathbf{R}|t]$, onde $\mathbf{R}$ é a matriz de rotação, e t é o vetor de translação. Enquanto os parâmetros intrínsecos descrevem as propriedades internas da câmera, como distância focal e posição do centro óptico, os parâmetros extrínsecos definem a orientação e a posição da câmera no espaço tridimensional em relação ao mundo real.

A calibração de câmera é um processo para determinar esses parâmetros intrínsecos e extrínsecos que descrevem um modelo de câmera.

2.1.1 Parâmetros Intrínsecos e Extrínsecos

Os parâmetros intrínsecos e extrínsecos desempenham papéis complementares na modelagem de câmeras. Enquanto os parâmetros intrínsecos descrevem as propriedades internas da câmera, como distância focal e posição do centro óptico, os parâmetros extrínsecos definem a orientação e a posição da câmera no espaço tridimensional em relação ao mundo real (HARTLEY; ZISSERMAN, 2003)..

Os parâmetros intrínsecos são representados por uma matriz de calibração $\mathbf{K}$, que encapsula as propriedades internas da câmera. Essa matriz é definida como:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.1}$$

onde f_x e f_y são as distâncias focais em pixels nos eixos x e y, e c_x e c_y representam as coordenadas do centro óptico no plano da imagem. Essa matriz projeta os pontos em coordenadas de câmera no plano de imagem, em coordenadas da imagem.

A projeção de um ponto tridimensional $P_w = [x_w, y_w, z_w, 1]^T$ no sistema de coordenadas do mundo para a coordenada de câmera é feita pelos parâmetros extrínsecos, que são representados pela concatenação de uma matriz de rotação $\mathbf{R}$ e um vetor de translação t. Esses parâmetros transformam o ponto em coordenada de mundo da seguinte forma:

$$P_c = [\mathbf{R}|t] \cdot P_w = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \end{bmatrix} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \tag{2.2}$$

onde $P_c = [x_c, y_c, z_c]^T$ são as coordenadas do ponto no sistema da câmera. A matriz $\mathbf{R}$ é uma matriz 3×3 que define a orientação da câmera, enquanto o vetor t representa sua posição.

A combinação dos parâmetros intrínsecos e extrínsecos resulta na matriz de projeção $\mathbf{M}$, que mapeia as coordenadas tridimensionais do mundo diretamente para as coordenadas da

imagem. Essa matriz é definida como:

$$\mathbf{M} = \mathbf{K} \cdot [\mathbf{R}|t]. \tag{2.3}$$

O mapeamento completo para as coordenadas da imagem $p = [u, v, 1]^T$ é então dado por:

$$p = \mathbf{M} \cdot P_w = \mathbf{K} \cdot [\mathbf{R}|t] \cdot P_w. \tag{2.4}$$

Apesar de sua simplicidade, o modelo pinhole consegue ser útil em varias situações, reforçando a sua popularidade. Porém, ele apresenta limitações por não considerar as distorções ópticas introduzidas por lentes reais, o que afeta a precisão do mapeamento geométrico. Porém, modelos complementares podem ser integrados a ele. Isso permite a inclusão de parâmetros de distorção, por exemplo, aumentando a precisão em situações práticas.

2.1.2 Parâmetros de Distorção

As lentes das câmeras reais introduzem inevitavelmente distorções ópticas que comprometem a precisão da projeção geométrica idealizada pelo modelo pinhole. Entre essas distorções, as distorções radiais são as mais comuns. A correção da distorção radial é representada por uma função que ajusta as coordenadas da imagem distorcida (u, v) para as coordenadas corrigidas (u_c, v_c). Essa relação é definida por meio da introdução de coeficientes de distorção radial $k_1, k_2, k_3, \ldots$, e depende da distância radial r, dada por:

$$r = \sqrt{(u - c_x)^2 + (v - c_y)^2}, \tag{2.5}$$

onde c_x e c_y são as coordenadas do ponto central da câmera. A correção é aplicada às coordenadas normalizadas u e v por meio das seguintes equações:

$$u_c = u(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \tag{2.6}$$

$$v_c = v(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \tag{2.7}$$

onde u_c e v_c são as coordenadas corrigidas após o ajuste radial.

Além das distorções radiais, a distorção tangencial também pode comprometer a precisão da projeção geométrica de uma câmera (HEIKKILA; SILVÉN, 1997). Esse tipo de distorção ocorre

devido a imperfeições no alinhamento das lentes, o que faz com que os pontos na imagem sejam deslocados tangencialmente em relação ao centro óptico. Ela pode ser descrita por dois coeficientes p_1 e p_2, que modelam o deslocamento tangencial das coordenadas da imagem, conforme as seguintes equações:

$$u_c = u + \left[2p_1uv + p_2\left(r^2 + 2u^2\right)\right], \tag{2.8}$$

$$v_c = v + \left[p_1\left(r^2 + 2v^2\right) + 2p_2uv\right], \tag{2.9}$$

onde r é a mesma distância radial dada pela Equação 2.5.

A distorção tangencial é frequentemente tratada em conjunto com a distorção radial, compondo um modelo completo de correção. O processo de calibração determina os coeficientes p_1 e p_2 juntamente com os coeficientes de distorção radial k_1, k_2, k_3, permitindo uma correção combinada que melhora significativamente a qualidade geométrica da projeção dada pelas equações:

$$u_c = u(1 + k_1r^2 + k_2r^4 + k_3r^6) + \left[2p_1uv + p_2\left(r^2 + 2u^2\right)\right], \tag{2.10}$$

$$v_c = v(1 + k_1r^2 + k_2r^4 + k_3r^6) + \left[p_1\left(r^2 + 2v^2\right) + 2p_2uv\right], \tag{2.11}$$

2.2 GEOMETRIA EPIPOLAR

A geometria epipolar descreve as restrições geométricas entre duas imagens de uma cena capturadas a partir de diferentes pontos de vista (HARTLEY; ZISSERMAN, 2003; FAUGERAS; LUONG; PAPADOPOULO, 2001). Ela é fundamental para a calibração de câmeras e a reconstrução 3D, pois impõe relações matemáticas entre os pontos correspondentes das imagens.

A geometria epipolar se baseia na noção de epipolo, que é o ponto onde a linha que conecta os centros de câmera intersectam o plano de imagem 2.

Outro conceito importante é o do plano contendo os centros das câmeras e o ponto 3D que está sendo observado, chamado de plano epipolar. Quando o ponto 3D é projetado em uma das imagens, o seu correspondente na outra imagem está restrito a uma linha, chamada de linha epipolar. Essa linha é a projeção da reta que passa pelo centro da primeira câmera e o ponto 3D. Ou seja, se temos o ponto p_1 na primeira imagem, seu correspondente na segunda

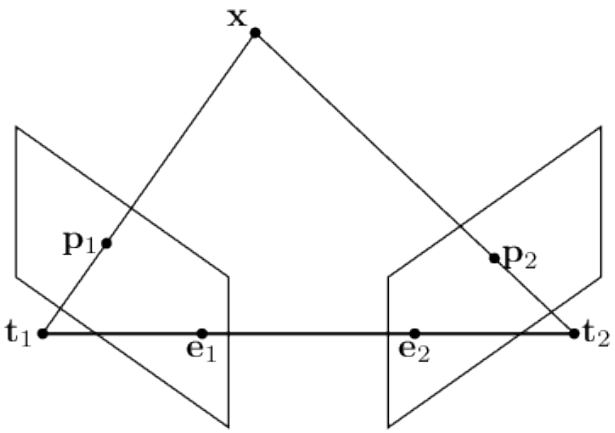

Figura 2 – Conceitos bésicos de geometria epipolar. Fonte: (TRUCCO; VERRI, 1998).

imagem está restrito à linha epipolar. É importante notar que todas as linhas epipolares passam pelo epipolo. Matematicamente, essa restrição é dada pela equação:

$$p_2^T \cdot \mathbf{F} \cdot p_1 = 0, \tag{2.12}$$

onde p_1 e p_2 são pontos correspondentes na primeira e segunda câmeras, respectivamente, e $\mathbf{F}$ é a matriz fundamental, que encapsula a relação geométrica entre as duas câmeras.

A matriz fundamental $\mathbf{F}$ descreve a relação geométrica entre duas câmeras. Ela é usada quando não se tem informação dos parâmetros intrínsecos e pode ser calculada a partir de um conjunto de pontos correspondentes em coordenadas de câmera. A matriz essencial $\mathbf{E}$ é similar à matriz fundamental, mas ela assume que o sistema está calibrado e os pontos expressos em coordenadas de imagem. A matriz essencial está relacionada à matriz fundamental pela seguinte equação:

$$\mathbf{E} = \mathbf{K_2}^T \cdot \mathbf{F} \cdot \mathbf{K_1}, \tag{2.13}$$

onde $\mathbf{K_1}$ e $\mathbf{K_2}$ são as matrizes de parâmetros intrínsecos das duas câmeras.

2.2.1 Erro Epipolar

O erro epipolar mede a distância entre um ponto projetado e sua linha epipolar correspondente. Essa distância é um indicativo da precisão da calibração das câmeras e da reconstrução 3D (HARTLEY; ZISSERMAN, 2003). Uma métrica para esse erro é a distância simétrica epipolar:

$$d(p_1, l_2) + d(p_2, l_1), \tag{2.14}$$

onde $d(p, l)$ é uma função que calcula a menor distância entre um ponto p em coordenada de imagem e uma reta na imagem e l_1 e l_2 são as linhas epipolares na primeira e segunda câmera.

2.3 MÉTODOS DE CALIBRAÇÃO DE CÂMERA

O processo de calibração de câmeras é uma etapa fundamental para garantir a precisão de aplicações de visão computacional. Ele envolve o uso de características facilmente distinguível na imagem para calcular as matrizes $\mathbf{K}$ e $\mathbf{R}$ e o vetor t. Essas características podem ser extraídas de objetos artificiais inseridos na cena, como um template xadrez, assim como elementos naturais, como linhas e pedestres. Após isso, os dados extraídos são utilizados para estimar as matrizes de calibração.

2.3.1 Extração de Características por Elementos Artificiais

O método mais comum de calibração utiliza padrões facilmente reconhecidos, como tabuleiros de xadrez ou alvos de pontos circulares, colocados na cena. A posição e dimensão desses padrões são medidos, que permite correlacionar os pontos extraídos na imagem com uma referência em coordenadas de mundo. Por serem padrões de alto contraste, esses padrões são mais facilmente identificados e os pontos na imagem correspondentes a eles são extraídos na imagem. Esse tipo de calibração é feito capturando múltiplas imagens do padrão em diferentes posições e momentos.

2.3.2 Extração de Características por Elementos Naturais

Esta forma de extrair características da imagem explora elementos naturais pertencentes à cena, como pontos de alto destaque, linhas e até pedestres. A grande vantagem dessa forma é que ela pode automatizar o processo de calibração, uma vez que não é necessário interferir na cena. Por outro lado, essas características são mais difíceis de serem extraídas, que pode resultar num aumento da imprecisão, por exemplo.

Uma das formas de extrais características da imagem sem a necessidade de inserir elementos artificiais é identificar pontos de alto destaques, também chamados de features. Eles podem ser encontrados usando extratores de features, como o SIFT (Scale-Invariant Feature Transform) (LOWE, 2004). Para encontrar uma referência com as coordenadas de mundo,

normalmente são usadas features de objetos cujas dimensões e posições são conhecidas.

2.3.3 Calibração de Câmeras

As características extraídas, seja utilizando elementos artificiais ou naturais, são utilizadas para estimar os parâmetros intrínsecos e extrínsecos da câmera. Essas características estão presentes vários frames e as correspondências delas durante a sequência e os seus correspondentes em coordenada de mundo podem ser usados para realizar a calibração. Uma das formas é usando a Transformação Linear Direta (cuja sigla do termo em inglês é DLT) (ABDEL-AZIZ; KARARA; HAUCK, 2015). Ela utiliza as correspondências 2D-3D para calcular a matriz de projeção $\mathbf{M}$. Posteriormente, a matriz $\mathbf{M}$ pode ser decomposta em seus componentes intrínsecos $\mathbf{K}$ e extrínsecos $[\mathbf{R}|t]$ usando a Decomposição em Valores Singulares (cuja sigla do termo em inglês é SVD) (KLEMA; LAUB, 1980), seguida por uma etapa de refinamento por métodos de otimização não linear, como o algoritmo de Levenberg-Marquardt.

Casos onde os pontos 3D são coplanares podem ocorrer quando as características são extraídas usando elementos artificiais. Nesse caso, são calculadas as homografias que mapeiam os pontos 2D das imagens para os pontos 3D do plano. A partir dessas homografias, é possível extrair uma estimativa inicial dos parâmetros, que pode ser posteriormente refinada através de uma otimização não linear, minimizando o erro de reprojeção (ZHANG, 2002).

2.3.4 Calibração de Rede Multicâmeras

A calibração de câmeras pode ser aplicada tanto a sistemas com uma única câmera quanto a redes multicâmeras. No caso de uma única câmera, o processo envolve a determinação dos parâmetros intrínsecos e extrínsecos em relação a um referencial, permitindo mapear pontos do mundo tridimensional para a imagem bidimensional. Entretanto, em uma rede de câmeras, além da calibração individual de cada dispositivo, é necessário estimar as relações espaciais entre elas, ou seja, determinar as matrizes de rotação e os vetores de translação que alinham os diferentes sistemas de coordenadas em um referencial comum (FAUGERAS; LUONG; PAPADOPOULO, 2001).

Em redes multicâmeras, surgem desafios adicionais, como a necessidade de sincronização temporal e de um alinhamento espacial preciso, especialmente em ambientes dinâmicos onde os cenários podem mudar rapidamente. Técnicas avançadas são empregadas para ajus-

tar os parâmetros de cada câmera e minimizar discrepâncias entre as diferentes perspectivas, considerando fatores como iluminação variável, oclusões e movimentos rápidos. Parâmetros compartilhados, como a distância entre as câmeras e a orientação relativa, são fundamentais para garantir a consistência geométrica e a precisão na reconstrução tridimensional, enquanto a sincronização dos frames é crucial para a integração confiável dos dados, especialmente em aplicações de monitoramento em tempo real.

Além disso, a sobreposição das áreas capturadas por diferentes câmeras oferece redundância e melhora a precisão dos algoritmos de visão computacional. Métodos robustos de calibração aproveitam a correspondência de pontos em múltiplas vistas e a utilização de objetos com trajetórias conhecidas para refinar as estimativas. A aplicação de técnicas de otimização, como a minimização do erro de reprojeção, permite ajustar globalmente os parâmetros intrínsecos e extrínsecos, enquanto modelos matemáticos avançados incorporam regularizações para lidar com diferenças de resolução e sobreposição de campos de visão, transformando o problema em uma complexa otimização multidimensional.

2.4 MÉTODOS DE CALIBRAÇÃO BASEADO EM PEDESTRES

Um outro elemento que pode ser usado na extração de características naturais são pedestres, especialmente quando se trata de cenários urbanos (GUAN et al., 2016; TEMPELAAR, 2022). Essas técnicas exploram pontos da anatomia humana, como articulações e extremidades do corpo humano, como correspondências para estimar relações entre câmeras. Essas abordagem baseia-se na premissa de que uma pessoa manterá sua estrutura corporal, como distância da cabeça aos pés, ao ser capturada em diferentes instantes de tempo. Assim, é possível extrair uma relação entre as características extraídas da imagem e as coordenadas de mundo. Dois tipos de técnicas são usados como base para ilustrar como pedestres podem ser usados para estimar a calibração extrínseca.

2.4.1 Calibração Extrínseca Baseada em Torsores de Pedestres

O TorsorCalib (TRUONG et al., 2019) usa o conceito do torso de um pedestre para obter os pontos que serão usados na calibração. O torso é um segmento de reta que vai do pescoço até os pés de uma pessoa. Para obter a posição de um pedestre na imagem, é aplicado um método de detecção de pose humana, que fornece o esqueleto das principais articulações do

corpo. Esses métodos podem ser o AlphaPose (FANG et al., 2022) ou o OpenPose (CAO et al., 2019). A articulação do pescoço é utilizada como referência para a parte superior do torso, enquanto a base é definida como o ponto médio entre os tornozelos esquerdo e direito. Essa técnica assume que os frames estão sincronização e a calibração intrínseca já foi realizada.

Seja uma rede de câmeras composta por n câmeras $C_1, C_2, \ldots, C_n$ e o ponto $P_w = [x_w, y_w, z_w, 1]^T$. Cada câmera possui seu próprio sistema de coordenadas locais. Assim, o ponto P_w no sistema da câmera i é dado por $P_w^{(i)} = [x_w^{(i)}, y_w^{(i)}, z_w^{(i)}, 1]^T$. Assim, a transformação que leva do sistema de coordenadas de mundo para o sistema de coordenadas de qual quer uma das câmeras é dada por $P_w^{(i)} = \mathbf{R}^{(i)} \cdot P_w + t^{(i)}$.

No TorsoCalib, a calibração é realizada por pares de câmera. Assim, nessas duas câmeras observadas, supõe-se que o pedestre moveu-se em m frames e manteve a mesma postura. Sejam $\tilde{\mathbf{u}}_{\text{bottom}}^{(i)}(f)$ e $\tilde{\mathbf{u}}_{\text{top}}^{(i)}(f)$ as posições de imagem da base e do topo do pedestre na câmera i no frame f. Essas coordenadas são normalizadas como $\tilde{\mathbf{x}}_{\text{bottom}}^{(i)}(f)$ e $\tilde{\mathbf{x}}_{\text{top}}^{(i)}(f)$, permitindo a recuperação das coordenadas 3D (GUAN et al., 2016).

Assumindo que o pedestre tem altura h, define-se as coordenadas tridimensionais da base e do topo como:

$$P_{w\_\text{top}}^{(i)}(f) = z_{w\_\text{top}}^{(i)}(f)\tilde{\mathbf{x}}_{\text{top}}^{(i)}(f), \tag{2.15}$$

$$P_{w\_\text{bottom}}^{(i)}(f) = z_{w\_\text{bottom}}^{(i)}(f)\tilde{\mathbf{x}}_{\text{bottom}}^{(i)}(f). \tag{2.16}$$

Dessa forma, tem-se:

$$P_{w\_\text{top}}^{(i)}(f) - P_{w\_\text{bottom}}^{(i)}(f) = he_z^{(i)}, \tag{2.17}$$

onde $e_z^{(i)}$ é o vetor unitário do pedestre na câmera i.

Pode-se então calcular um vetor 3D perpendicular ao plano vertical contendo a origem da câmera e os pontos de topo e base:

$$\mathbf{m}^{(i)}(f) = \tilde{\mathbf{x}}_{\text{bottom}}^{(i)}(f) \times \tilde{\mathbf{x}}_{\text{top}}^{(i)}(f). \tag{2.18}$$

A interseção desses planos define a direção vertical comum, e aplica-se SVD à matriz $\mathbf{M}^{(i)}$ para determinar $e_z^{(i)}$. Finalmente, usa-se Análise de Procrustes para estimar a transformação rígida entre os conjuntos de pontos 3D e, assim, chega-se a matriz de rotação $\mathbf{R}^{(i)}$ e a de translação $t^{(i)}$.

2.4.2 Calibração Extrínseca Baseada em Articulações Orientadas de um Corpo em Movimento

O MovingCalib (LEE et al., 2022) considera a movimentação do corpo humano para realizar a calibração. Diferente de abordagens convencionais que utilizam apenas pontos correspondentes 2D ou 3D, este método considera as posições e orientações das articulações corporais. Dessa forma, cada ponto possui uma posição r_f e uma direção v_f em coordenadas de mundo. Assim, para uma câmera i com rotação matriz de rotação $\mathbf{R^{(i)}}$ e vetor de translação $t^{(i)}$, esses pontos são transformados para o sistema da câmera por meio das equações:

$$P_w^{(i)} = \mathbf{R}^{(i)} \cdot r_f + t^{(i)}, \tag{2.19}$$

$$v_f^{(i)} = \mathbf{R}^{(i)} \cdot v_f \tag{2.20}$$

e projetados na imagem usando a matriz intrínseca $K^{(i)}$, obtida previamente.

A calibração inicia com a estimativa da rotação, que é obtida formando-se uma matriz de observação a partir das direções $v_f^{(i)}$ medidas. A decomposição em valores singulares (SVD) dessa matriz permite recuperar rotações normalizadas para cada câmera. Com as rotações determinadas, a translação é estimada utilizando restrições de colinearidade e coplanaridade, que garantem a consistência entre os pontos 3D e suas projeções.

Após essas etapas, é aplicado um bundle adjustment para minimizar o erro de reprojeção, refinando simultaneamente os parâmetros extrínsecos. Complementarmente, um fine-tuning auto-supervisionado utiliza as poses humanas 3D trianguladas como pseudo ground-truth para aprimorar o estimador de pose 3D monocular, garantindo uma calibração robusta mesmo em condições de entrada ruidosa.

3 REVISÃO DA LITERATURA

Existe uma vasta e diversa literatura na área de calibração extrínseca automática de câmeras e redes de câmeras. Elas variam em relação à quantidade de pedestres necessários, passando por abordagens baseadas correspondência entre linhas epipolares de silhuetas em movimento e até métodos não supervisionados baseados em trajetórias de pessoas.

(HöDLMOSER; KAMPEL, 2010) propõem um método para calibrar redes de câmeras de vigilância utilizando o movimento de um único pedestre como referência. A abordagem calcula os parâmetros intrínsecos e extrínsecos das câmeras, além de um fator de escala para estimar a altura real do pedestre. Os experimentos, realizados com dados sintéticos e reais, evidenciam a precisão e robustez do método, mesmo na presença de ruído ou variação nos pontos detectados. A técnica também se destaca pela eficiência computacional, com um incremento de apenas 1,4 segundos no tempo de processamento por câmera adicionada. Esta solução oferece uma alternativa prática e eficaz para calibração em cenários de segurança, possibilitando reconstrução 3D e estimativas de altura em ambientes desafiadores.

Já a técnica proposta por (BEN-ARTZI, 2017) utiliza correspondência entre linhas epipolares das silhuetas de pessoas em movimento. Melhorando em duas vezes o desempenho de métodos semelhantes e reduzindo outliers. Os pontos positivos deste estudo incluem o uso de um modelo de grafos que melhora a capacidade de realizar correspondências de pontos em diferentes vistas. Além de que, o uso de estimadores de probabilidade condicional permite um fine-tuning das correspondências, o que pode ser especialmente útil em cenas com movimento complexo. No entanto, a abordagem é dependente do movimento das silhuetas, o que pode ser problemático se os objetos possuem contornos pouco definidos. Além disso, a técnica pode falhar em situações em que os epípolos estão dentro do casco convexo, o que dificulta a recuperação de pontos de correspondência precisos.

O trabalho de (POSSEGGER et al., 2012) propõe um método de calibração extrínseca não supervisionada para redes de câmeras estáticas e PTZ (sigla para pan-tilt-zoom), baseado em correspondências entre trajetórias de pessoas em movimento. A abordagem utiliza a extração de localizações de cabeça e pés de pedestres a partir de imagens e realiza uma otimização não linear do erro de reprojeção para determinar os parâmetros extrínsecos das câmeras. Os experimentos demonstraram que o método consegue fornecer estimativas precisas dos parâmetros de câmeras em cenários variados, incluindo cenários externos. Pontos positivos incluem capa-

cidade do método de lidar com câmeras PTZ e cenas com múltiplos objetos em movimento. No entanto, pontos negativos envolvem a dependência da qualidade dos rastreamentos das pessoas, o que pode ser comprometido por oclusões ou movimentações rápidas.

O estudo de (LIU; COLLINS; LIU, 2013) propõe uma abordagem para a autocalibração de câmeras em redes de vigilância, utilizando uma estrutura de otimização conjunta combinada com estatísticas para obter calibração precisa. Para esta técnica não são necessários o rastreamento ou pontos de correspondência da mesma pessoa ao longo do tempo ou entre diferentes vistas. O algoritmo se destaca por sua robustez em cenários desafiadores, como ambientes com densidades moderadas de multidões e com ruído significativo proveniente de elementos de primeiro plano. No entanto, sua eficácia depende da qualidade das detecções e pode exigir recursos computacionais consideráveis, o que pode representar um desafio para implementações em tempo real.

(TEIXEIRA; MAFFRA; BADII, 2014) apresentam um framework para a autocalibração de câmeras de vigilância em cenários reais. O método proposto é generalizável para cenários do mundo real e utiliza segmentação semântica para gerar um mapa de ocupação, identificando áreas de interesse na cena, destacando pedestres e minimizando o impacto de oclusões. Essa abordagem permite ao framework lidar de forma eficaz com desafios como oclusões e a presença de objetos inesperados na cena. Além disso, o método integra a detecção de pedestres com a aplicação do algoritmo RANSAC para identificar linhas verticais e estimar o ponto de fuga vertical. Um aspecto adicional destacado no framework é a capacidade de refinar as estimativas de altura dos pedestres, o que melhora a precisão para aplicações que dependem dessa funcionalidade.

O método proposto por (PUWEIN et al., 2015) oferece uma abordagem robusta para a calibração de câmeras e a estimativa das posições 3D das articulações humanas em cenas onde há a movimentação de pedestres. As posições das articulações são estimadas para, em seguida, realizar uma otimização conjunta dos parâmetros extrínsecos das câmeras e das posições 3D das articulações. Os pontos fortes do método incluem sua capacidade de proporcionar uma calibração precisa. Além de que, a otimização considera múltiplos fatores, como continuidade temporal, fluxo óptico e visibilidade das partes do corpo. A abordagem demonstrou boa generalização, com bons resultados em diferentes conjuntos de dados.

(LETTRY; DRAGON; GOOL, 2017) propõem um método para a calibração de redes de câmeras baseado em correspondências de planos e amostragem probabilística, abordando desafios como planos cruzados e a presença de outliers. A metodologia utiliza técnicas estatísticas

combinadas com algoritmos de Monte Carlo via Cadeia de Markov (MCMC) em camadas para estimar matrizes de homografia entre planos observados por diferentes câmeras. O processo inicia com a calibração de pares de câmeras que apresentam boa sobreposição e baixa movimentação, aproveitando esses resultados para calibrar, de forma progressiva, pares com menor sobreposição ou maior movimento em um esquema em cascata. A robustez do método é destacada pela sua capacidade de lidar com correspondências imprecisas e outliers por meio da modelagem probabilística, além de abordar planos cruzados utilizando caminhos triangulares mais curtos para estabelecer correspondências. Entretanto, o uso de MCMC em camadas aumenta significativamente a complexidade computacional, tornando o método potencialmente menos eficiente para conjuntos de dados extensos. Além disso, o desempenho depende fortemente de uma parametrização adequada, demandando ajustes criteriosos para obter resultados satisfatórios.

(HALPERIN; WERMAN, 2018) apresentam um método eficiente para calcular a geometria epipolar em cenas dinâmicas de redes de câmeras. A base do método considera a correspondência entre pixels e linhas epipolares em vistas diferentes. Avaliado em vídeos reais, o método demonstrou superioridade em relação a abordagens semelhantes, graças aos refinamentos aplicados e à significativa redução da complexidade computacional. Além disso, a técnica se destaca por sua eficácia em cenários onde as câmeras possuem ângulos de visão substancialmente diferentes, superando os desafios associados à correspondência de pontos nesses casos.

O estudo de (TRUONG et al., 2019), fundamentado na técnica investigada por (GUAN et al., 2016), propõe uma solução para a calibração de câmeras em cenários urbanos com mais de uma pessoa e oclusões parciais. O método se baseia na detecção de poses humanas em imagens de câmeras, modelando os pedestres como bastões verticais, para estabelecer correspondências entre pessoas em diferentes imagens. A robustez do método inclui calibrações com erros de reprojeção de 3,76 a 3,69 pixels. Além disso, o método foi integrado com uma estratégia de amostragem aleatória, o que aumenta sua resistência a ruídos e outliers nos dados de pose humana, além de reduzir significativamente o tempo de coleta de dados. No entanto, o método não foi testado em conjunto de dados com alta densidade de pessoas.

O estudo de (NOWAK et al., 2021) apresenta uma metodologia para percepção multimodal, utilizando câmeras RGB e de profundidade (RGB-D), e introduz a biblioteca OpenHSML (JORDAN, 2021). A abordagem proposta utiliza calibração de câmeras RGB-D para determinar as matrizes fundamentais e de projeção, através de grafos e estimadores de probabilidade con-

dicional e similaridade, oferecendo uma solução de código aberto, prática e independente de configurações específicas de câmeras. A OpenHSML destaca-se pela simplicidade e versatilidade. Porém, um problema identificado é a ocorrência de "buracos"nos mapas de profundidade, áreas em que informações de profundidade estão ausentes, inviabilizando medições nessa região. Além disso, erros de projeção podem surgir quando as câmeras estão distantes entre si, permitindo que pontos atrás do objeto de interesse sejam visualizados e causem inconsistências. Nos experimentos realizados, a distância máxima entre as câmeras foi de 2 metros, limitando a aplicação em cenários mais amplos, como o monitoramento de pedestres, onde pontos no fundo da cena são frequentemente capturados.

(MOLINER; HUANG; ASTROM, 2021) apresentam um método para estimar os parâmetros extrínsecos de câmeras, incluindo escala, rotação e translação, utilizando apenas imagens de vídeos sincronizados e correspondência de poses humanas. A técnica aprimora a precisão da estimativa ao considerar fontes de erro associadas à detecção de pose e ao utilizar articulações com maior confiabilidade. Além disso, incorpora uma função objetivo baseada no bundle adjustment, que combina erro de reprojeção e restrições relacionadas à plausibilidade de movimentos humanos, como ângulos realistas entre membros. Embora o método tenha demonstrado redução significativa no erro de reprojeção, ele é limitado a cenas com um único pedestre, dificultando sua aplicação em cenários reais mais complexos. Como trabalho futuro, os autores sugerem a extensão da abordagem para lidar com múltiplos pedestres, ampliando seu potencial em aplicações práticas.

O estudo de (LEE et al., 2022) propõe o uso da orientação de articulações corporais humanas para estimar os parâmetros extrínsecos em redes de câmeras. Para cada articulação, são estimados pontos 3D e, a partir deles, são encontrados os parâmetros extrínsecos através de correspondência geométrica através de um algorítmo linear. Após isso, se iniciam ciclos de: ajustes dos parâmetros de calibração com *bundle adjustment* e refinamento das estimativas de coordenadas 3D das articulações com a calibração ajustada. A técnica destaca-se pela capacidade de generalização a diferentes ambientes e pela robustez a ruídos e pequenos movimentos. No entanto, o método enfrenta limitações, como a ambiguidade de escala, que requer um objeto de referência conhecido para ser resolvida. Além disso, a alta complexidade computacional pode dificultar sua aplicação em tempo real ou em dispositivos com recursos limitados. Outra restrição é que a abordagem considera apenas uma pessoa na cena, limitando sua aplicabilidade em cenários mais complexos.

(TEMPELAAR, 2022) apresenta um modelo inovador para a calibração automática de câ-

meras em cenários com a presença de pessoas, utilizando estimativas de poses humanas e reidentificação automática para ajustar os parâmetros extrínsecos das câmeras. A metodologia emprega um estimador de poses humanas para detectar pontos-chave nos pedestres, cujas características são, em seguida, processadas por um algoritmo de reidentificação (re-ID) baseado em afinidade, OSNet (ZHOU et al., 2021). Essa abordagem automatiza a correspondência entre visualizações de diferentes câmeras, permitindo uma calibração eficiente em redes multicâmeras.

Nos experimentos, este modelo demonstrou resultados promissores, atingindo calibração com elevada precisão no conjunto de dados SALSA e calibrando três das sete câmeras no conjunto de dados WildTrack (CHAVDAROVA et al., 2018). A limitação na calibração completa das câmeras no WildTrack foi atribuída à insuficiência de memória da GPU utilizada, destacando um desafio técnico enfrentado na implementação. Apesar dessa limitação, o método representa um avanço significativo na calibração extrínseca automatizada em cenários com pedestres, oferecendo uma abordagem robusta e aplicável a contextos reais.

Apesar das contribuições significativas, essas abordagens ainda apresentam limitações, como a dependência de padrões visuais específicos, que não estão usualmente presentes no cenário real. Além disso, as técnicas estudadas ainda carecem de validação mais ampla em cenários reais, que incluem movimentações imprevisíveis e multidões densas. Muitos métodos foram avaliados em ambientes controlados ou pouco dinâmicos, cenas com menos de dez pessoas, o que limita sua aplicabilidade prática em alguns tipos de locais. Assim, existe a necessidade de expandir o conhecimento para verificar o desempenho dessas soluções em cenários mais desafiadores.

4 TORSORCALIB

4.1 MÉTODO

A pesquisa avalia técnicas de calibração seguindo uma abordagem composta pelas seguintes etapas: seleção das técnicas, implementação ou adaptação de código, anotação e verificação de dados, experimentos e análise dos resultados obtidos. Ilustrada na Figura 3, essas etapas buscam promover o entendimento de elementos que influenciam a calibração automática em cenário real, escolhendo os conjuntos de dados de modo a obter informações sobre características e limitações dos métodos.

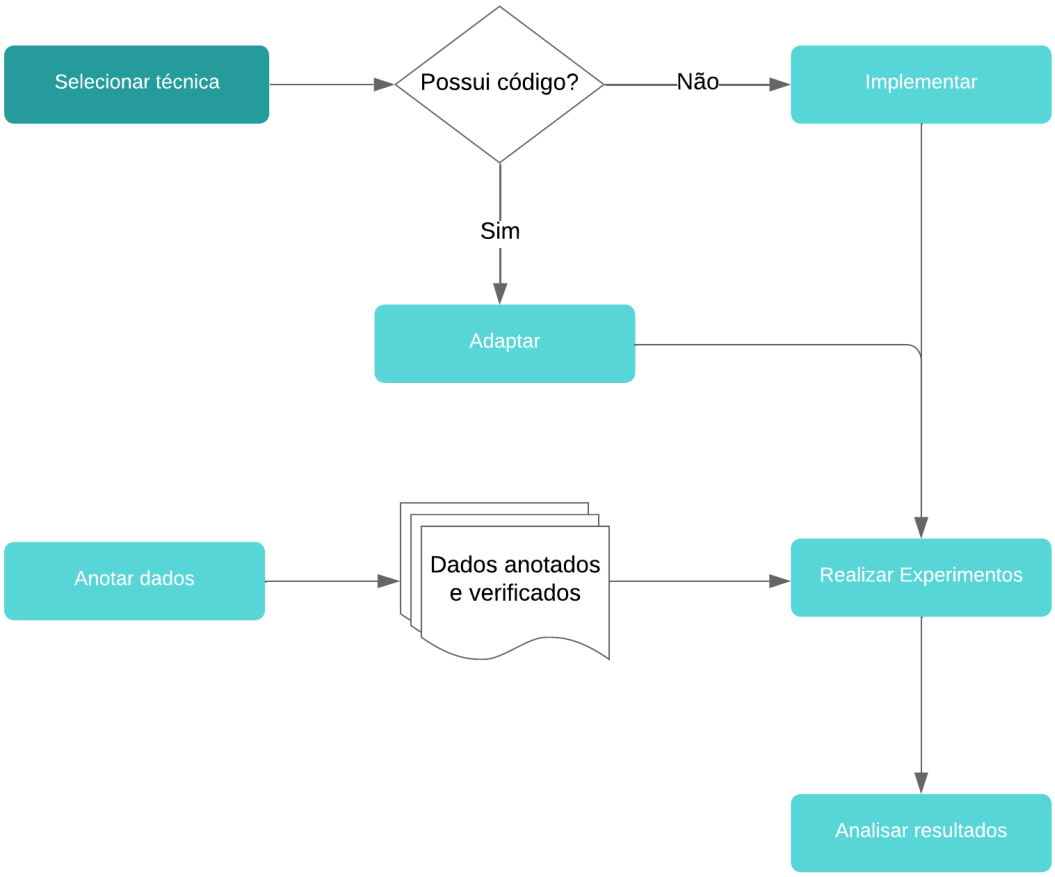

Figura 3 – Etapas da metodologia. Fonte: Elaborado pelo autor.

A primeira etapa é a seleção da técnica a ser experimentada, considerando os critérios de adequação aos objetivos da pesquisa e as limitações dos dados disponíveis. Em seguida,

procede-se à implementação ou adaptação do código, permitindo que a técnica selecionada esteja funcional e compatível com os dados e métricas do experimento.

Na etapa de anotação e de verificação de dados, são preparados os dados necessários para a avaliação, com a definição dos pedestres de calibração, anotação manual dos dados e verificação da correspondência dos esqueletos anotados. Após isso, realiza-se os experimentos, aplicando a técnica aos dados e gerando as métricas de avaliação. Por fim, ocorre a análise dos resultados, permitindo identificar pontos fortes, limitações e possibilidades de ajustes para os próximos testes.

4.1.1 Seleção de Técnica

O processo de seleção das técnicas ocorre a partir de um mapeamento da literatura, em que 21 técnicas de calibração automática de câmeras são levantadas. Para cada uma delas, são identificados os aspectos positivos e negativos de cada método, bem como os resultados de precisão publicados. Com base nessa análise, é selecionada uma técnica para experimentação considerando os seguintes critérios de seleção:

1. Utilizar pedestres como alvos para calibração;
2. Calibrar sistemas compostos por múltiplas câmeras;
3. Apresentar um baixo erro de calibração, tendo com referência o estado da arte de técnicas de calibração;
4. Ser aplicável a cenários com múltiplos pedestres presentes simultaneamente.

Com base nesses critérios, a primeira técnica é selecionada:

- **TorsorCalib:** Propõe uma solução para a calibração de câmeras em cenários complexos baseada em torsores gerados por pedestres na cena (TRUONG et al., 2019; LEE et al., 2022).

4.2 IMPLEMENTAÇÃO OU ADAPTAÇÃO DE TÉCNICA

Após a seleção, cada técnica é analisada em termos de seus algoritmos, requisitos computacionais e limitações com o objetivo de entender como ela pode ser implementada e adaptada

para este estudo. Nesta análise são identificadas ou escolhidas a linguagem de programação e ferramentas utilizadas. Vale ressaltar que as adaptações realizadas buscam não apenas possibilitar a execução técnica dos experimentos, mas também assegurar que o método opere de forma condizente com os objetivos do estudo.

O TorsorCalib, que não possui código-fonte disponível publicamente, é reimplementado a partir do método descrito nos seus dois artigos base (TRUONG et al., 2019; LEE et al., 2022). De forma específica, a implementação dessa técnica leva em consideração a modelagem dos pedestres como torsores verticais, formados a partir da ligação entre os pontos médios dos ombros e dos pés, para cada frame. Isso com o processo de calibração sendo realizado por meio das correspondências destes torsores nas diferentes vistas, obtendo os parâmetros extrínsecos para cada câmera da rede. Além disso, são feitas adaptações para as avaliações realizadas neste trabalho. Uma destas é a que permite que a entrada da técnica não seja a saída direta do detector de pose humana, mas sim o output do código de anotação de dados (mais informações na Seção 5.1.4).

cada técnica é adaptada para mostrar o erro de reprojeção de duas formas: o erro de reprojeção de cada vista calibrada e o erro de reprojeção médio do sistema de câmeras.

4.2.1 Avaliação de Técnica

São utilizados três datasets para os experimentos, selecionados considerando cenários que se aproximam de condições reais. Esses conjuntos de dados contêm pedestres em diferentes movimentos e posturas, além de variarem em aspectos como: trajetórias percorridas, densidade e nível de oclusão dos pedestres, posição das câmeras, áreas de sobreposição das vistas, quantidade de dados e outros. Essa diversidade permite avaliar o desempenho das técnicas de calibração em diferentes configurações e explorar diferentes características do método estudado.

4.2.1.1 EPFL Dataset - Campus Sequence

O EPFL Campus Sequence (CHAVDAROVA; FLEURET, 2017) é um dataset amplo e com sequências projetadas para a avaliação de técnicas de visão computacional em ambientes urbanos, caracterizados por diferentes cenários universitários. A sequência selecionada foi a Campus 4, que possui vídeos capturados por três câmeras com resolução de 360 x 288 pixels

distribuídas em uma área externa do campus universitário, conforme ilustrado na Figura 4.

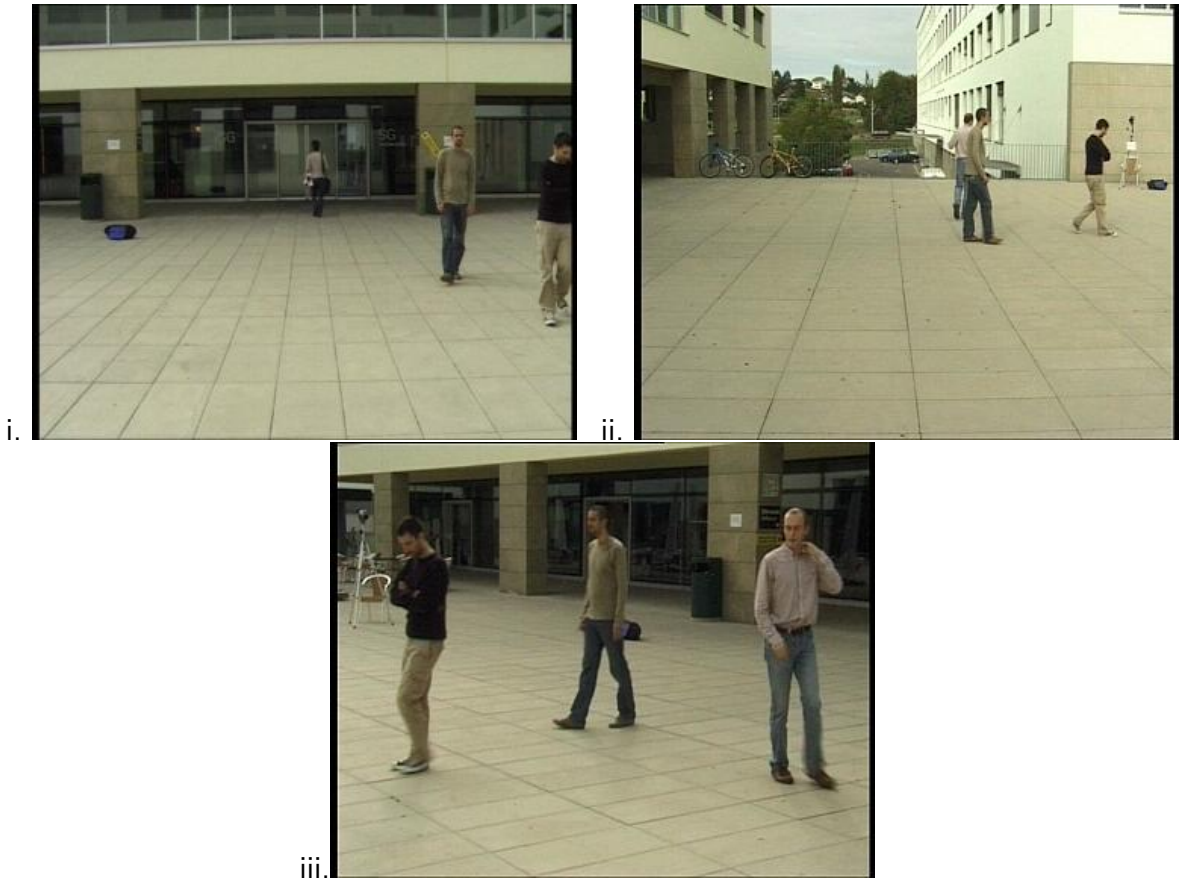

Figura 4 – Representação das vistas c0, c1 e c2 do EPFL Campus Sequence - Campus 4 usadas para calibração. Fonte: (CHAVDAROVA; FLEURET, 2017)

Esta sequência é caracterizada por um ambiente dinâmico, que inclui múltiplos pedestres se deslocando em postura ereta em direções diferentes. As câmeras apresentam campos de visão parcialmente sobrepostos, o que possibilita análises multivisão, como rastreamento de pedestres, calibração de câmeras e reconstrução tridimensional.

Porém, apesar de incluir aspectos dinâmicos, a sequência Campus 4 não apresenta uma elevada densidade de pedestres. São três pessoas se movimentando em trajetórias com aparência linear e em um espaço aberto. Essa configuração resulta em um cenário menos complexo que os próximos datasets apresentados, mas ainda desafiador pelas oclusões referentes à saída dos pedestres da área de sobreposição das vistas. Essas características tornam o dataset uma ferramenta relevante para avaliar o desempenho de técnicas em condições urbanas com menor densidade de pedestres.

4.2.1.2 Wildtrack

O Wildtrack Dataset (CHAVDAROVA et al., 2018) é amplamente reconhecido na área de visão computacional por sua aplicação em cenários não controlados e de alta densidade de pedestres. Ele consiste em capturas realizadas por sete câmeras estáticas, com resolução de 1920 x 1080 e dispostas ao redor de uma área de interesse, conforme ilustrado na Figura 5. As câmeras capturam imagens sincronizadas e o dataset fornece anotações detalhadas, incluindo as localizações tridimensionais dos pedestres. Essa configuração multicâmera permite explorar desafios característicos de ambientes reais, como oclusões severas causadas por pedestres e objetos no ambiente.

Figura 5 – Representação das três vistas do Wildtrack Dataset usadas para calibração. As imagens são capturadas simultaneamente por câmeras estáticas dispostas em torno de uma área central. As imagens destacam diferentes ângulos de visão, evidenciando a configuração multicâmera utilizada para a coleta de dados em cenários de monitoramento pedestre. Fonte: (CHAVDAROVA et al., 2018)

O Wildtrack se destaca pela diversidade de trajetórias dos pedestres, abrangendo desde movimentos rápidos e lentos até momentos de parada, além de apresentar, em muitos casos, percursos mais longos. Essas características resultam em um número elevado de detecções de poses, aumentando a quantidade de dados disponíveis para calibração, aspectos relevantes para esta pesquisa. A riqueza desse dataset, aliada à complexidade das condições presentes, torna-o uma ferramenta para validar métodos em tarefas como calibração multicâmera, detecção de poses e rastreamento de pedestres.

4.2.1.3 Si.U Dataset

O Si.U Dataset é um dataset próprio, capturado no pátio central do Centro de Informática (CIn) da Universidade Federal de Pernambuco (UFPE), com o objetivo de fornecer dados multicâmera para pesquisas voltadas à calibração de câmeras e detecção de pedestres em cenários reais. O conjunto de dados é composto por imagens de resolução 1920 x 1080 a partir de quatro vistas, posicionadas para capturar diferentes ângulos e foi criado para refletir a complexidade de ambientes não controlados, variando elementos como densidade de pessoas e iluminação. Imagens do dataset podem ser vistas Figura 6.

i.  ii. 

iii. 

Figura 6 – Imagens das três câmeras do Si.U Dataset utilizadas nos experimentos. As vistas capturam diferentes ângulos do pátio central do Centro de Informática da UFPE, evidenciando os desafios do cenário real, como oclusões e caminhos percorridos pelos pedestres. Fonte: Elaborado pelo autor.

Este conjunto de dados se destaca por abranger diferentes condições de densidade de pedestres. Nas cenas mais movimentadas, os principais desafios incluem oclusões causadas por pedestres e objetos no ambiente, além de trajetórias curtas, pois os indivíduos frequentemente entram e saem rapidamente do campo de visão das câmeras ou assumem posturas diferentes da postura ereta, essencial para a calibração baseada em torsores. Neste estudo, o Si.U Dataset representa um ambiente realista com alta variabilidade no movimento dos pedestres, caracterizado por transições frequentes entre diferentes movimentos e posturas. Essas particularidades, aliadas à configuração multicâmera, fazem dele uma base de dados desafiadora e relevante para as avaliações conduzidas neste trabalho.

O conjunto de dados originalmente utilizado para validar a técnica TorsorCalib não foi incluído nos experimentos deste trabalho, uma vez que não se encontra publicamente disponível. Essa indisponibilidade inviabilizou a replicação dos testes, com testes de sanidade, e a análise direta dos resultados apresentados pelos autores, limitando a possibilidade de comparação com as avaliações realizadas neste estudo.

4.2.2 Anotação dos Dados

A técnica é testada em cenários significativamente mais desafiadores do que aqueles apresentados nos estudos de (TRUONG et al., 2019). Enquanto as abordagens avaliadas não oferecem garantias de bons resultados em cenários com mais de sete pedestres, os ambientes analisados neste trabalho contam com dezenas de indivíduos em movimento simultâneo. Esse fator aumenta a complexidade da calibração, pois o grande número de interações e possíveis oclusões entre pedestres dificulta a obtenção de amostras confiáveis e a correspondência precisa dos pontos de calibração entre as diferentes câmeras.

Dado esse contexto, a anotação manual dos dados dos pontos-chave dos pedestres será adotada para isolar fatores de influência e garantir a precisão da análise. Primeiramente, a escolha do pedestre-alvo para calibração precisa ser bem definida em um cenário real, onde os critérios de seleção impactam diretamente nos resultados. A anotação permite avaliar visualmente as rotas percorridas e a qualidade das detecções, tornando os dados das juntas mais precisos e coerentes entre as vistas. Além disso, ambientes não controlados frequentemente limitam a diversidade das trajetórias dos pedestres, tendo trajetórias predominantemente retilíneas, tornando a calibração mais desafiadora.

Outro fator é a correspondência de pontos-chave em todas as câmeras. As técnicas dependem fortemente dessas estimativas, e qualquer inconsistência no rastreamento pode comprometer a calibração. Além disso, a coleta de dados de forma consecutiva permite modelar melhor o comportamento natural dos pedestres. Isso gera uma exigência muito grande para os métodos de rastreamento de pedestres e re-identificação de pessoas, que não garantem bons resultados em cenários tão dinâmicos. Também, a anotação evita erros causados pelo registro de dados em deslocamentos descontínuos, causados, por exemplo, por oclusões prolongadas, ou movimentos em superfícies irregulares, como escadas e desníveis, que podem distorcer as estimativas 3D.

Assim, a anotação de dados, feita por um anotador, ao permitir isolar fatores de influên-

cia, torna possível relacioná-los com os resultados obtidos, proporcionando uma investigação direcionada para o entendimento de elementos que impactam a calibração no cenário real.

Nesse contexto, para o processo de anotação e verificação de dados é desenvolvida uma ferramenta de anotação, que recebe como entrada os pontos-chaves extraídos dos pedestres e gera dados anotados prontos para calibração pelas técnicas TorsorCalib ou MovingCalib. Os pontos-chaves são extraídos das imagens de todas as vistas selecionadas, pelos detectores de pose humana empregados em cada técnica. Além disso, a ferramenta emprega a visualização e registro de informações, permitindo um processo estruturado e detalhado de anotação.

Inicialmente, as imagens do dataset utilizado no experimento são analisadas visualmente para identificar as rotas percorridas pelos pedestres. Esse procedimento tem como objetivo selecionar o pedestre-alvo de calibração. A escolha do pedestre-alvo para calibração deve garantir amostras bem distribuídas e variadas, evitando redundância de dados causadas por trajetórias muito curtas ou lineares. Em ambientes não controlados, a movimentação restrita dos pedestres pode dificultar a obtenção de bons pontos de calibração. Além disso, a coleta deve ser contínua para minimizar perdas por oclusões prolongadas. Pedestres muito pequenos ou com trajetórias curtas podem comprometer a calibração devido a possíveis imprecisões nas detecções dos pontos chave. Assim, é essencial garantir um número adequado de amostras para compensar outliers e variações sutis na postura dos pedestres. Adota-se como 20 a quantidade mínima de frames para um pedestre-alvo de calibração.

Após esta escolha, o fluxo de trabalho para anotação é iniciado e está representado na Figura 7, que ilustra as etapas realizadas para processar e registrar os dados de cada frame de maneira sistemática.

As etapas de anotação dos dados são:

1. **Plotar dados de detecção no frame atual:**

   Inicialmente, os dados do esqueletos dos pedestres no frame corrente são visualizados, um por vez. Cada detecção de pedestre é exibida sequencialmente, permitindo ao anotador inspecionar as informações com clareza e precisão.

2. **Verificar se os dados pertencem ao pedestre selecionado:**

   Para cada detecção, avalia-se se ela corresponde ao pedestre-alvo de calibração.

   - **Não:** Caso os dados não pertençam ao pedestre-alvo, o sistema avança para o próximo pedestre detectado no mesmo frame.

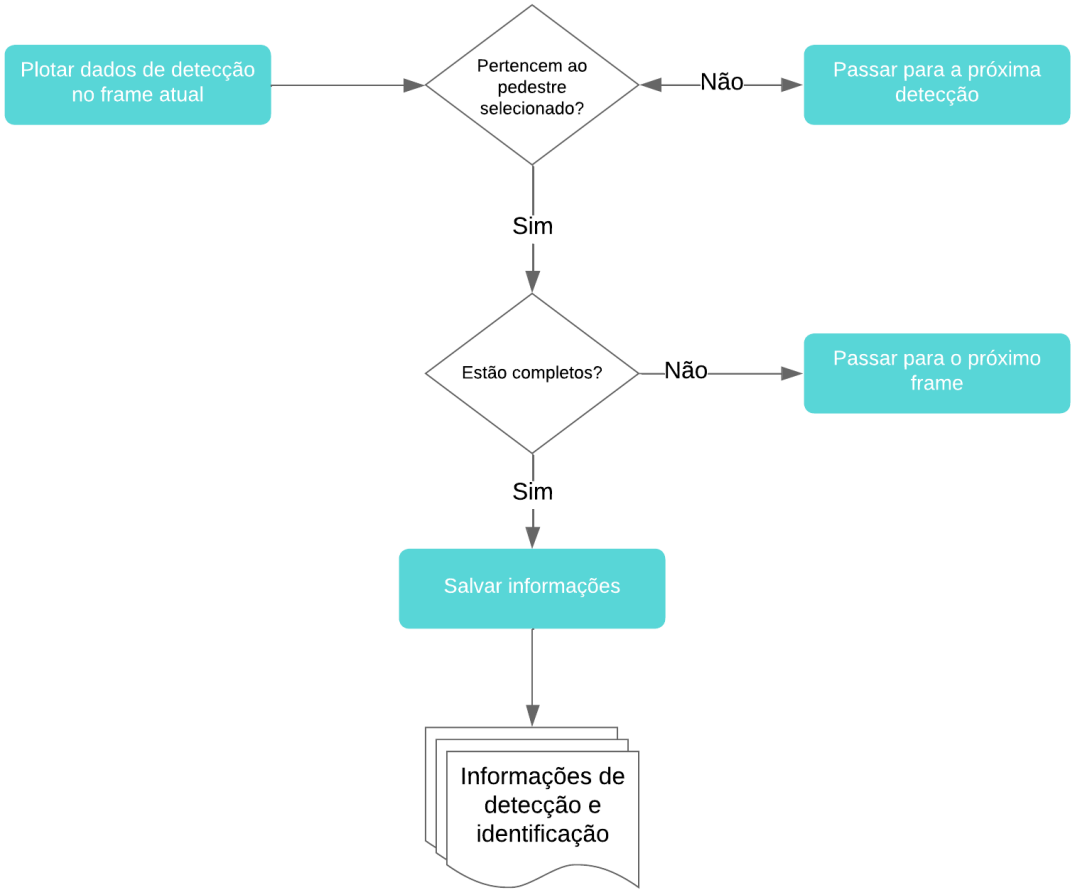

Figura 7 – Processo sequencial de anotação dos dados de detecção em cada frame, começando pela plotagem dos dados do torsor ou do esqueleto completo. A seguir, são verificados se os dados pertencem ao pedestre selecionado para a calibração. Caso positivo, o próximo passo é verificar se os dados de detecção estão completos. Se as articulações estão completas, as informações são registradas e salvas; caso contrário, o processo avança para o próximo frame. Esse processo garante a seleção de dados completos para a calibração precisa das câmeras. Fonte: Elaborado pelo autor.

- **Sim:** Se os dados forem do pedestre-alvo, o processo segue para a próxima etapa.

3. **Verificar se os dados de detecção estão completos:**

   Uma vez identificado o pedestre-alvo, o anotador confirma e o código analisa se o esqueleto inclue todas as articulações válidas. Essa verificação é importante para garantir que apenas informações completas sejam usadas no processo de calibração.

   - **Não:** Se os dados estiverem incompletos, o sistema avança para o próximo frame, descartando a detecção atual.
   - **Sim:** Se os dados estiverem completos, a próxima etapa é realizada.

4. **Salvar as informações de detecção e identificação:**

   Para as detecções validadas como pertencentes ao pedestre selecionado e com dados completos, as informações são registradas. Isso inclui as coordenadas das articulações, o ID do pedestre, o número do frame e o número da câmera.

A repetição deste ciclo para cada frame permite a anotação de rotas ao longo do tempo em todas as vistas. A Figura 8 mostra uma representação do torsor de um pedestre anotado para um frame em todas as vistas do Wildtrack.

Figura 8 – Anotação de dados de um frame para todas as vistas. Fonte: Elaborado pelo autor.

4.2.2.1 Verificação das Anotações

Para garantir a consistência e correspondência das anotações, é implementado um processo de verificação dos dados. Esse processo inclui a visualização dos dados de detecção dos pedestres anotados na etapa anterior, sobrepostos às imagens capturadas por cada câmera. A plotagem direta dos dados sobre as imagens permite uma análise visual detalhada para confirmar que as articulações anotadas correspondem corretamente aos pedestres presentes na cena. Esse procedimento identifica e corrige possíveis erros nos dados gerados pelo anotador

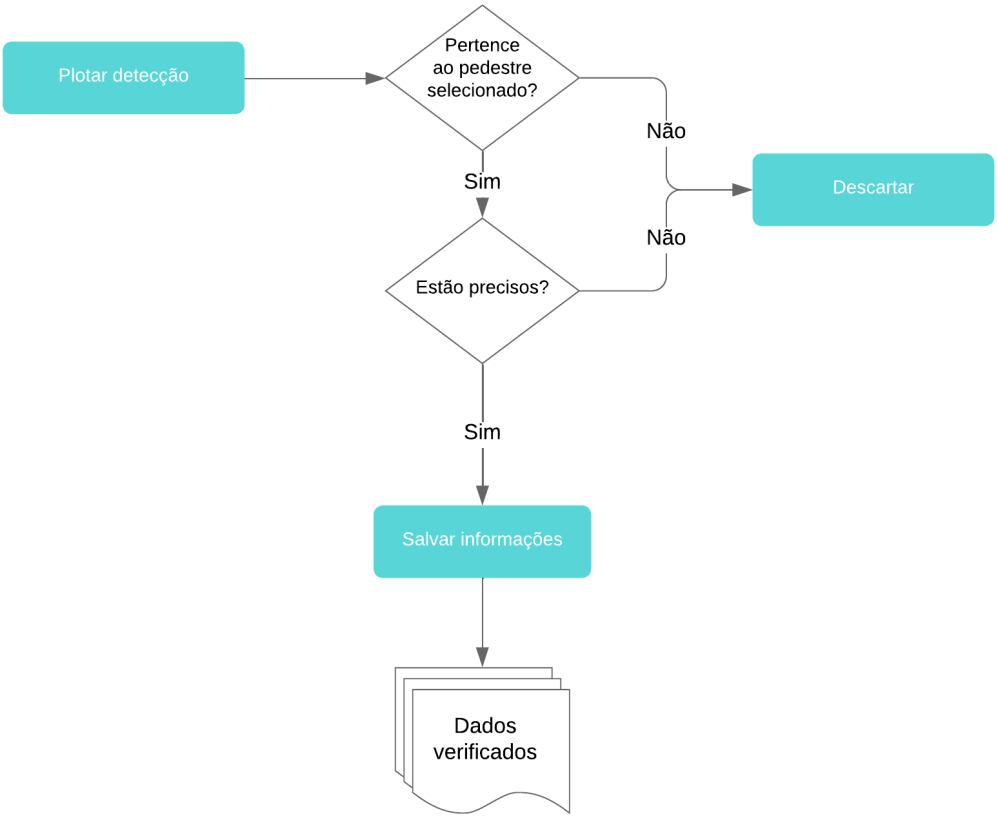

Figura 9 – No processo de verificação inicial dos dados, cada detecção de pedestre anotada é verificada visualmente para validar se: as articulações anotadas correspondem corretamente ao pedestre-alvo de calibração e elas não estão altamente imprecisas. Caso os dados não pertençam ao pedestre selecionado ou não sejam razoavelmente precisos, eles são descartados. Caso sejam, eles são salvos. Fonte: Elaborado pelo autor.

e pelo detector de pose, como falhas de anotação ou grandes imprecisões na detecção de articulações.

O fluxo do processo de verificação inicial é apresentado na Figura 9, que descreve as etapas para verificação das anotações quanto à correspondência com o pedestre-alvo de calibração e precisão das articulações. Além disso, existe o passo de verificação da correspondência dos dados nas vistas, descrito posteriormente.

As etapas de verificação inicial dos dados são:

1. **Plotar a detecção no frame correspondente:** Cada detecção salva durante a etapa de anotação é visualizada no frame correspondente.

2. **Verificar se a detecção pertence ao pedestre-alvo de calibração:** A correspon-

dência entre as articulações plotadas e os pedestre-alvo é visualmente avaliada.

- **Não:** Se a detecção não corresponde ao pedestre, ela é descartada.
- **Sim:** Caso positivo, a análise prossegue para a próxima etapa.

3. **Verificar a precisão dos dados:** A correspondência entre as posições das articulações detectectadas e às posições reais do pedestre na cena são analisadas visualmente.
   - **Não:** Caso sejam identificadas altas inconsistências ou falhas na detecção de qualquer articulação, os dados são descartados.
   - **Sim:** Se os dados forem considerados precisos ou razoavelemente precisos, eles são salvos.

Após concluir o processo de verificação inicial dos dados, começa a etapa de verificação dos frames válidos para o processo de calibração. Nesse estágio, são analisados os frames em que os dados de detecção do pedestre-alvo estão simultaneamente disponíveis nas vistas selecionadas. Apenas os frames que apresentam dados correspondentes em todas as vistas são incluídos no conjunto de calibração.

4.3 EXPERIMENTOS

A técnica TorsorCalib é reimplementada nesse trabalho utilizando a linguagem Python e com as principais bibliotecas sendo: OpenCV (BRADSKI, 2000), Scipy (VIRTANEN et al., 2020) e Numpy (HARRIS et al., 2020). As adaptações realizadas incluem: a integração da técnica com a ferramenta de anotação de dados, com o detector de pose humana AlphaPose (FANG et al., 2022), ao invés do OpenPose (CAO et al., 2019), usado na implementação original do artigo. A escolha do AlphaPose é justificada pelo uso deste detector em outras tecnologias utilizadas em pesquisas parceiras (LIMA et al., 2021), objetivando assim a facilidade na integração entre os scripts de calibração e essas tecnologias. Além de que o AlphaPose possui performance equivalente ao OpenPose, o que não prejudica a qualidade do experimento.

A partir disso, os experimentos são conduzidos utilizando o código implementado, aplicando-o aos quatro datasets apresentados na metodologia. Também, é utilizada a amostragem aleatória de dados empregada pelos autores da técnica. As tabelas apresentadas a seguir mostram as configurações dos experimentos com a técnica TorsorCalib, bem como os resultados obtidos para cada dataset. A Tabela 1 descreve os datasets utilizados, as vistas das câmeras envolvidas

em cada experimento e a quantidade de frames, que corresponde ao número de detecções de conjuntos de ombros e pés dos pedestres utilizados para calibração. Já a Tabela 2 apresenta os erros de reprojeção em pixels para cada vista de câmera e a média dos erros para cada dataset.

Tabela 1 – Configuração dos experimentos com a TorsorCalib. Fonte: Elaborado pelo autor.

Dataset	**Vistas**	**Quantidade de frames**
EPFL Campus Sequence - Campus 4	c0, c1, c2 (Figura 4)	85
Si.U	gopro01, gopro02, gopro03 (Figura 6)	56
Wildtrack	c1, c5, c7 (Figura 5)	106

Tabela 2 – Resultados dos experimentos com TorsorCalib. Fonte: Elaborado pelo autor.

Dataset	**Erro de Reprojeção (pixels)**			
	Vista I	**Vista II**	**Vista III**	**Médio**
EPFL Campus Sequence	1008,5	1852,3	1582,3	1537,5
Si.U	420,5	950,1	620,3	681,7
Wildtrack	700,4	430,6	340,3	489,1

Devido aos resultados com altos erros de reprojeção apresentados na Tabela 2, pode-se ver que o desempenho da TorsorCalib pode ser melhor investigado, especialmente por meio da seleção de dados que isolem fatores de influência.

No experimento realizado com o EPFL Campus Sequence, o erro de reprojeção médio das três vistas é de 1537,5 pixels, um valor elevado que pode ter relação com fatores relacionados às condições do dataset. O pedestre alvo de calibração está representado na Figura 10, assim como os resultados estão representados na figura 11. Um dos fatores que pode ter contribuição para esse erro é o número de frames utilizados na calibração, que totaliza 85. Esse número reduzido de frames reflete diretamente a quantidade de dados disponíveis para estimar os parâmetros de calibração e é neles que está descrita a rota do pedestres. Os fatores observados que contribuem para essa baixa quantidade de esqueletos são a área de sobreposição entre as câmeras e as oclusões enfrentadas pelo pedestre durante a sua trajetória. Uma menor área de sobreposição, em um ambiente dinâmico, tende a gerar pontos cegos quando os pedestres não são vistos por todas as câmeras, impedindo sua visualização e detecção de seus esqueletos correspondentes, limitando a quantidade de informações úteis para a calibração, além de que

pode limitar a amplitude da rota. Já oclusões não são desejadas, porque quanto mais ocluso o pedestre está, menos resultados de detecção dos pontos médios de ombros e pés é possível extrair.

Figura 10 – Pedestre alvo de calibração - EPFL Campus Sequence. Fonte: Elaborado pelo autor.

Figura 11 – Resultados de calibração automática usando o EPFL dataset. Fonte: Elaborado pelo autor.

Além disso, as trajetórias contidas neste dataset estão caracterizadas visualmente como retilíneas. Isso contribui para que elas fiquem concentradas em algumas áreas das imagens, sem grande amplitude de trajetória quando consideramos sua projeção 2D, o que também pode influenciar os resultados segundo os autores da técnica (TRUONG et al., 2019). Também,

para eles, se as amostras forem muito próximas umas das outras, como alguém caminhando lentamente em uma área pequena, as informações podem ser redundantes e não agregar valor significativo à calibração. Uma hipótese é que essa limitação na variedade das trajetórias pode ter comprometido a abrangência das estimativas dos parâmetros de calibração, influenciando negativamente a precisão dos resultados obtidos. Esses aspectos indicam a necessidade de mais dados e maior diversidade nas trajetórias para tentativa de melhoria na precisão da calibração em cenários com características semelhantes.

O experimento realizado com o Si.U Dataset, tem o pedestre alvo de calibração da Figura 12, e o erro de reprojeção médio é de 681,7 pixels, figura 13, um valor inferior ao encontrado no EPFL Campus Sequence, mas ainda elevado e considerado insatisfatório para calibração precisa. A principal vantagem do Si.U Dataset em relação ao EPFL é a maior variabilidade das trajetórias dos pedestres, que resulta em uma distribuição mais ampla das projeções 2D nas imagens. Essas trajetórias mais amplas são acompanhadas por variações nas posturas e transições dos pedestres, que caracterizam um cenário mais dinâmico.

Figura 12 – Pedestre alvo de calibração - Si.U. Fonte: Elaborado pelo autor.

No entanto, isso prejudica a quantidade de frames disponíveis para calibração, que é de apenas 56, uma vez que, para a calibração, o pedestre precisa estar na mesma postura durante toda trajetória, não sendo considerados os frames em que ele se senta, por exemplo. O alto nível de oclusão também representa um desafio significativo na quantidade de esqueletos identificados. Além da saída dos pedestres da área de sobreposição das vistas, o dataset inclui obstruções adicionais, como objetos e vegetação, que reduzem ainda mais a quantidade de dados utilizáveis para a calibração. Esse cenário de oclusões e posturas não eretas limita a quantidade de dados e reitera a necessidade de uma amostra mais ampla para investigação na

Figura 13 – Resultados de calibração automática usando o Si.U dataset. Fonte: Elaborado pelo autor.

precisão dos resultados.

No experimento realizado com o Wildtrack Dataset e com pedestre alvo da Figura 14, o erro de reprojeção médio é de 489,1 pixels, Figura 15, um valor inferior ao encontrado nos experimentos anteriores, mas ainda elevado. Esse conjunto de dados possui trajetórias visivelmente amplas e diversificadas, uma vez que os pedestres se deslocam em várias direções, incluindo quatro direções principais e outras direções oblíquas. No entanto, embora as trajetórias no Wildtrack sejam variadas e visualmente amplas, a presença de oclusões severas devido à movimentação de outros pedestres próximos prejudica a quantidade e a qualidade de dados utilizáveis para a calibração. As oclusões, causadas pelas interações entre os pedestres nas vistas calibradas, afetam a detecção dos pontos de referência necessários para o processo de calibração. Além disso, casacos e mochilas também confundem o detector de pose humana. Ainda assim, o número de frames utilizáveis no Wildtrack foi maior, totalizando 106, o que oferece uma quantidade mais substancial de dados em comparação aos experimentos anteriores.

O artigo original do TorsorCalib apresenta erros de reprojeção menores, variando de 3.69 pixels com 300 frames a 3.98 pixels com 20 frames. No entanto, esses resultados são obtidos em um dataset próprio, indisponível para testes neste trabalho. A ausência dessas imagens impede a análise das rotas dos pedestres utilizadas nos experimentos originais. Com base nas imagens

Figura 14 – Pedestre alvo de calibração - Wildtrack. Fonte: Elaborado pelo autor.

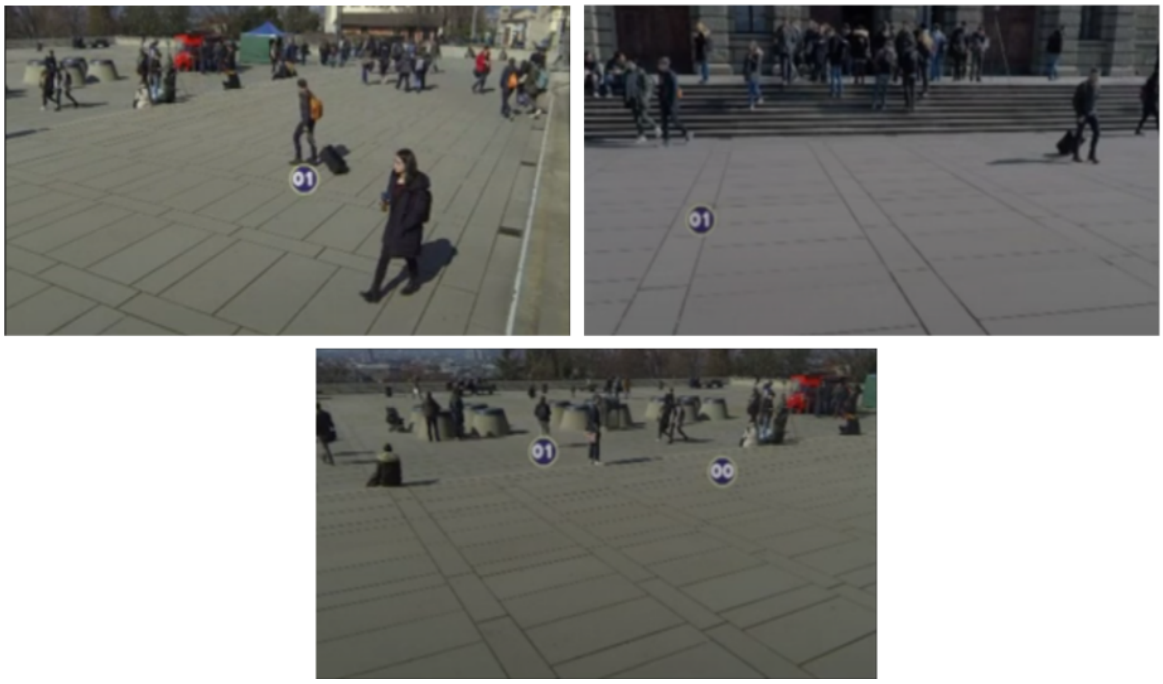

Figura 15 – Resultados de calibração automática usando o Wildtrack dataset. Fonte: Elaborado pelo autor.

publicadas no artigo, que possuem resolução VGA, os cenários aparentam ser mais controlados, com baixa incidência de oclusões e câmeras posicionadas em ângulos elevados. Esses fatores podem ter favorecido o desempenho da técnica. Dessa forma, uma possível abordagem para melhorar os resultados do TensorCalib seria aplicá-lo em cenários semelhantes. No entanto, este estudo busca avaliar a técnica em condições mais próximas da realidade, testando sua robustez em ambientes desafiadores.

Portanto, a TorsorCalib pode ter seu desempenho melhor investigado, principalmente através da seleção de dados que isolem fatores de influência. Ou seja, a aplicação deste método

deve seguir com atenção à seleção criteriosa dos dados, buscando testar de forma isolada fatores que podem impactar negativamente a calibração. Aspectos como a variedade e distribuição de trajetórias podem ser estudados para verificar a interferência destes nos parâmetros de calibração. Além disso, a incidência de oclusões pode ser reduzida sempre que possível, pois limita a quantidade de informações úteis extraídas das imagens para testes.

5 MOVINGCALIB

Ao longo do desenvolvimento deste trabalho, optou-se por uma reformulação no enfoque metodológico. Em vez de priorizar uma experimentação extensa em múltiplos cenários, como inicialmente proposto, foi adotada uma abordagem voltada para o isolamento e análise mais controlada dos fatores que influenciam a calibração automática. Essa decisão foi motivada pela necessidade de compreender com maior profundidade as variáveis críticas que afetam o desempenho das técnicas avaliadas, incluindo inclusive a análise de uma nova técnica.

Nesse contexto, dois tipos distintos de avaliação foram conduzidos utilizando o conjunto de dados Wildtrack, cuja escolha se justifica pelo seu equilíbrio entre quantidade de dados, qualidade das anotações e diversidade de rotas percorridas pelos pedestres. O uso deste dataset permitiu realizar análises mais direcionadas com o método MovingCalib, focando na investigação específica do comportamento da calibração em condições bem controladas, sem comprometer a representatividade dos dados.

Essa mudança metodológica não visa à generalização estatística, mas sim à geração de insights sobre os fatores que limitam ou favorecem o uso prático de calibração baseada em pedestres em cenários reais.

5.1 MÉTODO

As etapas de seleção da técnica, implementação ou adaptação do código, anotação e de verificação de dados e experimentos, análise dos resultados descritas no capítulo anterior correspondem também ao método adotado nos experimentos com a segunda técnica selecionada para os testes, (LEE et al., 2022). A seguir serão especificados apenas os pontos relativos a especificidades da técnica selecionada.

5.1.1 Seleção de Técnica

Com base nos critérios apresentados no capítulo anterior, a segunda técnica selecionada é:

- **MovingCalib:** Usa o movimento de articulações corporais humanas para estimar os parâmetros extrínsecos em redes de câmeras (LEE et al., 2022).

5.1.2 Implementação ou Adaptação de Técnica

No caso do MovingCalib o código-fonte está disponível publicamente [1] e são necessárias apenas adaptações. Esta também recebe como entrada os dados vindos do código de anotação de dados, ao invés da saída do detector de pose humana.

5.1.3 Avaliação das Técnicas

Foram realizados testes com o objetivo de isolar melhor fatores de influência da calibração, ao invés de priorizar a experimentação extensiva em diferentes cenários. Nesse intuito, dois tipos de avaliações são realizadas usando o Wildtrack. Este dataset foi escolhido para os experimentos com o MovingCalib devido ao seu equilíbrio entre a quantidade de dados e a variação de rotas disponível.

Na primeira avaliação, a técnica é aplicada e o erro de reprojeção é calculado e discutido. Para isso, são encontrados dois pedestres: aquele cuja calibração apresenta o menor erro de reprojeção e aquele que aparece em mais frames. O primeiro pedestre ajuda a entender características que fazem uma calibração boa. O segundo ajuda a explorar fatores como a quantidade de frames, tipo de rotas e qualidade da detecção de pedestres. O segundo pedestre apareceu em 400 frames consecutivos. Assim, além de calcular o erro de reprojeção na totalidade dos 400 frames, é feito experimentos para medir isso também nos 10, 20, 50, 100, 200 e 300 frames iniciais. Também é avaliado se há relação entre o erro de reprojeção e a qualidade da detecção dos pedestres. Para isso, é usado o score de detecção de esqueletos 2D como uma medida de qualidade (Subseção 5.1.6).

A segunda avaliação focou em validar a consistência geométrica da cena quando observada com a câmera calibrada usando geometria epipolar. Para isso, foi calculada a distância média do ponto para a sua linha epipolar.

5.1.3.1 Dataset de Avaliação

Os experimentos ocorrem com três vistas do Wildtrack dataset 4.2.1.2 e com um único pedestre-alvo para calibração. Como pontuado, o Wildtrack se destaca pela diversidade de trajetórias dos pedestres e apresenta percursos longos na trajetória dos pedestres. Este da-

[1] Disponível em <https://github.com/kyotovision-public/extrinsic-camera-calibration-from-a-moving-person>

taset foi escolhido para os experimentos com o MovingCalib devido ao seu equilíbrio entre a quantidade de dados e a variação de rotas disponível.

Optou-se por não utilizar outros datasets, como o Panoptic Dataset, empregado no trabalho original, devido a menor complexidade em comparação aos cenários urbanos reais. Embora estes conjuntos de dados sejam bem estruturados e amplamente utilizados, eles não representam ambientes controlados, com menor densidade de pedestres e baixa ocorrência de oclusões, o que pode comprometer a validade dos resultados no contexto desta investigação.

5.1.4 Anotação dos Dados

Os dados são anotados com a ferramenta de anotação de dados descrita em 5.1.4.

5.1.5 Métricas de Avaliação

As métricas utilizadas para avaliação são o erro de reprojeção, a distância epipolar, que foram explicadas no Capítulo 2 e o score de detecção médio que está explicado a seguir.

5.1.6 Score de Detecção Médio

O detector de pose humana OpenPose (CAO et al., 2019) determina uma nota de confiança para cada uma das articulações que formam o esqueleto de um pedestre. Essa nota é o score de detecção das articulações e, quanto maior o valor, maior a confiança na qualidade da detecção. Consequentemente, esse score pode ser usado para determinar a qualidade das detecções.

Neste trabalho, o score do pedestre é definido como a média dos scores de todas as juntas do esqueleto:

$$S_p = \frac{1}{J} \sum_{j=1}^{J} s_j, \tag{5.1}$$

onde J representa o número total de juntas do esqueleto e s_j é o score atribuído à junta j.

O score de detecção médio para um conjunto com n frames é então definido como:

$$S_f = \frac{1}{n} \sum_{i=1}^{n} S_{p,i}, \tag{5.2}$$

onde S_f é o score do pedestre no frame.

Por fim, o score de detecção médio para um experimento que envolve k pares de vistas é calculado como:

$$S_m = \frac{1}{k} \sum_{v=1}^{k} S_{f,v}, \tag{5.3}$$

onde S_m é o score médio do pedestre nos frames do sistema de câmeras.

OpenPose retorna scores no intervalo de $[0, 1]$, onde valores próximos de 1 indicam alta confiança na detecção, enquanto valores baixos sugerem maior incerteza ou erros potenciais.

5.2 EXPERIMENTOS

A técnica MovingCalib é adaptada a partir do seu código original, em Python, para integrar-se ao código de anotação de dados. O código original não suporta a saída do detector de poses humanas quando mais de uma pessoa é detectada na cena, pois está desenvolvido para um único pedestre, o que impede sua execução. Para os experimentos com a MovingCalib, o código de anotação de dados é adaptado para corresponder à entrada exigida pela técnica. O OpenPose é mantido como detector de pose humana, pois já é a ferramenta utilizada na implementação original da MovingCalib, propiciando a inclusão de alterações mínimas nos códigos para integração. Com as adaptações, a saída de dados do código de anotação passa todos os pontos-chave do esqueleto do pedestre anotado como entrada para a técnica.

Para os experimentos com esta técnica é utilizado apenas o Wildtrack Dataset devido ao seu equilíbrio entre o número de frames e a variação de rotas disponíveis nesse conjunto de dados. Considera-se que a diversidade de cenários explorados nos experimentos com a técnica TorsoCalib ajudou a identificar possíveis fatores de influência presentes nos conjuntos de dados, mas com a técnica MovingCalib e o Wildtrack pode-se explorar mais especificamente alguns desses fatores.

São utilizadas três vistas (c1, c6 e c7) do Wildtrack, que objetivam ampliar a quantidade de dados utilizáveis. Os pedestres presentes nas três vistas são anotados e utilizados em experimentos aplicando a técnica de calibração. Dentre os pedestres anotados, dois foram selecionados para análises mais detalhadas. O primeiro é identificado como Pedestre 1, Figura 16, e seu experimento tem as configurações resumidas na Tabela 3. Este pedestre é escolhido por apresentar o menor erro de reprojeção dentre todos os pedestres anotados. Seu valor é de 7.18 pixels, Tabela 4, considerando os 66 frames utilizados. Este erro é considerado baixo dentro

do contexto dos experimentos, tornando-o um caso relevante para avaliação da técnica. Esta técnica obtém erros de reprojeção de até 1,569 pixels, com Panoptic Dataset (JOO et al., 2017) , em seus testes originais.

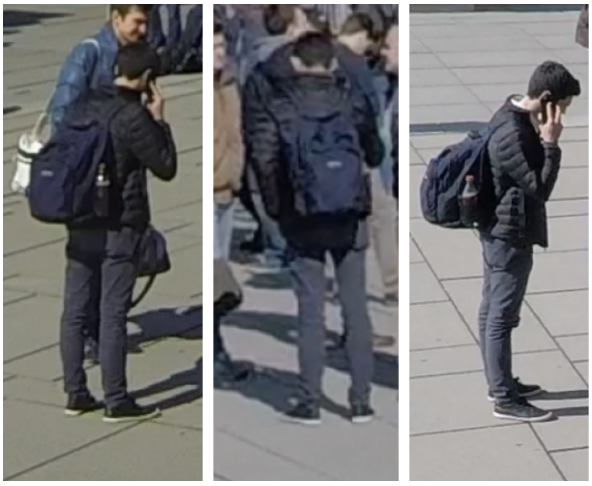

Figura 16 – Pedestre alvo de calibração 1, para experimentos com a técnica MovingCalib. Fonte: Elaborado pelo autor.

A rota deste pedestre é composta por duas partes distintas: uma dinâmica e uma com características estacionárias. Durante a parte dinâmica, o pedestre se desloca caminhando junto a outros pedestres, mas, em determinado momento, ele para e permanece próximo a outras pessoas na cena, interagindo com elas. Esse período estacionário é caracterizado por momentos de oclusão, embora sua posição varie muito pouco durante esse tempo.

Tabela 3 – Configuração do Experimento com o Pedestre 1. Fonte: Elaborado pelo autor.

Parâmetro	**Valor**
Dataset	Wildtrack
Vistas utilizadas	c1, c6, c7
Quantidade de frames usados	66

Tabela 4 – Resultados do experimento o Pedestre 1. Fonte: Elaborado pelo autor.

Métrica	**Valor**
Experimento 1 - Erro de Reprojeção Médio	
Erro de Reprojeção Médio (pixels)	7,18

O segundo pedestre, Figura 17, foi selecionado por possuir a maior quantidade de dados disponíveis, possibilitando uma análise mais detalhada dos fatores que influenciam a calibração. Em particular, foram investigadas a relação entre o número de frames utilizados e o erro de reprojeção, bem como a relação entre o score de detecção médio e o erro de reprojeção.

Figura 17 – Pedestre alvo de calibração 2, para experimentos com a técnica MovingCalib. Fonte: Elaborado pelo autor.

O pedestre 2 permanece visível por um período de tempo prolongado nas imagens. Embora ocorra oclusão em alguns momentos, a maior parte dos dados é aproveitada, resultando no uso de 400 frames para a calibração. Durante parte do trajeto, o pedestre interage com outras pessoas, permanecendo em uma posição mais estacionária na cena. A MovingCalib requer uma trajetória menos dinâmica em comparação à TorsorCalib (TRUONG et al., 2019).

Nesse sentido, é realizada uma análise para verificar se a quantidade de frames, que segue a progressão 10, 20, 50, 100, 200, 300 e 400, influencia significativamente os resultados desta técnica. Como sabemos, o número de dados pode ser importante para a calibração em técnicas que realizam a correspondência de dados (SVOBODA; MARTINEC; PAJDLA, 2005). O resultado está mostrado na Figura 18, onde o resultado do erro de reprojeção de cada teste pode ser visualizado no eixo Y, enquanto se mostra o número incremental de frames, correspondente a cada erro, deste conjunto de dados no eixo X.

O gráfico não possui uma tendência decrescente do valor de erro na medida que a quantidade de frames aumenta. Uma hipótese é que se as detecções de pedestres não estão precisas, inserir mais frames vai acrescentar mais ruído e, consequentemente, aumentando o erro. Trabalhos como o de (POSSEGGER et al., 2012) discutem especialmente precisão dos dados de entrada e até propõe um processo sistemático de retirada de outliers, mostrando a importância dessa investigação.

Para validar essa hipótese é analisada a qualidade da detecção do pedestre em relação ao erro de reprojeção. A métrica de qualidade escolhida é o score de detecção médio. Na Figura 19, o resultado do score de detecção médio de cada teste pode ser visualizado no eixo

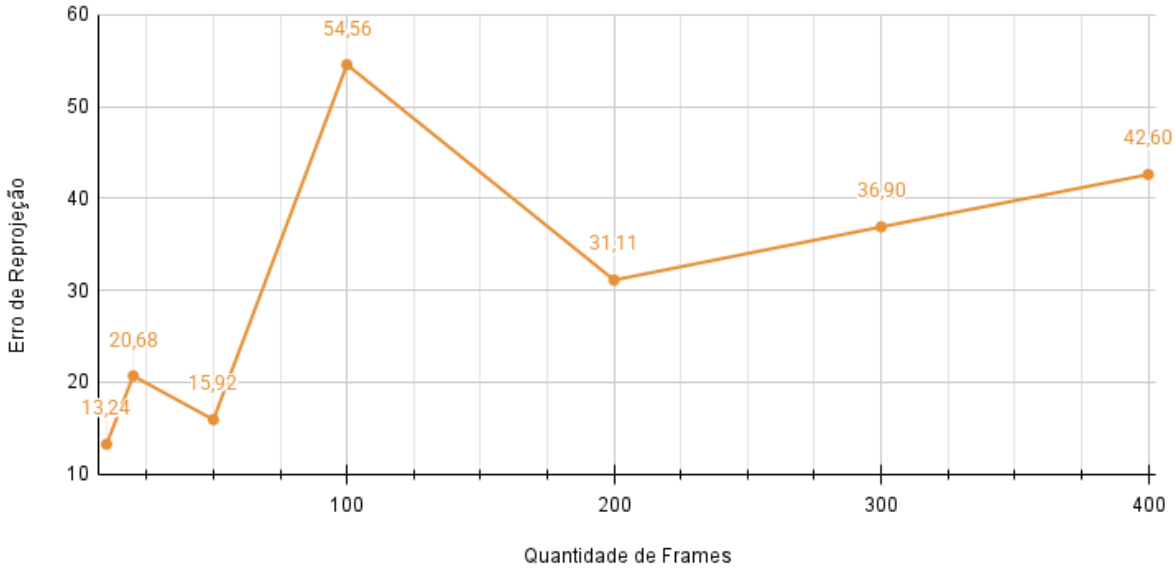

Figura 18 – Relação entre o erro de reprojeção e a quantidade de frames, com uma progressão de 10, 20, 50, 100, 200, 300 e 400 frames. Fonte: Elaborado pelo autor.

Y, enquanto a quantidade de frames em ordem crescente é mostrada no eixo X.

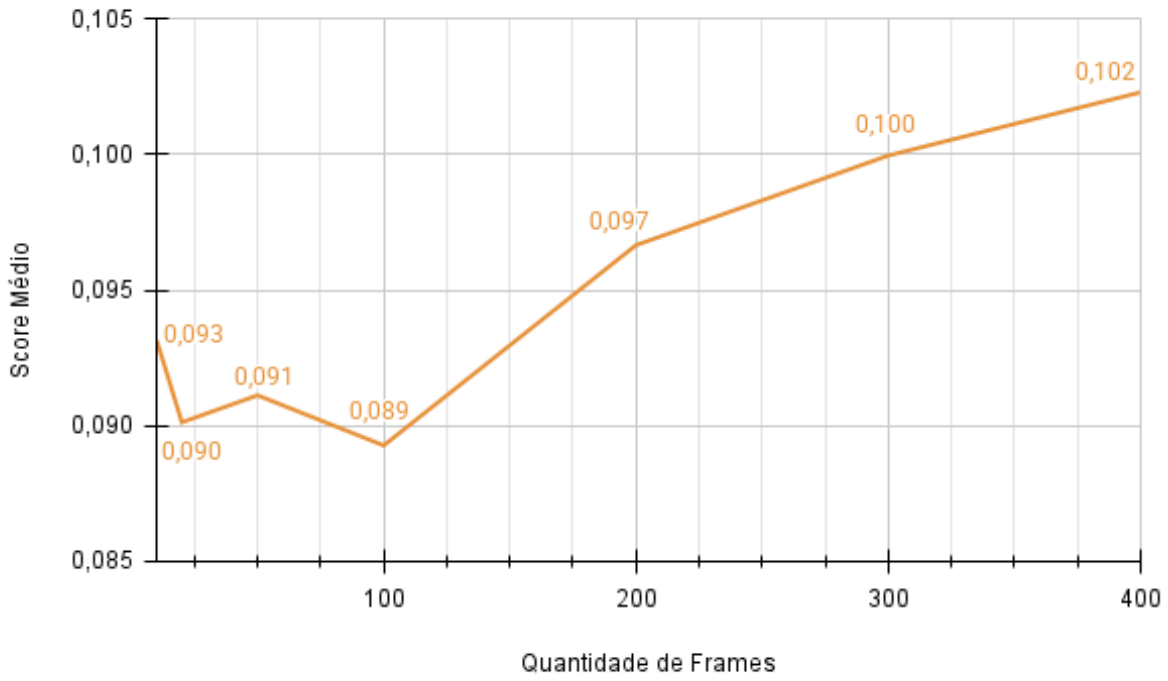

Figura 19 – Relação entre o score médio e a quantidade de frames, com uma progressão de 10, 20, 50, 100, 200, 300 e 400 frames. Fonte: Elaborado pelo autor.

Observa-se que os valores do score médio de detecção para os conjuntos de frames ana-

lisados do Wildtrack Dataset estão entre 0,089 e 0,102. Esses valores são consideravelmente baixos, o que indica uma baixa precisão na detecção de pedestres. Esse score médio é cerca de cinco vezes menor do que o obtido quando a calibração é feita usando o Panoptic (JOO et al., 2017), dataset em que o MovingCalib foi testado originalmente. Esse dataset apresenta um ambiente controlado, com pouca oclusão e o pedestre próximo à câmera. Essa diferença evidencia uma degradação significativa na qualidade da detecção em condições não controladas. Além disso, o fato de o score médio estar mais próximo de 0 que de 1 sugere que, apesar da visibilidade contínua dos pedestres ao longo do trajeto, a qualidade da detecção pode ser baixa no cenário real, mesmo após a verificação e refinamento dos dados.

Adicionalmente, ao analisar as Figuras 18 e 19, observa-se uma relação inversamente proporcional entre o score médio de detecção e o erro de reprojeção para até 200 frames. Ou seja, nos intervalos em que o score de detecção melhora, o erro de reprojeção tende a diminuir. Também, para um conjunto de 100 frames, o menor score médio registrado (0,089) está associado ao maior erro de reprojeção (54,56 pixels). Esse comportamento sugere que a redução da qualidade da detecção impacta negativamente a calibração em conjuntos menores de frames, que, por conterem menos amostras, são mais suscetíveis à introdução de ruídos nas detecções. Além disso, ao utilizar um número maior de frames, há um risco de que detecções imprecisas aumentem o ruído no processo de calibração, comprometendo a precisão final da calibração.

Esses resultados sugerem que a simples inclusão de um maior número de frames não é suficiente para garantir uma calibração mais precisa. A confiabilidade das detecções deve ser levada em consideração. Uma hipótese é que o tamanho da projeção dos pedestres na imagem possa comprometer a calibração. Já que quanto maior o tamanho do pedestre (ou o ponto de interesse), mais fácil é para os algoritmos de detecção prever suas posições. Em cenas com pedestres pequenos, o score médio pode tender a diminuir, devido à falta de detalhes. Isso pode ser investigado, por exemplo, através de estratégias que filtrem frames com baixa confiabilidade ou atribuam pesos diferentes às detecções conforme seu nível de confiança.

5.2.1 Experimentos da Distância Epipolar

Quando os parâmetros de calibração obtidos com o bom resultado do experimento com o Pedestre 1 foram testados em uma técnica de detecção de pedestres (LIMA et al., 2021), o desempenho da detecção não foi satisfatório. Os resultados da avaliação da distância do ponto

à linha epipolar estão resumidos na Tabela 5.

Tabela 5 – Resultados dos experimentos da distância epipolar. Fonte: Elaborado pelo autor.

Média Ground Truth	327.49
Desvio Padrão Ground Truth	2.55
Média Calibração Automática	710.43
Desvio Padrão Calibração Automática	484.53
Erro Absoluto Médio (MAE)	406.42
Erro Quadrático Médio (RMSE)	579.91
Teste t (p-valor)	0.1534

Para esse experimento, a hipótese nula considerada foi de que não há diferença significativa entre as distâncias epipolares obtidas com a calibração automática e o ground truth. Com base nos resultados apresentados, observou-se uma discrepância considerável entre as duas estimativas, evidenciada pelos altos valores de erro absoluto médio (406.42) e erro quadrático médio (579.91). Além disso, o desvio padrão elevado na calibração automática indica uma alta variabilidade nos erros, sugerindo falta de consistência na reconstrução da geometria epipolar.

O teste t revela que, embora as distribuições das distâncias epipolares apresentem diferenças relevantes, o p-valor obtido (0.1534) não é suficientemente baixos para rejeitar a hipótese nula em um nível de significância tradicional. Isso significa que, com base nesse teste, não é possível concluir que a calibração automática tenha um desempenho significativamente diferente da calibração baseada no ground truth. Essa discrepância e a falta de consistência nos resultados sugerem que, para melhorar o entendimento da calibração automática, seria necessário estudar mais o impacto dos fatores de influência.

Em geral, a diferença significativa entre as estimativas automáticas e o ground truth indica que a técnica MovingCalib ainda apresenta altos erros em sua calibração. O desvio padrão elevado nos resultados reforça a ideia de que a técnica pode ser vulnerável à qualidade dos dados usados para calibração. Para esta técnica, aspectos de rota do pedestre são considerados menos importantes, já que esta se mostrou robusta a pequenas quantidade de movimento em seus testes originais (LEE et al., 2022).

6 DISCUSSÕES GERAIS E APRENDIZADOS DO PROCESSO EXPERIMENTAL

6.1 DISCUSSÕES GERAIS

Os experimentos mostram que as técnicas avaliadas apresentam dificuldades de adaptação a cenários distintos daqueles em que foram originalmente validadas. Seja em ambientes mais controlados ou em contextos mais dinâmicos, desafios como oclusões frequentes tem impacto direto na qualidade da calibração.

Esta qualidade dos dados de entrada parece ser um fator de influência relevante no desempenho das técnicas, principalmente para conjuntos de frames menores, até 200 frames. Condições adversas, como iluminação inadequada e a presença de múltiplos pedestres, são elementos que podem comprometer a detecção das poses humanas e, consequentemente, afetar a calibração (GHARI et al., 2024). Assim, a dependência da técnica em relação à robustez dos detectores de pose deve ser considerada ao avaliar sua aplicabilidade em cenários reais.

Até aqui, dois fatores de influência principais podem ser explorados para o maior entendimento destes resultados: a qualidade das detecções das juntas e a rota do pedestre, ou seja, a distribuição de movimento presente nos frames utilizados. Como uma primeira hipótese, pode-se afirmar que, a calibração automática pode não ser precisa se os frames analisados não capturarem rotas adequadas, ou seja, uma trajetória bem distribuída no espaço e com detecções de qualidade para a técnica TorsorCalib. Isso pode enviesar os parâmetros de calibração. O teste sugerido é investigar como diferentes subconjuntos de frames, com variação na distribuição, afetam o erro epipolar e se a distribuição dos frames impacta a precisão da calibração. Além de estudar a qualidade da detecção de cada experimento, relacionando o score médio com o erro de reprojeção.

Uma segunda hipótese pode ser que, se as detecções das articulações forem ruidosas ou inconsistentes, o erro epipolar será maior e mais variável para a técnica MovingCalib. Isso ocorre porque as detecções de baixa qualidade afetam diretamente a triangulação dos pontos no espaço 3D, prejudicando a calibração. O teste proposto é analisar a relação entre o erro epipolar e o score médio das detecções das juntas para verificar esse impacto.

Com isso, os resultados obtidos reforçam que a calibração automática de redes de câmeras baseada em pedestres enfrenta desafios significativos na adaptação a ambientes dinâmicos reais. A variabilidade dos dados exige técnicas mais generalistas e capazes de lidar com diferentes

fontes de erro.

Por fim, embora as técnicas apresentem potencial para calibração automática em redes de câmeras multi-visão, avanços substanciais ainda são necessários para garantir sua aplicação em contextos reais. A coleta de dados mais controlados pode ser uma estratégia para isolar e entender melhor os fatores que afetam a calibração. O próximo capítulo apresenta recomendações e direções para estudos futuros, visando aprimorar a robustez dessas abordagens.

6.2 APRENDIZADOS DO PROCESSO EXPERIMENTAL

Ao longo dos experimentos realizados são identificadas limitações que fornecem uma base para orientar trabalhos futuros. Os principais fatores que influenciaram a calibração foram:

1. **Rotas dos pedestres:** As trajetórias impactam diretamente os resultados. Em especial para técnicas como a TorsorCalib, a presença de uma rota bem distribuída entre as câmeras contribui para a precisão da calibração, pois fornece uma amostragem espacial mais representativa. Este aspecto sugere que futuros trabalhos devem priorizar datasets onde seja possível controlar ou mapear as rotas dos pedestres. O objetivo é aprofundar os estudos para identificar se há melhores rotas para calibração de câmeras. No entanto, como os datasets existentes frequentemente carecem dessa característica, seria interessante criar um dataset onde rotas possam ser controladas e ajustadas para calibração automática. Isso poderá resultar em *guidelines* que orientem desenvolvedores e usuários a como realizar calibrações de maneira automática em cenários reais.

2. **Oclusões e densidade de pedestres:** Ambientes com muitas oclusões ou alta densidade de pedestres dificultaram a correspondência de pontos, comprometendo a eficácia das técnicas por perda de dados e diminuição da qualidade da detecção de pose humana. Uma direção importante seria investigar soluções que aumentem a robustez da calibração em cenários com alta densidade, como técnicas baseadas em aprendizado profundo para prever e corrigir detecções perdidas.

3. **Método de estimação da pose humana:** A escolha do método de pose pode ter impacto sobre os resultados. Métodos mais precisos tendem a contribuir para uma melhor estimativa da posição das articulações, o que melhora a triangulação e, por consequência,

a calibração. Por outro lado, métodos menos robustos geram detecções ruidosas, o que pode comprometer a precisão, especialmente na técnica TorsorCalib. Trabalhos futuros podem explorar a influência de diferentes detectores de pose de última geração, avaliando como suas acurácias se refletem nos erros de calibração.

4. **Quantidade de frames utilizados:** Os experimentos realizados neste trabalho não permitiram concluir de forma definitiva que a quantidade de dados influencia diretamente na precisão da calibração. Em alguns casos, conjuntos menores com detecções mais confiáveis produziram resultados mais consistentes do que sequências mais longas com dados ruidosos. Esse resultado sugere que a qualidade e a diversidade das informações nos frames podem ser mais determinantes que a quantidade absoluta de dados. Trabalhos futuros podem aprofundar essa análise, controlando melhor a variação nos dados e comparando diretamente subconjuntos com tamanhos distintos, mas qualidade similar de detecção.

5. **Quantidade e posicionamento das câmeras:** O número de câmeras e sua distribuição no ambiente também influenciaram diretamente os resultados. Trabalhos futuros devem explorar redes de câmeras mais densas e diversificadas em termos de ângulos de visão, especialmente em cenários controlados, para entender como maximizar a qualidade da calibração. Um ponto importante é garantir que a área de sobreposição dos equipamentos inclua a rota do(s) pedestre(s) de interesse por um período de duração maior que os datasets atuais proporcionam.

6. **Qualidade dos dados de entrada:** A análise mostrou que a qualidade das detecções, medidas pelo score médio de confiança, pode impactar direto nos resultados, como também é observado por (MOLINER; HUANG; ASTROM, 2021). Futuros experimentos podem se concentrar em melhorar os pré-processamentos dos dados, utilizando técnicas de filtragem ou mesmo aprendizado profundo para refinar as detecções antes da calibração.

6.3 DIRECIONAMENTO DE TRABALHOS FUTUROS

Os experimentos realizados proporcionaram uma compreensão valiosa das limitações e potencialidades das técnicas analisadas. Para avançar no desenvolvimento de calibração automática em redes de câmeras, é essencial adotar abordagens que combinem controle experimental

rigoroso com robustez em condições reais. A criação de datasets controlados, o uso de técnicas avançadas de aprendizado de máquina e a adaptação a cenários reais são caminhos promissores para superar os desafios observados e aprimorar a tecnologia para aplicações futuras.

6.3.1 Criação de um Dataset Controlado

Um dos principais desafios identificados foi a falta de datasets que ofereçam controle suficiente sobre variáveis como rotas de pedestres, iluminação, densidade de pessoas e oclusões. Propor a criação de um dataset sintético ou real, com controle rigoroso sobre essas condições, poderia avançar significativamente o campo. Esse dataset poderia incluir:

- Rotas pré-definidas e bem documentadas.
- Variedade de cenários de iluminação (natural e artificial).
- Número ajustável de pedestres.
- Anotação e variação sistemática nas posições e orientações das câmeras.

6.3.2 Exploração de Métodos Avançados

Futuras pesquisas podem integrar técnicas de aprendizado profundo para superar limitações observadas, como:

- Uso de modelos baseados em estimativa de pose humana para melhorar a correspondência de pontos.
- Aplicação de algoritmos de reidentificação de pedestres para lidar com oclusões e trajetórias dinâmicas.
- Desenvolvimento de modelos híbridos que combinem abordagens geométricas tradicionais com técnicas baseadas em deep learning.

6.3.3 Adaptação a Cenários Reais

Trabalhos futuros podem buscar expandir a aplicabilidade das técnicas a cenários reais, como áreas urbanas dinâmicas. Isso incluiria:

- Testes em condições climáticas e de iluminação variáveis.
- Análise de performance em ambientes com grande densidade populacional.
- Uso de redes de câmeras parcialmente sobrepostas, como as encontradas no Wildtrack Dataset.

7 CONCLUSÃO

Com o aumento da complexidade dos cenários de aplicação da visão computacional, especialmente em ambientes dinâmicos e com uma rede de câmeras, como áreas urbanas, a necessidade de técnicas automáticas de calibração tornam-se mais necessárias. Este trabalho, portanto, propôs-se a avaliar métodos automatizados de calibração em redes de câmeras usando pedestres em cenários reais.

A revisão da literatura mostrou ser possível a utilização de técnicas de detecção de pedestres para calibrar câmeras sem a necessidade de inserir padrões artificiais. Esses métodos exploram a capacidade de abordagens tradicionais e dos modelos de aprendizado profundo de extrair características robustas de pedestres. Essas são usadas para calibrar de formar automática um conjunto de câmeras.

Assim, os experimentos realizados ao longo deste trabalho foram projetados para avaliar duas técnicas de calibração automática em cenários reais. Eles buscaram levantar os principais elementos que podem influenciar este processo. O TorsorCalib foi testado em diferentes datasets. Os resultados mostraram que a sua precisão foi comprometida devido à quantidade de dados disponíveis devido a oclusões, assim como a qualidade das rotas. Isso indica que ajustes são necessários para lidar com a complexidade dos ambientes reais, especialmente em relação às rotas dos pedestres e à variação nas condições do cenário.

Já o MovingCalib apresentou um comportamento inversamente proporcional entre a qualidade da detecção de pose humana e o erro de reprojeção, como conjuntos com até 200 frames. Isso sugere que a qualidade de dados, que foi inferior no cenário real e dinâmico do Wildtrack, influencia os resultados de calibração significativamente, exigindo uma abordagem mais refinada para melhorar a calibração.

De modo geral, os experimentos mostraram que, embora as técnicas automáticas tenham evoluído, sua eficácia ainda é limitada em ambientes não estruturados. Questões como oclusões, densidade de pedestres e variação na qualidade dos dados impactam a calibração, tornando essencial o desenvolvimento de métodos mais robustos. O uso de estratégias híbridas, combinando abordagens tradicionais e aprendizado profundo, pode oferecer melhorias.

7.1 TRABALHOS FUTUROS

Para trabalhos futuros, recomenda-se a investigação de novas métricas que captem melhor as falhas da calibração em cenários dinâmicos. Além disso, testar as técnicas em diferentes datasets e explorar modelos de fusão de dados podem contribuir para a adaptação dos métodos às condições reais. A incorporação de redes neurais para prever e corrigir erros também pode ser um caminho promissor para aumentar a precisão da calibração automática.

7.2 CONTRIBUIÇÕES

As principais contribuições derivadas desta pesquisa são:

- Ampliação dos conhecimento da área de visão computacional, através da análise de fatores de influência, do cenário real, na calibração automática de câmeras, deixando lições aprendidas como ponto de partida para que pesquisas futuras desenvolvam técnicas mais robustas;
- Código de anotação dados de detecção integrado aos detectores de pose humana: AlphaPose e OpenPose;
- Uma publicação em planejamento a partir deste estudo, para comunidade da área de Visão Computacional e Calibração Automática de Câmeras;
- Publicação científica nas áreas de Reconhecimento de Atividades Humanas e Redes Neurais (CAVALCANTE et al., 2023), propiciada por conhecimentos derivados desta pesquisa;
- Outras duas publicações sendo planejadas nas áreas de Reconhecimento de Atividades Humanas e Redes Neurais.

REFERÊNCIAS

ABDEL-AZIZ, Y. I.; KARARA, H. M.; HAUCK, M. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Photogrammetric engineering & remote sensing*, Elsevier, v. 81, n. 2, p. 103–107, 2015.

BEN-ARTZI, G. Camera Calibration by Global Constraints on the Motion of Silhouettes. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017. p. 5344–5353. ISBN 978-1-5386-1032-9. Disponível em: <http://ieeexplore.ieee.org/document/8237832/>.

BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

CAO, Z.; HIDALGO, G.; SIMON, T.; WEI, S.-E.; SHEIKH, Y. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 43, n. 1, p. 172–186, 2019.

CAVALCANTE, A. F.; KUNST, V. H. d. L.; CHAVES, T. d. M.; SOUZA, J. D. de; RIBEIRO, I. M.; QUINTINO, J. P.; SILVA, F. Q. da; SANTOS, A. L.; TEICHRIEB, V.; GAMA, A. E. F. da. Deep learning in the recognition of activities of daily living using smartwatch data. *Sensors*, MDPI, v. 23, n. 17, p. 7493, 2023.

CHAVDAROVA, T.; BAQUÉ, P.; BOUQUET, S.; MAKSAI, A.; JOSE, C.; BAGAUTDINOV, T.; LETTRY, L.; FUA, P.; GOOL, L. V.; FLEURET, F. Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 5030–5039.

CHAVDAROVA, T.; FLEURET, F. Deep multi-camera people detection. In: IEEE. *2017 16th IEEE international conference on machine learning and applications (ICMLA)*. [S.l.], 2017. p. 848–853.

DIAS, L. A. *Estudo e Análise de Diferentes Métodos de Calibração de Câmeras*. 2015.

FANG, H.-S.; LI, J.; TANG, H.; XU, C.; ZHU, H.; XIU, Y.; LI, Y.-L.; LU, C. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

FAUGERAS, O.; LUONG, Q.-T.; PAPADOPOULO, T. *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. [S.l.]: MIT press, 2001.

GHARI, B.; TOURANI, A.; SHAHBAHRAMI, A.; GAYDADJIEV, G. Pedestrian detection in low-light conditions: A comprehensive survey. *Image and Vision Computing*, Elsevier, p. 105106, 2024.

GUAN, J.; DEBOEVERIE, F.; SLEMBROUCK, M.; HAERENBORGH, D. V.; CAUWELAERT, D. V.; VEELAERT, P.; PHILIPS, W. Extrinsic Calibration of Camera Networks Based on Pedestrians. *Sensors*, v. 16, n. 5, p. 654, maio 2016. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/16/5/654>.

HALPERIN, T.; WERMAN, M. An Epipolar Line from a Single Pixel. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe,

NV: IEEE, 2018. p. 983–991. ISBN 978-1-5386-4886-5. Disponível em: <https://ieeexplore.ieee.org/document/8354217/>.

HARRIS, C. R.; MILLMAN, K. J.; WALT, S. J. van der; GOMMERS, R.; VIRTANEN, P.; COURNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; KERKWIJK, M. H. van; BRETT, M.; HALDANE, A.; R'ıO, J. F. del; WIEBE, M.; PETERSON, P.; G'eRARD-MARCHANT, P.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E. Array programming with NumPy. *Nature*, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <https://doi.org/10.1038/s41586-020-2649-2>.

HARTLEY, R.; ZISSERMAN, A. *Multiple view geometry in computer vision*. [S.l.]: Cambridge university press, 2003.

HEIKKILA, J.; SILVÉN, O. A four-step camera calibration procedure with implicit image correction. In: IEEE. *Proceedings of IEEE computer society conference on computer vision and pattern recognition*. [S.l.], 1997. p. 1106–1112.

HöDLMOSER, M.; KAMPEL, M. Multiple Camera Self-calibration and 3D Reconstruction Using Pedestrians. In: BEBIS, G.; BOYLE, R.; PARVIN, B.; KORACIN, D.; CHUNG, R.; HAMMOUND, R.; HUSSAIN, M.; KAR-HAN, T.; CRAWFIS, R.; THALMANN, D.; KAO, D.; AVILA, L. (Ed.). *Advances in Visual Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. v. 6454, p. 1–10. ISBN 978-3-642-17273-1 978-3-642-17274-8. Series Title: Lecture Notes in Computer Science. Disponível em: <http://link.springer.com/10.1007/978-3-642-17274-8_1>.

JOO, H.; SIMON, T.; LI, X.; LIU, H.; TAN, L.; GUI, L.; BANERJEE, S.; GODISART, T. S.; NABBE, B.; MATTHEWS, I.; KANADE, T.; NOBUHARA, S.; SHEIKH, Y. Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

JORDAN, N. *OpenHSML - Open-source Hybrid Stereovision Matching Library*. 2021. <https://github.com/jordan-nowak/OpenHSML>. Acessado em 28 Fevereiro de 2025.

KLEMA, V.; LAUB, A. The singular value decomposition: Its computation and some applications. *IEEE Transactions on automatic control*, IEEE, v. 25, n. 2, p. 164–176, 1980.

LEE, S.-E.; SHIBATA, K.; NONAKA, S.; NOBUHARA, S.; NISHINO, K. Extrinsic Camera Calibration From a Moving Person. *IEEE Robotics and Automation Letters*, v. 7, n. 4, p. 10344–10351, out. 2022. ISSN 2377-3766, 2377-3774. Disponível em: <https://ieeexplore.ieee.org/document/9834083/>.

LETTRY, L.; DRAGON, R.; GOOL, L. V. Markov chain monte carlo cascade for camera network calibration based on unconstrained pedestrian tracklets. In: SPRINGER. *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II 13*. [S.l.], 2017. p. 400–415.

LIMA, J. P.; ROBERTO, R.; FIGUEIREDO, L.; SIMOES, F.; TEICHRIEB, V. Generalizable multi-camera 3d pedestrian detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2021. p. 1232–1240.

LIU, J.; COLLINS, R. T.; LIU, Y. Robust autocalibration for a surveillance camera network. In: *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. Clearwater Beach, FL, USA: IEEE, 2013. p. 433–440. ISBN 978-1-4673-5054-9 978-1-4673-5053-2 978-1-4673-5052-5. Disponível em: <https://ieeexplore.ieee.org/document/6475051>.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, p. 91–110, 2004.

MOLINER, O.; HUANG, S.; ASTROM, K. Better Prior Knowledge Improves Human-Pose-Based Extrinsic Camera Calibration. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. Milan, Italy: IEEE, 2021. p. 4758–4765. ISBN 978-1-72818-808-9. Disponível em: <https://ieeexplore.ieee.org/document/9411927/>.

NOWAK, J.; FRAISSE, P.; CHERUBINI, A.; DAURÈS, J.-P. Point clouds with color: A simple open library for matching rgb and depth pixels from an uncalibrated stereo pair. In: IEEE. *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. [S.l.], 2021. p. 1–7.

POSSEGGER, H.; RÜTHER, M.; STERNIG, S.; MAUTHNER, T.; KLOPSCHITZ, M.; ROTH, P. M.; BISCHOF, H. Unsupervised calibration of camera networks and virtual ptz cameras. In: *Proceedings of the Computer Vision Winter Workshop*. [S.l.: s.n.], 2012.

PUWEIN, J.; BALLAN, L.; ZIEGLER, R.; POLLEFEYS, M. Joint camera pose estimation and 3d human pose estimation in a multi-camera setup. In: SPRINGER. *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part II 12*. [S.l.], 2015. p. 473–487.

SVOBODA, T.; MARTINEC, D.; PAJDLA, T. A convenient multicamera self-calibration for virtual environments. *Presence: Teleoperators & virtual environments*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 14, n. 4, p. 407–422, 2005.

TEIXEIRA, L.; MAFFRA, F.; BADII, A. Scene understanding for auto-calibration of surveillance cameras. In: SPRINGER. *International Symposium on Visual Computing*. [S.l.], 2014. p. 671–682.

TEMPELAAR, W. J. Extrinsic Camera Calibration using Human-pose Estimations and Automatic Re-identification. 2022.

TRUCCO, E.; VERRI, A. *Introductory techniques for 3-D computer vision*. [S.l.]: Prentice Hall Englewood Cliffs, 1998. v. 201.

TRUONG, A. M.; PHILIPS, W.; DELIGIANNIS, N.; ABRAHAMYAN, L.; GUAN, J. Automatic Multi-Camera Extrinsic Parameter Calibration Based on Pedestrian Torsors †. *Sensors*, v. 19, n. 22, p. 4989, nov. 2019. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/19/22/4989>.

VIRTANEN, P.; GOMMERS, R.; OLIPHANT, T. E.; HABERLAND, M.; REDDY, T.; COURNAPEAU, D.; BUROVSKI, E.; PETERSON, P.; WECKESSER, W.; BRIGHT, J.; van der Walt, S. J.; BRETT, M.; WILSON, J.; MILLMAN, K. J.; MAYOROV, N.; NELSON, A. R. J.; JONES, E.; KERN, R.; LARSON, E.; CAREY, C. J.; POLAT, İ.; FENG, Y.; MOORE, E. W.; VanderPlas, J.; LAXALDE, D.; PERKTOLD, J.; CIMRMAN, R.; HENRIKSEN, I.; QUINTERO, E. A.; HARRIS, C. R.; ARCHIBALD, A. M.; RIBEIRO, A. H.; PEDREGOSA, F.;

van Mulbregt, P.; SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, p. 261–272, 2020.

ZHANG, Z. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 22, n. 11, p. 1330–1334, 2002.

ZHOU, K.; YANG, Y.; CAVALLARO, A.; XIANG, T. Learning generalisable omni-scale representations for person re-identification. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 44, n. 9, p. 5056–5069, 2021.