

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**METODOLOGIA FORENSE DE ESTIMATIVA
SEMIAUTOMÁTICA DE ALTURA HUMANA EM VÍDEOS
DE SISTEMAS DE VIGILÂNCIA**

ADMILSON GONÇALVES JÚNIOR

ORIENTADOR: HÉLIO PEDRINI

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA
ÁREA DE CONCENTRAÇÃO INFORMÁTICA FORENSE E
SEGURANÇA DA INFORMAÇÃO**

PUBLICAÇÃO: PPGENE.DM - 667/2017

BRASÍLIA / DF: JULHO/2017

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**METODOLOGIA FORENSE DE ESTIMATIVA
SEMIAUTOMÁTICA DE ALTURA HUMANA EM VÍDEOS DE
SISTEMAS DE VIGILÂNCIA**

ADMILSON GONÇALVES JÚNIOR

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:



**HÉLIO PEDRINI, Dr., UNICAMP
(ORIENTADOR)**



**FLÁVIO ELIAS GOMES DE DEUS, Dr., ENE/UNB
(EXAMINADOR INTERNO)**



**ANA PAULA BAERNARDI DA SILVA, Dra., UCB
(EXAMINADORA EXTERNA)**

Brasília, 06 de Julho de 2017.

FICHA CATALOGRÁFICA

GONÇALVES JÚNIOR, ADMILSON.
METODOLOGIA FORENSE DE ESTIMATIVA SEMIAUTOMÁTICA DE ALTURA HUMANA EM VÍDEOS DE SISTEMAS DE VIGILÂNCIA [Distrito Federal] 2017.
xxii, 73 p., 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2017).

Dissertação de Mestrado – Universidade de Brasília, Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

1. Estimativa de Altura 2. Calibração de Câmera 3. Python 4. OpenCV 5. Biometria Branda
6. Computação Visual 7. Criminalística 8. Ciências Forenses

I. ENE/FT/UnB. II. Título (Série)

REFERÊNCIA BIBLIOGRÁFICA

GONÇALVES JÚNIOR, ADMILSON. (2017). METODOLOGIA FORENSE DE ESTIMATIVA SEMIAUTOMÁTICA DE ALTURA HUMANA EM VÍDEOS DE SISTEMAS DE VIGILÂNCIA. Dissertação de Mestrado, Publicação PPGENE.DM - 667/2017, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 73 p.

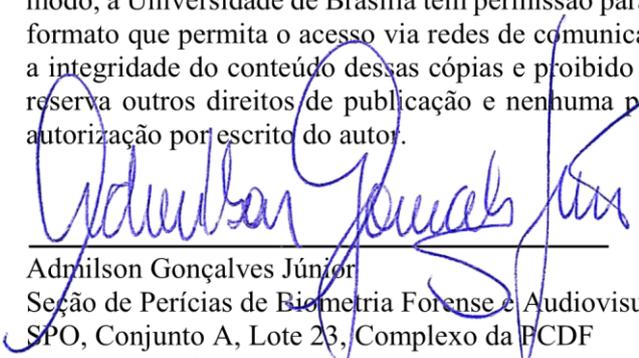
CESSÃO DE DIREITOS

NOME DO AUTOR: Admilson Gonçalves Júnior

TÍTULO DA DISSERTAÇÃO: METODOLOGIA FORENSE DE ESTIMATIVA SEMIAUTOMÁTICA DE ALTURA HUMANA EM VÍDEOS DE SISTEMAS DE VIGILÂNCIA.

GRAU/ANO: Mestre/2017.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, a Universidade de Brasília tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.


Admilson Gonçalves Júnior

Seção de Perícias de Biometria Forense e Audiovisuais/IC/DPT/PCDF
SPO, Conjunto A, Lote 23, Complexo da PCDF
Setores Complementares
CEP 70.610-907 – Brasília – DF – Brasil

Dedico a Deus este trabalho, porque capacitou-me e permitiu a sua realização.

AGRADECIMENTOS

Agradeço imensamente a minha esposa Eloisa Cláudia Vieira Gouvêa Gonçalves pelas orações, companheirismo e fé. Ela que esteve sempre presente, especialmente nos momentos mais difíceis.

Agradeço aos meus filhos Isaías, Isadora, Isabel e Isaura por terem suportado, sempre com muito amor, a ausência do papai.

Agradeço aos meus pais Admilson Gonçalves e Joana Marlene Casoni Gonçalves pelo apoio dado à distância e por sempre acreditaram no meu sucesso.

Agradeço ao meu orientador, Prof. Dr. Hélio Pedrini, por ter aceitado este desafio e pelo constante apoio, paciência, dedicação e disponibilidade, essenciais para a realização deste trabalho.

Agradeço à Diretoria Técnico-Científica da Polícia Federal pela iniciativa do projeto e ao Programa de Pós-Graduação em Engenharia Elétrica da UnB pela forma que conduziu os trabalhos.

Agradeço ao Instituto de Criminalística do Departamento de Polícia Técnica da Polícia Civil do Distrito Federal (IC/DPT/PCDF) pela permissão da participação no curso.

Agradeço à Perita Criminal Nelice Roberta da Silva Costa, Chefe da Seção de Perícias em Biometria Forense e Audiovisuais do IC/DPT/PCDF, pela confiança depositada.

Agradeço à Divisão de Apoio e Serviços Gerais (DASG/PCDF), em nome de seus diretores, os quais permitiram o uso das instalações e equipamentos desta Divisão, bem como autorizam a realização de serviços de produção dos instrumentos utilizados neste trabalho.

A todos, os meus sinceros agradecimentos.

RESUMO

METODOLOGIA FORENSE DE ESTIMATIVA SEMIAUTOMÁTICA DE ALTURA HUMANA EM VÍDEOS DE SISTEMAS DE VIGILÂNCIA

Autor: Admilson Gonçalves Júnior

Orientador: Hélio Pedrini

Programa de Pós-Graduação em Engenharia Elétrica

Brasília, julho de 2017.

Na Criminalística, as aplicações forenses de biometria são relacionadas a problemas de determinação de fonte ou de origem (confronto entre material padrão e questionado), exames estes caracterizados por se basearem em traços biométricos tradicionais únicos (DNA, lançamentos manuscritos, voz, face, etc). Identificadores biométricos possuem graus de distinção diferentes ou podem não estar suficientemente evidentes ou disponíveis, de maneira a limitar ou até mesmo inviabilizar o exame pericial. Embora os traços biométricos brandos (gênero, altura, etnia, etc) não sejam tão permanentes e confiáveis quanto os tradicionais, eles complementam as informações do indivíduo, elevando a precisão no estabelecimento de sua identidade. Este trabalho propõe uma metodologia forense de estimativa semiautomática de altura humana em vídeos de sistemas de vigilância destinada à Criminalística, cujos métodos atualmente empregados para estimar altura são restritos à projeção reversa (caracterizada pela subjetividade) e a técnicas de fotogrametria baseadas em uma única imagem (dependentes de cenas estruturadas). A metodologia proposta foi submetida a experimentos em condições reais, com câmeras de diferentes marcas, modelos e configurações. Situações adversas, como grande distância do alvo em relação às câmeras, baixa qualidade dos vídeos, pisos não planos ou irregulares, presença de ruídos nos quadros dos vídeos também foram consideradas. Os resultados de estimativas de altura dos alvos apresentaram erro médio absoluto de 0,013 m em relação aos dados de validação, o que comprova a robustez da solução para o uso forense.

ABSTRACT

FORENSIC METHODOLOGY OF SEMIAUTOMATIC ESTIMATION OF HUMAN HEIGHT IN VIDEOS OF SURVEILLANCE SYSTEMS

Author: Admilson Gonçalves Júnior

Supervisor: Hélio Pedrini

Programa de Pós-Graduação em Engenharia Elétrica

Brasília, July of 2017.

In Criminalistics, biometric forensic applications are related to problems of source or origin determination (confrontation between standard and questioned material). These tests are based on unique traditional biometric traits (DNA, handwriting, voice, face, among others). Biometric identifiers have different degrees of distinction or may not be sufficiently obvious or available, in order to limit or even render unfeasible the forensic expert report. Although soft biometric traits (gender, height, ethnicity, etc.) are not as permanent and reliable as traditional ones, they complement the individual's information, increasing the accuracy in establishing his/her identity. This work proposes a forensic methodology of semi-automatic human height estimation in videos of surveillance systems for Criminalistics, whose methods currently used to estimate height are restricted to the reverse projection (characterized by subjectivity) and single-image based photogrammetry techniques (dependent on structured scenes). The proposed methodology was submitted to experiments in real conditions, with cameras of different brands, models and configurations. Adverse situations, such as a large distance from the target to the cameras, poor video quality, uneven or irregular floors, presence of noise in the video frames were also considered. The results of height estimates of the targets presented an absolute mean error of 0.013 m in relation to the validation data, which demonstrates the robustness of the solution for the forensic use.

SUMÁRIO

1.	INTRODUÇÃO.....	1
1.1.	PROBLEMA DE PESQUISA.....	2
1.2.	MOTIVAÇÕES E JUSTIFICATIVAS.....	4
1.3.	OBJETIVOS E CONTRIBUIÇÕES.....	5
1.4.	LIMITAÇÕES E DESAFIOS.....	6
1.5.	ESTRUTURA DO TEXTO.....	6
2.	ESTUDO BIBLIOGRÁFICO – ESTADO DA ARTE.....	9
2.1.	GEOMETRIA PROJETIVA.....	9
2.2.	MODELO DE CÂMERA ESTENOPEICA	10
2.2.1.	Parâmetros Extrínsecos	12
2.2.2.	Parâmetros Intrínsecos	13
2.2.3.	Generalização do Parâmetros Intrínsecos	16
2.2.4.	Correção de Distorções	17
2.3.	CALIBRAÇÃO DE CÂMERA COM OPENCV	20
2.4.	MÉTODO DE REGRESSÃO NÃO LINEAR PARA ESTIMAR ALTURA..	22
2.5.	TRABALHOS RELACIONADOS	25
3.	METODOLOGIA	31
3.1.	AQUISIÇÃO DE DADOS.....	32
3.1.1.	Coletar Informações da Câmera de Origem e do Local de sua Instalação	32
3.1.2.	Capturar o Vídeo do Padrão Planar de Calibração	33
3.1.2.1.	Preparação do padrão planar de calibração	34
3.1.2.2.	Aquisição do vídeo do padrão planar de calibração.....	34

3.1.3.	Capturar o Vídeo do da Escala de Referência.....	35
3.2.	ANÁLISE DE DADOS	35
3.2.1.	Calibrar a Câmera de Origem.....	35
3.2.2.	Realizar as Marcações de Coordenadas da Escala de Referência	41
3.2.3.	Realizar as Marcações de Coordenadas do Alvo	42
3.2.3.1.	Ciclo de Marcha.....	42
3.2.3.2.	Escolhendo os Melhores Quadros	44
3.2.3.3.	Marcando as Coordenadas do Alvo	46
3.2.4.	Estimar a Altura do Alvo	48
4.	RESULTADOS EXPERIMENTAIS.....	49
4.1.	COLETA DE INFORMAÇÕES DA CÂMERA DE ORIGEM E DO LOCAL DE SUA INSTALAÇÃO.....	49
4.2.	CAPTURE DO VÍDEO DO PADRÃO PLANAR DE CALIBRAÇÃO	49
4.3.	CAPTURE DO VÍDEO DA ESCALA DE REFERÊNCIA.....	52
4.4.	CALIBRAÇÃO DAS CÂMERAS DE ORIGEM.....	52
4.5.	MARCAÇÕES DE COORDENADAS DA ESCALA DE REFERÊNCIA	54
4.6.	MARCAÇÕES DE COORDENADAS DOS ALVOS.....	54
4.7.	ESTIMATIVAS DE ALTURA	56
4.8.	ALTERAÇÕES NO CAMPO DE VISÃO DA CÂMERA.....	60
4.9.	DISCUSSÃO	65
5.	CONCLUSÕES E TRABALHOS FUTUROS.....	67
	REFERÊNCIAS BIBLIOGRÁFICAS	69

LISTA DE TABELAS

Tabela 4.1 – Características Externas das Câmeras e de seus respectivos Locais de Instalação.	50
Tabela 4.2 – Características Internas das Câmeras (Configurações).	51
Tabela 4.3 – Matrizes das Câmeras e Coeficientes de Distorção com Respectivos Erros de Reprojecção.	53
Tabela 4.4 – Estimativas de Altura da ESCALA01.	54
Tabela 4.5 – Estimativas de Altura dos Alvos.	57
Tabela 4.6 – Estimativas de Altura do Alvo Considerando os Cenários de Posicionamento da Câmera Cam99.	64
Tabela 4.7 – Estimativas de Altura do Alvo Considerando os Cenários de Posicionamento da Câmera Cam99 após Ajustes de Perspectiva.	64

LISTA DE FIGURAS

Figura 2.1. Câmera estenopeica (<i>pinhole</i>).....	10
Figura 2.2. Modelo da câmera estenopeica (<i>pinhole</i>).	11
Figura 2.3. Imagem com uma inclinação de ângulo θ	16
Figura 2.4. a) distorção radial positiva (barril, do inglês “ <i>barrel</i> ”; b) distorção radial negativa (almofada, do inglês “ <i>pincushion</i> ”).....	18
Figura 2.5. a) câmera com lente e sensor não paralelos entre si; b) imagem com distorção tangencial.	19
Figura 2.6. Modelo de padrão planar de calibração do tipo tabuleiro clássico.	21
Figura 2.7. Instalação típica de uma câmera de sistema de vigilância e seu respectivo sistema de coordenadas, adaptado de (Li et al., 2015).	23
Figura 3.1. Diagrama contendo as etapas da metodologia proposta.	31
Figura 3.2. Exemplo de detecção dos cantos internos do padrão de calibração.	37
Figura 3.3. Erros de reprojeção e as respectivas raízes dos erros médios quadráticos (RMSE) antes e após eliminação de pontos com erros de reprojeção atípicos. O RMSE apresentado refere-se ao retornado pela função <code>cv2.CalibrateCamera()</code>	40
Figura 3.4. Exemplo do resultado do cálculo dos parâmetros de calibração. No destaque, erros de reprojeção encontrados em cada eixo, antes e após remoção de pontos atípicos.	40
Figura 3.5. Resultado das marcações das coordenadas da escala em um vídeo.	41
Figura 3.6. Terminologia do ciclo de marcha e posições das pernas durante o ciclo, considerando a perna direita, adaptado de Whittle (2007).	43
Figura 3.7. Variação da altura durante processo de caminhada normal.	45
Figura 3.8. Variação da altura durante processo de caminhada simultânea com o uso de um <i>smartphone</i>	45
Figura 3.9. Variação da altura durante processo de corrida.	46
Figura 3.10. Resultado de marcações de coordenadas do alvo em quadros de um vídeo.	47
Figura 3.11. Exemplo de marcação de coordenadas de alvo utilizando a ferramenta de acessibilidade “lupa” dos sistemas operacionais da Microsoft (2016).	47

Figura 4.1. Escala de referência: a) projeto da escala utilizada na metodologia; b) escala pronta para o uso.	53
Figura 4.2. Exemplo de marcação de coordenadas de alvos: a) quadro da câmera Cam13; b) quadro da câmera Cam15.....	55
Figura 4.3. Histograma dos erros de estimativa de altura.....	56
Figura 4.4. Histograma dos erros de estimativa de altura dos alvos.	57
Figura 4.5. Gráfico de correlações: a) erros de estimativa de altura e coordenadas superiores do eixo x; b) erros de estimativa de altura e coordenadas inferiores do eixo x; c) erros de estimativa de altura e coordenadas superiores do eixo y; b) erros de estimativa de altura e coordenadas inferiores do eixo y.....	60
Figura 4.6. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário B, onde a escala de referência foi capturada como ESCALA02.	62
Figura 4.7. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário C, onde a escala de referência foi capturada como ESCALA03.	62
Figura 4.8. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário D, onde a escala de referência foi capturada como ESCALA04.	63
Figura 4.9. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário E, onde a escala de referência foi capturada como ESCALA05.	63
Figura 4.10. Marcações de coordenadas da escala de referência, denominada, respectivamente, de ESCALA02, ESCALA03, ESCALA04 e ESCALA05 nos vídeos dos cenários B, C, D e E, após ajustes de perspectiva em cada um deles.	64
Figura 4.11. Quadro de vídeo produzido pela câmera Cam10. No destaque, exemplo de como um alvo é visto no vídeo.	66
Figura 4.12. À esquerda, um quadro de vídeo produzido pela câmera Cam13. À direita, e um recorte ampliado do mesmo quadro. Observa-se a continuidade das faixas da escala no quadro.	66

LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIACÕES

CCD	Dispositivos de Carga Acoplada (<i>Charged Coupled Device</i>)
CFTV	Circuitos Fechados de Televisão
CMYK	Ciano Magenta Amarelo Preto (<i>Cyan Magenta Yellow Black</i>)
DLT	Transformação Linear Direta (<i>Direct Linear Transform</i>)
DNA	Ácido Desoxirribonucleico (<i>Deoxyribonucleic Acid</i>)
dpi	Pontos por Polegada (<i>Dots per Inch</i>)
EM	Maximização de Expectativas (<i>Expectation Maximization</i>)
FISWG	Grupo de Trabalho Científico de Identificação Facial (<i>Facial Identification Scientific Working Group</i>)
FLPTDM	Medida de Distância por Triangulação utilizando um Ponto de Laser Fixo (<i>Fixed Laser Point Triangular Distance Measurement</i>)
HOG-LBP	Histograma de Gradientes Orientados e Padrão Binário Local (<i>Histogram of Oriented Gradients and Local Binary Pattern</i>)
LMedS	Mediana Mínima dos Quadrados (<i>Least Median of Squares</i>)
OpenCV	Biblioteca de Computação Visual de Código Aberto (<i>Open Source Computer Vision Library</i>)
PTZ	Oscilar-Inclinar-Aproximar (<i>Pan-Tilt-Zoom</i>)
RANSAC	Consenso de Amostras Aleatórias (<i>Random Sample Consensus</i>)
RGB	Vermelho Verde Azul (<i>Red Green Blue</i>)
RMSA	Aproximação Estocástica de Robbins-Monro (<i>Robbins-Monro Stochastic Approximation</i>)
RMSE	Raiz do Erro Médio Quadrático (<i>Root Mean Square Error</i>)
SVM	Máquina de Vetores de Suporte (<i>SVM</i>)

1. INTRODUÇÃO

A biometria é uma ciência que se fundamenta nos traços (características) fisiológicos e comportamentais de um indivíduo para determinar sua identidade¹. Traços fisiológicos incluem ácido desoxirribonucleico (DNA), odor, morfologia da face e orelhas, olhos (retina e íris), geometria e padrão das veias das mãos, voz, termogramas da face ou mãos e impressões digitais, enquanto traços comportamentais incluem lançamentos manuscritos, marcha (modo de caminhar) e dinâmica de digitação (Jain et al., 2004b, 2007).

Os traços humanos, para serem utilizados como características biométricas e, assim, comporem sistemas biométricos – sistemas que empregam a biometria para identificação ou verificação (por exemplo, para controle de acesso) de indivíduos –, devem atender aos seguintes requisitos: **universalidade** (comum na população), **distinção** (duas pessoas quaisquer devem ser suficientemente diferentes considerando uma determinada característica), **permanência** (invariável durante determinado tempo), **mensurabilidade** (pode ser contável), **performático** (permitir resultados rápidos e exatos), **aceitabilidade** (uso cotidiano permitido) e **anti-evasão** (quão resistente a métodos fraudulentos) (Jain et al., 2004b).

A necessidade de estabelecimento da identidade ou verificação de indivíduos fomentou o uso da biometria em diferentes aplicações, que, segundo Jain et al. (2007), podem ser subdivididas em três grupos principais: comerciais, governamentais e forenses.

Na Criminalística, as aplicações forenses de biometria são relacionadas a problemas de determinação de fonte² ou de origem: o material questionado – vestígio (dado biométrico) – é comparado com o material padrão – coletado de um indivíduo suspeito. Em problemas desse tipo, a perícia criminal oficial é inquirida a responder se o indivíduo suspeito é a fonte do dado biométrico, considerado como marca ou traço forense (Valente, 2017).

¹ Devido à ambiguidade do termo “identidade”, Meuwly (2006) esclarece que, na ciência, a operação de “identificação” ou “classificação” refere-se à **identificação qualitativa**, ou seja, é o agrupamento de um ente (indivíduo ou objeto de interesse) numa classe de entidades com características comuns entre elas. Por outro lado, a operação de “individualização” refere-se à **identificação quantitativa**, isto é, o agrupamento de um ente em uma classe unitária com características exclusivas, que, a rigor, só pode ser estabelecida pela garantia da continuidade no tempo. É comum a operação de “individualização” ser, erroneamente, tratada como “identificação” nas ciências forenses.

² Segundo Valente (2016), tanto sistemas de identificação biométrica quanto exames de determinação de fonte são **classificadores**, promovem a **identificação qualitativa** (identificação) de um ente, não realizam a **identificação quantitativa** (individualização), que depende da continuidade no tempo, algo não garantido nem por esses sistemas nem por esses tipos de exames.

São exemplos de exames periciais de determinação de fonte fundamentados em características biométricas: exame de DNA forense, em materiais biológicos; confronto grafotécnico (ou grafoscópico), a partir de assinaturas ou outros lançamentos manuscritos; comparação forense de locutor (Valente, 2012), em traços de voz; comparação facial, a partir de imagens da face.

Os exames anteriormente citados caracterizam-se por utilizar traços biométricos únicos. Porém, além de cada um desses traços (DNA, lançamentos manuscritos, voz, face) possuir graus de distinção diferentes, como demonstrado por Jain et al. (2006), eles podem não estar suficientemente evidentes (ou disponíveis) tanto no material padrão quanto no questionado, de maneira a limitar ou até mesmo inviabilizar o exame pericial.

Em seu trabalho, Jain et al. (2004a) demonstraram que “[...] a utilização de informações auxiliares, como gênero, altura e etnia, podem melhorar o desempenho dos sistemas biométricos tradicionais”. Aliás, explicam também que, embora esses traços biométricos, denominados de “biometria branda”, não sejam tão permanentes e confiáveis quanto os tradicionais (primários), como decadatilares e face, eles complementam as informações do indivíduo, elevando a precisão no estabelecimento de sua identidade. No entanto, as características biométricas brandas só seriam úteis na identificação de um indivíduo se fossem utilizadas de forma complementar e ponderada aos identificadores biométricos primários.

Este trabalho trata da obtenção da altura humana em vídeos produzidos por câmeras de sistemas de vigilância (Circuitos Fechados de Televisão – CFTV) localizados em cenas de crime, visando ao uso dessa informação biométrica branda em exames periciais de determinação de fonte, em substituição (caso não disponíveis) ou em conjunto com informações biométricas tradicionais.

1.1. PROBLEMA DE PESQUISA

As soluções utilizadas na Criminalística para estimar altura humana em vídeos de CFTV são, em geral, restritas à projeção reversa e a técnicas de fotogrametria baseadas em uma única imagem, como: matriz fotogramétrica de câmera, metrologia de visão única e projeção de pontos.

A projeção reversa envolve uma reprodução subjetiva da cena do crime. Um gabarito métrico é posicionado no local onde o alvo foi visto na cena e, com isso, é gravado pela câmera em questão. A sobreposição do quadro recriado (quadro padrão) com o quadro original (quadro

questionado) possibilita a visualização da altura do alvo. As limitações desse método se relacionam às mudanças da cena, como, por exemplo, a retirada ou movimentação da câmera e a mudança de posição de objetos do local, impossibilitando o enquadramento original e/ou o posicionamento do gabarito métrico (HOSDB, 2007; ANP, 2016).

A matriz fotogramétrica de câmera (ou foto adaptada), técnica normalmente disponível em ferramentas de modelagem 3D, necessita que pontos de controle sejam marcados (em termos de altura, largura e profundidade, considerando um ponto de origem) numa imagem (quadro de um vídeo) para representar coordenadas do mundo. Os pontos de controle devem fornecer informações suficientes para que a ferramenta possa calcular o mapeamento do mundo 3D na imagem 2D, a partir da matriz da câmera estimada. Quando as reprojeções dos pontos de controle forem suficientemente precisas, é possível extrair informações métricas da imagem. Este método fica limitado a cenas que contenham dois conjuntos de linhas paralelas, sendo estes conjuntos ortogonais entre si, e que pelo menos uma medida do mundo seja conhecida (HOSDB, 2007; ANP, 2016; Silva Filho & Melo, 2015).

A técnica de metrologia de visão única, utilizada por Criminisi et al. (1999, 2000), baseia-se num dos principais conceitos de geometria de perspectiva, a razão cruzada, onde certas proporções do mundo real são mantidas quando mapeadas na imagem 2D. Como a razão cruzada é invariante, o valor calculado a partir da imagem é o mesmo valor que seria calculado a partir dos pontos no mundo. Para sua aplicação, é necessária uma cena geometricamente estruturada, isto é, que contenha linhas paralelas (paredes, postes, linhas segmentadas em rodovias ou ruas, móveis, marcas no piso, etc) suficientes para permitir que pontos de fuga horizontais e verticais sejam localizados. Além disso, é necessário um objeto de referência de altura conhecida na cena, localizado no mesmo plano do alvo (HOSDB, 2007; ANP, 2016).

Na técnica de projeção de pontos é possível utilizar recursos na cena para projetar virtualmente o alvo para um objeto ou uma estrutura (parede, por exemplo) de altura conhecida ou mensurável. Esta técnica possui os mesmos requisitos que a técnica de metrologia de visão única: uma cena geometricamente estruturada, que possibilite a identificação de pontos de fuga horizontais e verticais, além de uma referência de altura do mundo real (HOSDB, 2007; ANP, 2016).

Com exceção da técnica de projeção reversa, que é caracterizada pela subjetividade no posicionamento do gabarito métrico, as técnicas de fotogrametria são dependentes de cenas

geometricamente estruturadas e da permanência no local de objetos ou estruturas de referência, que nem sempre estão disponíveis.

1.2. MOTIVAÇÕES E JUSTIFICATIVAS

A utilização de circuitos fechados de televisão (CFTV) nos projetos de segurança pertencentes às mais diversas entidades (que abrangem desde grandes corporações até residências comuns), como forma de coibir, monitorar, reunir provas, bem como identificar autores de ações criminosas, tem se tornado comum nos dias de hoje. Isso tem gerado um aumento expressivo de materiais contendo vídeos extraídos de CFTV de locais de crime, os quais são encaminhados aos Institutos de Criminalística para exames periciais de determinação de fonte.

Entre os exames de determinação de fonte mais solicitados em materiais que contêm vídeos de CFTV, destaca-se o de comparação facial. Este exame consiste na análise técnica de convergências e divergências das características da face entre dois conjuntos de registros visuais – imagens padrão (indivíduo suspeito de identidade conhecida) e imagens questionadas (indivíduo registrado nos vídeos) – a fim de se verificar se as evidências suportam ou se contrapõem à hipótese de as amostras de imagens analisadas retratarem uma única pessoa.

Consoante à natureza comparativa do exame de comparação facial e à necessidade de se empregarem métodos morfológicos, recomendados pelo *Facial Identification Scientific Working Group* – FISWG (2012), requisitos mínimos de qualidade do material (vídeos) devem ser atendidos, como (Gonçalves Júnior & Costa, 2017a): **contemporaneidade** (algumas estruturas faciais sofrem alterações com o passar do tempo e o tecido cutâneo sofre colapso e desidratação com o processo de envelhecimento); **enquadramento** (preferencialmente em norma frontal e próximo da câmera); **iluminação e contraste** (de modo que a topografia cutânea e os limites das estruturas faciais sejam adequadamente visualizados); **resolução espacial** (devem possuir resolução suficiente para permitir a correta detecção das estruturas faciais, da morfologia dos elementos constitutivos da face e das marcas e sinais dermatológicos nessa região). No entanto, Arruda e Morrison (2012) destacam que grande parte dos vídeos produzidos por câmeras de sistemas de CFTV não atendem, parcial ou totalmente, a esses requisitos mínimos, limitando ou até mesmo inviabilizando o exame pericial por meio das informações biométricas da face, conduzindo a resultados inconclusivos.

Desta forma, nos casos em que a análise técnica das características biométricas tradicionais em vídeos não é suficiente para promover o convencimento do perito criminal, o uso de identificadores biométricos brandos, como a altura, incrementa o elenco de evidências analisadas e, dependendo do caso concreto, pode direcionar a exames com resultados conclusivos.

1.3. OBJETIVOS E CONTRIBUIÇÕES

Este trabalho propõe uma metodologia forense de estimativa semiautomática de altura humana em vídeos de sistemas de vigilância, destinada à Criminalística.

Adaptada de Li et al. (2015) e implementada na linguagem de programação Python³ (2017), com utilização de recursos da biblioteca de computação visual e aprendizagem de máquina OpenCV (do inglês, *Open Source Computer Vision Library*) (OpenCV, 2017a), a metodologia utiliza um modelo de regressão não linear para ajustar, em termos de mínimos quadrados, a distância focal, a altura e o ângulo de inclinação da câmera, a partir de coordenadas inferiores e superiores (base e topo) de uma escala de referência de altura conhecida. Com apenas estes três parâmetros da câmera, é possível estimar a altura humana nos quadros dos vídeos, mesmo que estes não atendam a requisitos mínimos de qualidade – como distância do objeto de interesse, baixa iluminação/contraste da cena e resolução espacial insuficiente – e sem a necessidade de cenas geometricamente estruturadas e da permanência no local de objetos ou estruturas de referência.

A solução inclui explicações detalhadas: de como realizar a coleta de informações de câmeras de origem e de seus respectivos locais de instalação, bem como orientações de como produzir os vídeos do padrão planar de calibração e da escala de referência; do processo automático para o cálculo dos parâmetros de calibração de câmeras, a partir de vídeos do padrão planar de calibração; do módulo de marcações manuais de coordenadas de alvos e de escalas de referência nos quadros dos vídeos.

Com o módulo de marcações manuais de coordenadas é possível percorrer os quadros do vídeo, escolher aqueles que melhor representam a altura estática humana e realizar diversas marcações

³ Python 3.6.0 x64, com os pacotes: lmfit (0.9.5), matplotlib (2.0.0), numpy (1.12.0+mkl), opencv-python (3.2.0), pandas (0.19.2), scipy (0.18.1).

(medições). Com isso, a altura de um alvo é obtida a partir de seleção de múltiplos quadros de um vídeo e não apenas de uma imagem estática.

São discutidos também assuntos sobre os fenômenos óticos introduzidos pelas lentes das câmeras nos vídeos e as influências do ciclo de marcha, da postura e da posição da cabeça na altura estática humana.

1.4. LIMITAÇÕES E DESAFIOS

O uso da metodologia é condicionado à completa visualização do alvo nos quadros de um vídeo e à distinção de suas extremidades (pés e cabeça) na cena. Situações em que o alvo não é totalmente visível – provocadas, por exemplo, por iluminação insuficiente do ambiente, por capturas parciais (da cintura para cima) ou por oclusões de objetos – em um vídeo, não permitem a indicação das coordenadas de seus pés e cabeça, impossibilitando o uso desta solução.

Outro fator limitante é relacionado com a movimentação e a postura do alvo no momento em que é registrado pelas câmeras. Por exemplo, alvos com posturas não eretas, em processo de corrida ou com a cabeça inclinada para baixo, geram erros consideravelmente altos nos resultados de estimativa de altura.

Foi avaliada a hipótese da utilização de métodos automáticos de detecção de pessoas nos vídeos, como, por exemplo, o proposto por Dalal e Triggs (2005). Todavia, a aplicação forense dessa metodologia exige precisão nas marcações de coordenadas (do alvo e da escala), algo que é possível de ser realizado manualmente e que não é garantido por métodos automáticos, especialmente quando são utilizados vídeos de baixa qualidade (pouca resolução espacial, iluminação/contraste e enquadramento inadequados, grande distância do objeto de interesse da câmera, etc).

1.5. ESTRUTURA DO TEXTO

A organização do restante deste trabalho é apresentada nos parágrafos a seguir.

A base teórica utilizada neste trabalho – que inclui introdução sobre geometria projetiva, apresentação do modelo de câmera estenopeica, processo de calibração de câmera com OpenCV e método de regressão não linear para estimar altura em vídeos –, bem como os trabalhos relacionados são apresentados no Capítulo 2.

No Capítulo 3, são detalhadas as fases da metodologia – fase de Aquisição de Dados (as informações da câmera de origem e do local de sua instalação são coletadas e os vídeos do padrão planar de calibração e da escala de referência são produzidos) e fase de Análise de Dados (as informações adquiridas na fase anterior são processadas com o objetivo de estimar a altura do alvo) –, bem como a base teórica utilizada.

O Capítulo 4 apresenta os resultados experimentais realizados com a aplicação da metodologia, incluindo um caso de alteração de campo de visão da câmera, bem como a análise dos resultados obtidos.

Nas seções de 4.1 a 4.7, cada fase proposta é experimentada em condições reais. Dois resultados de estimativa de altura são apresentados: um para escala, com vista a verificar e validar as marcações de referência, e outra para os alvos.

Um caso de alteração de campo de visão da câmera é estudado na seção 4.8. É apresentada uma simulação utilizando cinco diferentes cenários para verificar qual o impacto das mudanças de perspectiva da cena capturada, ocasionadas por alterações do campo de visão da câmera, nos resultados de estimativa de altura da metodologia proposta.

Na análise dos resultados dos experimentos, apresentada na seção 4.9, são discutidas situações que podem ser encontradas em casos reais, como: câmeras de diversas marcas, modelos e configurações, distância do alvo da câmera, baixa qualidade dos vídeos, pisos não planos ou irregulares, compartilhamento de parâmetros de calibração entre câmeras de mesma marca/modelo e presença de ruído nos quadros dos vídeos.

Por fim, o Capítulo 5 conclui este trabalho, apresentando direcionamentos para trabalhos futuros.

2. ESTUDO BIBLIOGRÁFICO – ESTADO DA ARTE

Nas seções a seguir é apresentada a base teórica utilizada neste trabalho, que inclui uma introdução sobre geometria projetiva, apresentação do modelo de câmera estenopeica, o processo de calibração de câmera com OpenCV e um método de regressão não linear para estimar altura em vídeos.

2.1. GEOMETRIA PROJETIVA

A geometria projetiva provê uma estrutura matemática para o imageamento e a computação gráfica tridimensionais (3D). É empregada na modelagem de processos de formação de imagens, na geração de imagens sintéticas e na reconstrução 3D de objetos a partir de múltiplas imagens (Morvan, 2009).

Conforme Morvan (2009), embora a geometria Euclidiana seja útil para modelar linhas, planos ou pontos em espaços 3D, ela possui duas desvantagens. A primeira advém da incapacidade de modelar pontos e linhas paralelas no infinito (respectivamente, pontos de fuga e linhas de fuga). A segunda é devido à operação de dimensionamento de perspectiva (necessária na projeção de pontos 3D na imagem) tornar-se não linear na divisão pelo fator de escala (que é um parâmetro da operação de dimensionamento). Todavia, contorna-se essas desvantagens da geometria Euclidiana com o uso da geometria projetiva.

Seja um ponto tridimensional qualquer. No espaço Euclidiano, esse ponto é representado por um vetor de três elementos: $(X, Y, Z)^T$, cujas coordenadas são chamadas de coordenadas não homogêneas. No espaço projetivo, esse mesmo ponto é representado por um vetor de quatro elementos: $(X_1, X_2, X_3, X_4)^T$, cujas coordenadas são chamadas de coordenadas homogêneas. Para $X_4 \neq 0$, tem-se a equação (2.1):

$$X = \frac{X_1}{X_4}; Y = \frac{X_2}{X_4}; Z = \frac{X_3}{X_4}. \quad (2.1)$$

Considerando um fator de escala homogêneo λ , a generalização do mapeamento de um ponto no espaço Euclidiano n -dimensional para um espaço projetivo $(n+1)$ -dimensional pode ser definido pela equação (2.2):

$$\underbrace{(X_1, X_2, \dots, X_n)^T}_{\text{Espaço Euclidiano}} \rightarrow \underbrace{(\lambda X_1, \lambda X_2, \dots, \lambda X_n, \lambda)^T}_{\text{Espaço Projetivo}}. \quad (2.2)$$

2.2. MODELO DE CÂMERA ESTENOPEICA

A câmera estenopeica ou *pinhole* (buraco de agulha) é caracterizada pela sua simplicidade. Ao invés de uma lente, a câmera possui em sua face anterior uma pequena abertura, por onde os raios de luz passam e projetam, na sua face posterior interna, uma imagem invertida de um objeto do espaço tridimensional (MathWorks, 2017b) (Figura 2.1).

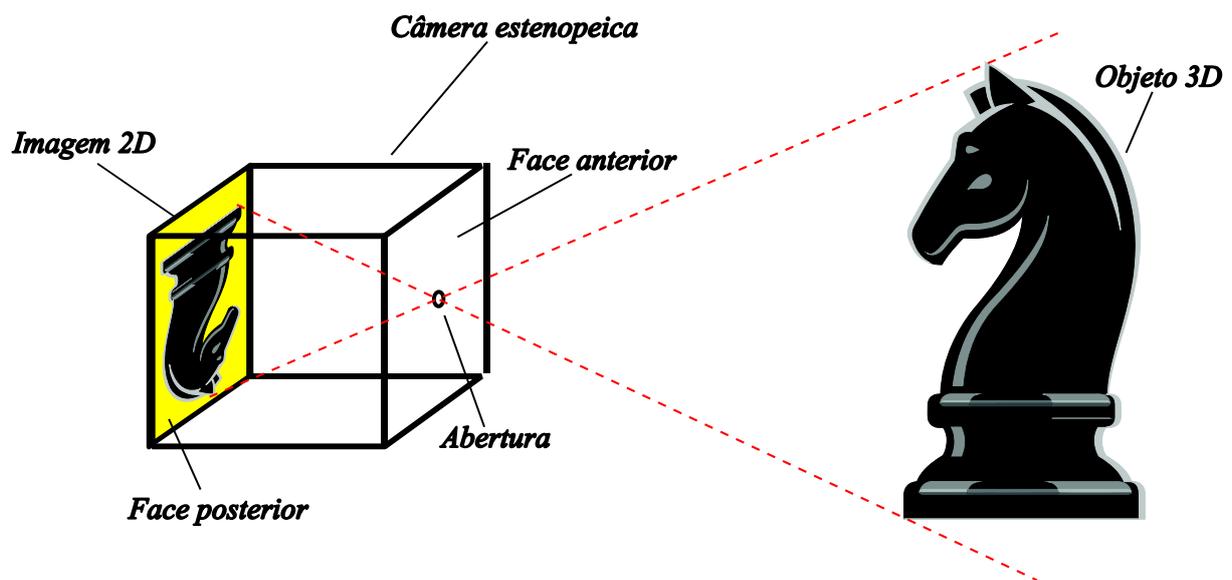


Figura 2.1. Câmera estenopeica (*pinhole*).

O modelo da câmera estenopeica é baseado no princípio da colinearidade, em que cada ponto do objeto no espaço 3D é projetado na imagem 2D por uma linha reta que passa pela abertura da câmera (Heikkila & Silven, 1997). Esse modelo define a relação geométrica – chamada de projeção perspectiva (Morvan, 2009) – entre um ponto 3D e seu pixel 2D correspondente projetado no plano da imagem, representado no sistema de coordenadas da imagem.

O processo de projeção perspectiva possui as seguintes fases (Wang et al., 2010):

- i. Um ponto B_m (3D) no sistema de coordenadas do mundo é traduzido para um ponto equivalente B_c (também 3D) no sistema de coordenadas da câmera;
- ii. O ponto B_c é projetado no modelo de câmera estenopeica em um ponto b_c (2D) no sistema de coordenadas do plano da imagem;

- iii. O ponto b_c , no sistema de coordenadas do plano da imagem, é convertido para as coordenadas do pixel b_i , também no sistema de coordenadas do plano de imagem;
- iv. São consideradas no pixel b_i características que podem ocorrer em equipamentos reais de captura, como: variações de razão de aspecto, inclinação e distorções.

A Figura 2.2 mostra o modelo de câmera estenopeica, onde (Morvan, 2009; Hartley & Zisserman, 2003): o centro da projeção perspectiva – abertura por onde todos os raios de luz se interceptam – é chamado de *centro de projeção da câmera* (C), também conhecido como *centro da câmera* ou *centro ótico*; o plano que passa pelo centro de projeção e é paralelo ao *plano da imagem*, ou *plano focal*, é o *plano principal da câmera*; a linha que passa pelo centro de projeção da câmera e incide perpendicularmente no plano da imagem é o *eixo principal*, chamado também de *eixo ótico* ou *raio principal da câmera*; o eixo principal incide no plano da imagem no *ponto principal* (c) ou *centro da imagem*; a distância entre o centro de projeção da câmera e o ponto principal é a *distância focal* (f) da câmera (normalmente expressa em milímetros).

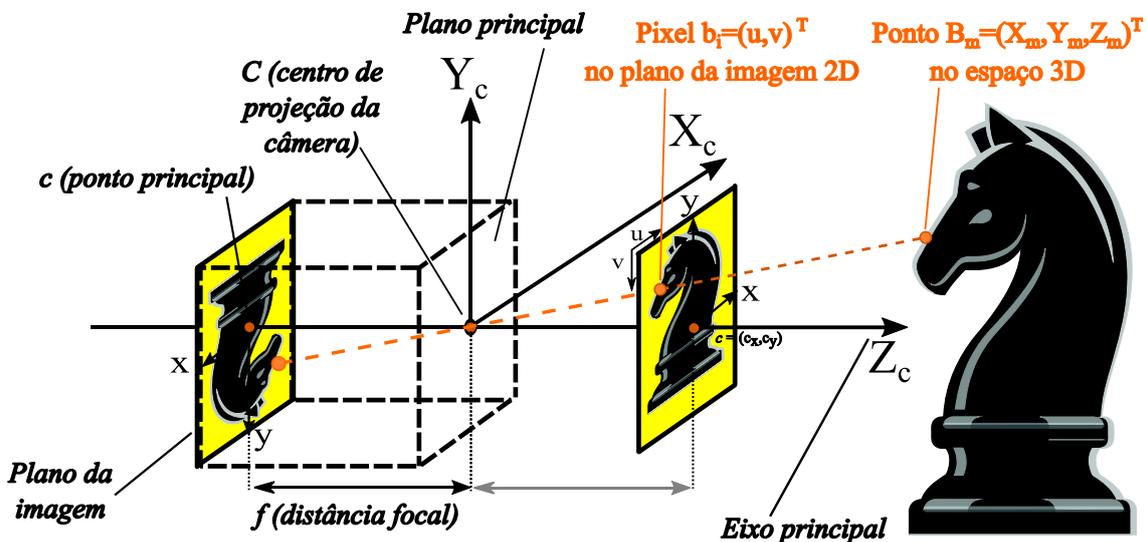


Figura 2.2. Modelo da câmera estenopeica (*pinhole*).

Para a demonstração do modelo, considera-se que a origem do sistema de coordenadas da câmera (X_c, Y_c, Z_c) é o centro de projeção (C), o qual está localizado na origem do sistema de coordenadas do mundo, em $\check{C} = (X_{m_0}, Y_{m_0}, Z_{m_0})$, e o *eixo principal* da câmera é colinear ao eixo Z_c (Morvan, 2009; Heikkila & Silven, 1997).

2.2.1. Parâmetros Extrínsecos

Os parâmetros extrínsecos descrevem o posicionamento e a orientação tridimensional da câmera em relação a um determinado sistema de coordenadas do mundo, ou seja, relacionam o sistema de coordenadas da câmera com o sistema de coordenadas do mundo.

Seja $B_m = (X_m, Y_m, Z_m)^T$ um ponto no sistema de coordenadas do mundo e $B_c = (X_c, Y_c, Z_c)^T$ esse mesmo ponto representado no sistema de coordenadas da câmera. Para transformar o ponto B_m no ponto B_c , a câmera deve ser previamente transladada para a origem das coordenadas do mundo e, em seguida, rotacionada no seu próprio sistema de coordenadas. A equação (2.3) representa essa transformação (Hartley & Zisserman, 2003):

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \begin{bmatrix} R & -R\check{C} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_m \\ Y_m \\ Z_m \end{pmatrix}, \quad (2.3)$$

em que R , de dimensão 3×3 , é a *matriz de rotação*, que representa a orientação da câmera no seu sistema de coordenadas; \check{C} é o centro de projeção da câmera representado no sistema de coordenadas no mundo; $t = -R\check{C}$ é o *vetor de translação*, de dimensão 3×1 .

Segundo Heikkila e Silvén (1997), a matriz de rotação é representada utilizando-se os ângulos de Euler ω , φ e κ , os quais definem uma sequência de três rotações elementares no sistema de coordenadas da câmera, todas de sentido horário: primeiro ao redor do eixo X_c , depois ao redor do eixo Y_c e duas vezes ao redor do eixo Z_c durante as rotações anteriores. Assim, R fica conforme a equação (2.4):

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad (2.4)$$

em que:

$$\begin{aligned} r_{11} &= \cos \varphi \cos \kappa; & r_{23} &= \cos \omega \sin \varphi \sin \kappa - \sin \omega \cos \kappa; \\ r_{12} &= \sin \omega \sin \varphi \cos \kappa - \cos \omega \sin \kappa; & r_{31} &= -\sin \varphi; \\ r_{13} &= \cos \omega \sin \varphi \cos \kappa + \sin \omega \sin \kappa; & r_{32} &= \sin \omega \cos \varphi; \\ r_{21} &= \cos \varphi \sin \kappa; & r_{33} &= \cos \omega \cos \varphi. \\ r_{22} &= \sin \omega \sin \varphi \sin \kappa + \cos \omega \cos \kappa; \end{aligned}$$

A equação (2.3) pode ser reescrita conforme a equação (2.5):

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X_m \\ Y_m \\ Z_m \end{pmatrix} + t \rightarrow B_c = RB_m + t, \quad (2.5)$$

em que R e t são os parâmetros extrínsecos da câmera estenopeica.

2.2.2. Parâmetros Intrínsecos

Para projetar o ponto B_c – ponto do espaço 3D representado no sistema de coordenadas da câmera – no modelo de câmera estenopeica e transformá-lo no ponto b_c – no sistema de coordenadas do plano da imagem – é necessário calcular os parâmetros intrínsecos da câmera.

Os parâmetros intrínsecos descrevem as características geométricas e óticas internas da câmera e são compostos, basicamente, pela distância focal, por um fator de escala e pelo ponto principal da imagem.

Utilizando-se de similaridade de triângulos, o mapeamento do ponto $B_c = (X_c, Y_c, Z_c)^T$ para o ponto $b_c = (x, y)$, pode ser escrito conforme a equação (2.6):

$$\frac{f}{Z_c} = \frac{x}{X_c} = \frac{y}{Y_c} \rightarrow x = \frac{fX_c}{Z_c}; y = \frac{fY_c}{Z_c}. \quad (2.6)$$

A partir de (2.6), tem-se a equação (2.7):

$$(x, y)^T = \left(\frac{fX_c}{Z_c}, \frac{fY_c}{Z_c} \right)^T. \quad (2.7)$$

A equação (2.7) descreve o mapeamento – conversão do espaço Euclidiano \mathbb{R}^3 para o \mathbb{R}^2 – da projeção central do mundo em coordenadas do plano da imagem (Hartley & Zisserman, 2003).

Por meio da geometria projetiva, reformula-se a equação (2.7) em coordenadas homogêneas, em que $\lambda = Z$ é o fator de escala homogêneo, obtendo-se a equação (2.8):

$$(\lambda x, \lambda y, \lambda)^T = (fX_c, fY_c, Z_c, 1)^T. \quad (2.8)$$

Representando-se a equação (2.8) em termos de multiplicação de matrizes, obtém-se a equação (2.9):

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \rightarrow b_c = PB_c, \quad (2.9)$$

em que P é a *matriz de projeção da câmera* ou, simplesmente, *matriz da câmera* (Hartley & Zisserman, 2003).

Assumiui-se até agora que a origem do sistema de coordenadas do plano da imagem é o ponto principal, o qual está representado no sistema de coordenadas da câmera. Contudo, a maioria dos sistemas de imagens define a origem do sistema de coordenadas como sendo o pixel na extremidade superior esquerda. Uma conversão de coordenadas nesse caso é necessária (Morvan, 2009; Hartley & Zisserman, 2003).

Incluindo-se as coordenadas do ponto principal e dividindo-se a distância focal pelas dimensões (p_x, p_y) do pixel – em unidade de medida do mundo, geralmente o milímetro – nas equações de (2.6), o ponto $b_c = (x, y)$, representado no sistema de coordenadas do plano de imagem, é convertido para as coordenadas do pixel $b_i = (u, v)$, neste mesmo sistema de coordenadas, conforme a equação (2.10):

$$u = \frac{\left(\frac{f}{p_x}\right) X_c}{Z_c} + c_x; \quad v = \frac{\left(\frac{f}{p_y}\right) Y_c}{Z_c} + c_y. \quad (2.10)$$

A partir da equação (2.10), reescreve-se a equação (2.9) como:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}, \quad (2.11)$$

em que (f_x, f_y) são os fatores de escala da imagem ou a distância focal da câmera, nos termos de dimensões do pixel; (c_x, c_y) são as coordenadas do ponto principal.

Numa câmera estenopeica real, os valores de (f_x, f_y) são os mesmos. Entretanto, em termos práticos, esses valores podem se diferenciar por uma série de fatores, como: falhas no sensor da câmera, falhas no pós-processamento após a aquisição da imagem, distorções da lente,

compressão de uma cena de razão de aspecto maior que a do sensor ou erros durante o processo de calibração.

É importante destacar que os valores (f_x, f_y) e (c_x, c_y) são dependentes da dimensão utilizada e precisam ser devidamente escalados quando as imagens ou os quadros dos vídeos forem redimensionados (OpenCV, 2017b).

Considerando $K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$, a matriz 3x3 da equação (2.11) pode ser expressa como $P =$

$K[I_3|0_1]$, em que I_3 é uma matriz identidade de dimensão 3x3 e 0_1 é um vetor coluna nulo. A equação (2.11) fica conforme a equação (2.12):

$$\lambda b_i = K[I_3|0_1]B_c = PB_c. \quad (2.12)$$

A matriz K representa os parâmetros intrínsecos da câmera estenopeica, conhecida como *matriz de calibração da câmera* (Hartley & Zisserman, 2003).

A unificação das equações (2.5) e (2.12) representa a fórmula geral do modelo de câmera estenopeica para que um ponto $B_m = (X_m, Y_m, Z_m)^T$ do espaço 3D, representado no sistema de coordenadas do mundo, seja mapeado para um pixel $b_i = (u, v)^T$, no sistema de coordenadas do plano de imagem:

$$\lambda b_i = K[R|t]B_m = PB_m, \quad (2.13)$$

em que a matriz da câmera P , que contém os parâmetros de calibração da câmera (matriz de calibração K , matriz de rotação R e vetor de translação t), é representada como a equação (2.14):

$$P = K[R|t]. \quad (2.14)$$

Todavia, devido à simplicidade da câmera estenopeica (não possui lentes), outras características precisam ser inseridas no modelo, a fim de torná-lo compatível com os equipamentos de aquisição de imagens e vídeos atualmente utilizados.

2.2.3. Generalização do Parâmetros Intrínsecos

O modelo de câmera estenopeica até aqui apresentado considerou as coordenadas da imagem como coordenadas euclidianas, com escalas iguais em ambos os eixos – pixels quadrados, com *razão de aspecto* de 1: 1 (Hartley & Zisserman, 2003).

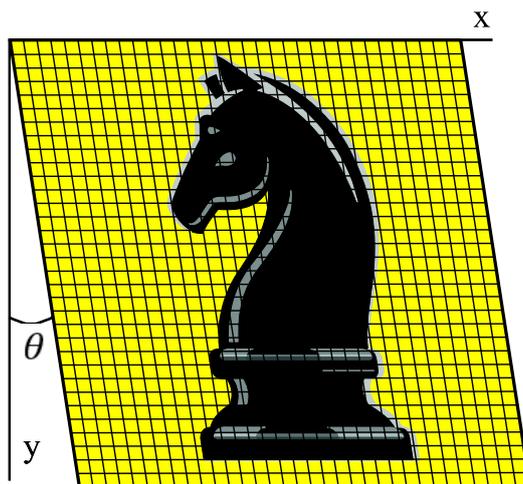


Figura 2.3. Imagem com uma inclinação de ângulo θ .

No entanto, em termos práticos, há a possibilidade de os pixels dos sensores das câmeras não serem quadrados, como pode ocorrer em câmeras com sensores de tecnologia CCD (do inglês, *Charged Coupled Device*), exigindo-se a aplicação de um fator de escala ρ em ambas as direções das coordenadas do plano da imagem. Fazendo-se isso na equação (2.10), tem-se a equação (2.15):

$$u = \rho_x \left(\frac{\left(\frac{f}{p_x}\right) X_c}{Z_c} \right); v = \rho_y \left(\frac{\left(\frac{f}{p_y}\right) Y_c}{Z_c} + c_y \right). \quad (2.15)$$

A matriz K representa, a partir da equação (2.15), a seguinte forma geral:

$$K = \begin{pmatrix} \rho_x f_x & 0 & \rho_x c_x \\ 0 & \rho_y f_y & \rho_y c_y \\ 0 & 0 & 1 \end{pmatrix}. \quad (2.16)$$

Outra característica que deve ser prevista na matriz K é a inclinação dos pixels da imagem. A inclinação pode ocorrer, por exemplo, em imagens adquiridas por placas de captura (escâneres) (Morvan, 2009) ou por falhas na fabricação dos sensores (Figura 2.3).

Incluindo-se o parâmetro γ , que descreve a inclinação entre os eixos da imagem, a matriz de calibração fica conforme a equação (2.17):

$$K = \begin{pmatrix} \rho_x f_x & \gamma & \rho_x c_x \\ 0 & \rho_y f_y & \rho_y c_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.17)$$

em que $\gamma = f_y \tan \theta$ quando os eixos da imagem não são perpendiculares entre si (MathWorks, 2017b).

Contudo, em relação à razão de aspecto e à inclinação, assume-se que os pixels de câmeras digitais recentes são quadrados ($\rho_x = \rho_y = 1$) e não possuem inclinação ($\gamma = 0$) (Morvan, 2009; Hartley & Zisserman, 2003).

2.2.4. Correção de Distorções

A inserção de distorções provocadas pelas lentes de equipamentos reais é inerente ao processo de captura de imagens e vídeos.

As distorções fazem com que linhas retas da cena real apareçam curvadas e objetos tridimensionais tenham suas dimensões afetadas quando projetados em quadros bidimensionais de um vídeo, por exemplo. Assim, antes da extração de quaisquer informações métricas, é necessário estimar os parâmetros de distorção e retificar os quadros de vídeos ou imagens.

As distorções de lentes são usualmente expressas da forma mostrada pela equação (2.18) (Heikkila & Silven, 1997; Zhang, 1998; Wang et al., 2008):

$$\begin{aligned} u_d &= u_s + \delta_{u_s}(u_s, v_s); \\ v_d &= v_s + \delta_{v_s}(u_s, v_s). \end{aligned} \quad (2.18)$$

Na equação (2.18), (u_d, v_d) são coordenadas reais da imagem, com distorções; (u_s, v_s) são coordenadas ideais⁴ da imagem, sem distorções e não observáveis; $\delta_{u_s}(u_s, v_s)$ e $\delta_{v_s}(u_s, v_s)$ são distorções em cada um dos eixos de coordenadas da imagem.

As distorções radiais, as mais comuns, ocorrem quando os raios de luz das bordas da lente se curvam mais intensamente que aqueles que passam pelo centro óptico. Essa falha de curvatura da lente pode ocorrer devido à qualidade do material utilizado (lentes de baixo custo). Em geral, quanto menor for a lente, mais perceptível será a distorção radial (MathWorks, 2017b; Hartley & Zisserman, 2003) (Figura 2.4).

Sabendo-se que o ponto principal (u_0, v_0) é centro da distorção radial, esta pode ser expressa conforme a equação (2.19) (Wang et al., 2008):

$$\begin{aligned}\delta_{u_s}(u_s, v_s)_{radial} &= u_s(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots); \\ \delta_{v_s}(u_s, v_s)_{radial} &= v_s(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots),\end{aligned}\tag{2.19}$$

em que k_1, k_2, k_3 são coeficientes de distorção radial e $r = \sqrt{(u_s - u_0)^2 + (v_s - v_0)^2}$ é a distância entre (u_s, v_s) e o ponto principal (centro da distorção radial).

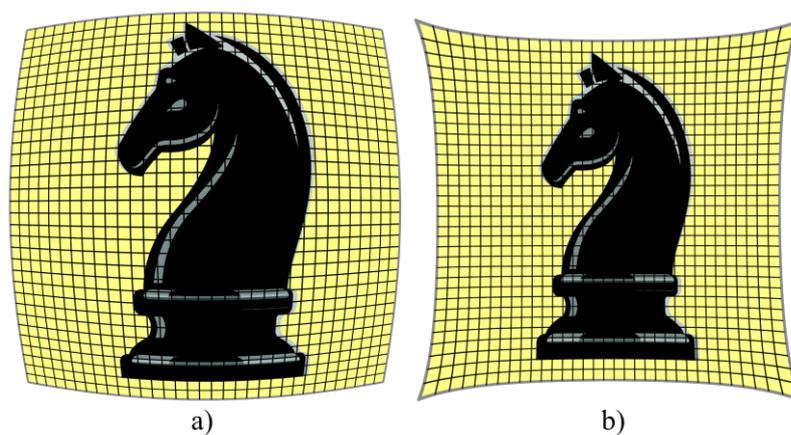


Figura 2.4. a) distorção radial positiva (barril, do inglês “*barrel*”; b) distorção radial negativa (almofada, do inglês “*pincushion*”).

⁴ Os pontos ideais são as projeções de acordo com o modelo de câmera estenopeica (Zhang, 1998).

Em geral, dois coeficientes de distorção radial são necessários para corrigir as imagens. Em distorções severas, como em lentes angulares grandes, pode-se utilizar ao menos mais um coeficiente (k_3) (MathWorks, 2017b; Heikkila & Silven 1997).

A distorção tangencial ocorre quando os centros de curvatura da superfície de lentes não são estritamente colineares ou quando a lente e o plano da imagem (sensor da câmera) não são paralelos (MathWorks, 2017b ; Heikkila & Silven 1997). (Figura 2.5).

A expressão para a distorção tangencial pode ser definida na equação (2.20) (Heikkila & Silven 1997):

$$\begin{aligned}\delta_{u_s}(u_s, v_s)_{tang} &= 2 p_1 u_s v_s + p_2 (r^2 + 2 u_s^2); \\ \delta_{v_s}(u_s, v_s)_{tang} &= p_1 (r^2 + 2 v_s^2) + 2 p_2 u_s v_s,\end{aligned}\tag{2.20}$$

em que p_1 e p_2 são os coeficientes de distorção tangencial da lente.

A partir das equações (2.19) e (2.20), apresenta-se, na equação (2.21), uma forma de tratar as distorções radial (com dois coeficientes) e tangencial:

$$\begin{aligned}u_d &= u_s + u_s(k_1 r^2 + k_2 r^4) + 2 p_1 u_s v_s + p_2 (r^2 + 2 u_s^2); \\ v_d &= v_s + v_s(k_1 r^2 + k_2 r^4) + p_1 (r^2 + 2 v_s^2) + 2 p_2 u_s v_s.\end{aligned}\tag{2.21}$$

Todavia, em calibrações não métricas de distorções de lente, uma fórmula inversa é utilizada, a qual transforma um pixel (u_d, v_d) , com distorções, em outro (u_s, v_s) , sem distorções.

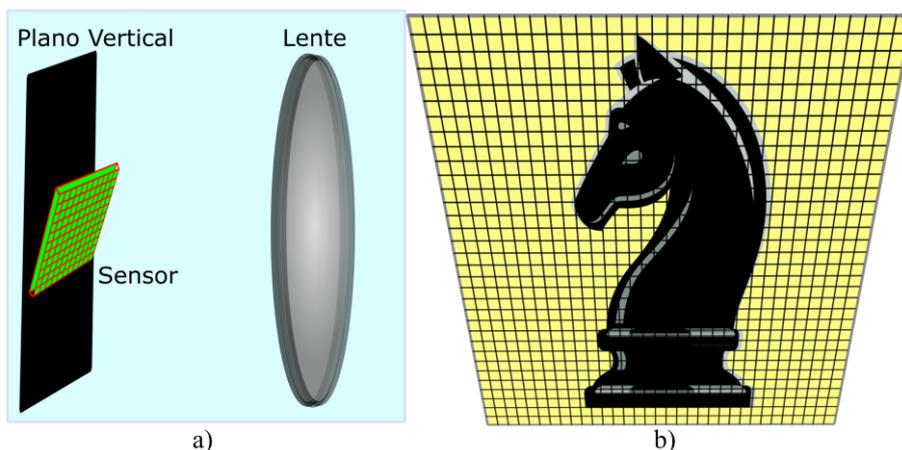


Figura 2.5. a) câmera com lente e sensor não paralelos entre si; b) imagem com distorção tangencial.

A equação (2.21) é então reescrita da forma apresentada na equação (2.22):

$$\begin{aligned}u_s &= u_d + u_d(k_1 r^2 + k_2 r^4) + 2 p_1 u_d v_d + p_2 (r^2 + 2 u_d^2); \\v_s &= v_d + v_d(k_1 r^2 + k_2 r^4) + p_1 (r^2 + 2 v_d^2) + 2 p_2 u_d v_d,\end{aligned}\tag{2.22}$$

em que $r = \sqrt{(u_d - u_0)^2 + (v_d - v_0)^2}$.

Como os coeficientes de distorção não dependem da cena capturada, são considerados como parâmetros intrínsecos. Uma vez estimados para uma determinada câmera, podem ser utilizados para retificar quaisquer imagens ou vídeos capturados por essa câmera, independentemente da resolução espacial utilizada (OpenCV, 2017b).

2.3. CALIBRAÇÃO DE CÂMERA COM OPENCV

Consoante às referências bibliográficas de sua documentação (OpenCV, 2017b), a biblioteca OpenCV baseia-se, dentre outros trabalhos, no de Zhang (1998, 2000), o qual é caracterizado pela simplicidade, baixo custo e flexibilidade em relação às técnicas clássicas de calibração.

No método de calibração de Zhang, um padrão de calibração é capturado por uma câmera em, no mínimo, duas diferentes orientações (tanto a câmera quanto o padrão de calibração podem ser movimentados, sem que as posições e os deslocamentos sejam previamente conhecidos). Os parâmetros intrínsecos e extrínsecos da câmera são estimados com método de Transformação Linear Direta (DLT, do inglês *Direct Linear Transform*), baseado no modelo de câmera estenopeica, utilizando-se dos pontos do padrão – pontos do objeto – e de seus correspondentes detectados nos quadros do vídeo – pontos de imagem. Esse processo é rápido (soluciona equações lineares), porém os resultados não são precisos, uma vez que não são previstas as componentes de distorções radiais e tangenciais da lente em seus cálculos. Esses valores iniciais de parâmetros de calibração, juntamente com as componentes de distorções, são submetidos a um processo de minimização não linear – usando Levenberg-Marquardt – das distâncias entre os pontos da imagem e seus respectivos pontos de objeto reprojatados (Wang et al., 2010).

A biblioteca OpenCV suporta atualmente três tipos de padrão de calibração, que são detectáveis automaticamente (OpenCV, 2017c): o de tabuleiro preto-branco clássico, o de círculos simétricos e o de círculos assimétricos.

Neste trabalho, é considerado como padrão planar de calibração o tabuleiro clássico (Figura 2.6). O processo de calibração que utiliza esse padrão requer como entrada ao menos dez quadros do vídeo contendo o padrão e em diferentes posições.

A detecção dos pontos do tabuleiro, cantos internos de seus quadrados, é realizada pela função `cv2.findChessboardCorners()`, a qual retorna os pontos de imagem. Como cada detecção de pontos é apenas uma aproximação das coordenadas dos cantos internos dos quadrados do tabuleiro, um refinamento das coordenadas dos pontos de imagem é feito pela função `cv2.cornerSubPix()`. Os pontos de imagem são desenhados nos quadros do vídeo pela função `cv2.drawChessboardCorners()`.

O cálculo dos parâmetros de calibração é realizado pela função `cv2.CalibrateCamera()`, que recebe, dentre outros argumentos, os pontos do objeto e os pontos de imagem. De forma geral, o cálculo segue os seguintes passos (OpenCV, 2017b): estimativa dos parâmetros intrínsecos; estimativa da posição da câmera no sistema de coordenadas do mundo, os parâmetros extrínsecos; por fim, a execução do algoritmo de otimização Levenberg-Marquardt para minimizar o erro de reprojeção dos pontos da imagem, considerando os parâmetros de calibração estimados. Como retorno, a função fornece a matriz da câmera, os coeficientes de distorção, vetor com vetores de rotação e vetor com vetores de translação, bem como a raiz do erro médio quadrático (RMSE, do inglês, *Root Mean Square Error*) de todo o processo de estimativa dos parâmetros de calibração.

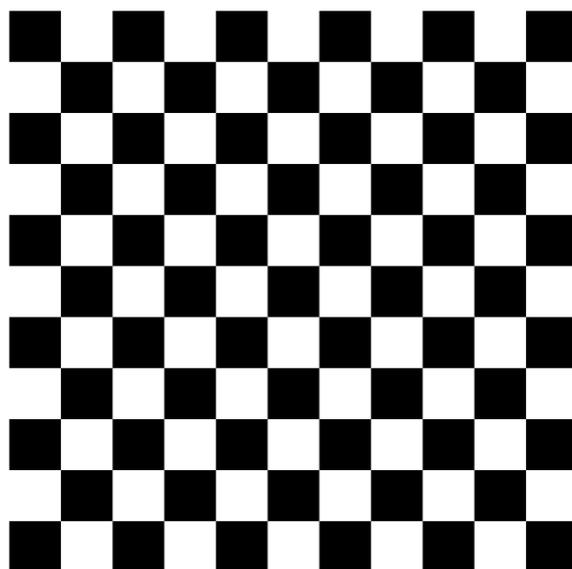


Figura 2.6. Modelo de padrão planar de calibração do tipo tabuleiro clássico.

A reprojeção dos pontos do objeto (3D) no plano de imagem (2D) utilizando-se os parâmetros de calibração encontrados, fornece uma boa estimativa de quão preciso foi o cálculo desses parâmetros. As reprojeções de pontos de objeto são feitas pela função `cv2.projectPoints()`.

A distância, medida em pixel, entre um ponto de imagem e seu respectivo ponto de objeto reprojeto no plano de imagem é conhecida como *erro de reprojeção*. É desejável que o erro de projeção seja próximo de 0, mas, como regra geral, erros menores que 1 pixel são aceitáveis (MathWorks, 2017a).

2.4. MÉTODO DE REGRESSÃO NÃO LINEAR PARA ESTIMAR ALTURA

Um método simples de calibração de câmera para estimar altura humana em vídeos produzidos por câmeras de vigilância foi proposto por Li et al. (2015). Segundo os autores, a maioria das câmeras de vigilância são instaladas em locais altos e levemente inclinadas, a fim de se obter um melhor campo de visão (Figura 2.7). Com isso, somente três parâmetros da câmera precisam ser conhecidos: a *distância focal*, o *ângulo de inclinação* e a *altura da câmera*. Esses parâmetros são estimados utilizando-se um modelo de regressão não linear – com ajustes baseados no método de mínimos quadrados – a partir das coordenadas de pontos observados da cabeça e do pé de um ser humano ou de uma escala de referência, como é o caso deste trabalho.

Com somente três parâmetros desconhecidos, o modelo de regressão não linear pode fornecer um ajuste eficiente de dados, sem a necessidade de se determinar linhas de fuga e pontos de fuga (os quais são dependentes de uma cena geometricamente estruturada, são difíceis de se estimar e são suscetíveis a erros) nos quadros de um vídeo e de se utilizar objetos de calibração (como linhas paralelas ou perpendiculares) no piso do local.

O método de Li et al. (2015) utiliza como referência a matriz da câmera P , dada pela equação (2.14), da fórmula geral do modelo da câmera estenopeica utilizada mapear pontos do espaço tridimensional no sistema de coordenadas do ponto de imagem.

Na instalação típica apresentada pela Figura 2.7, pode-se assumir que os ângulos de rotação ao longo dos eixos Y e Z e as translações ao longo dos eixos X e Z são nulos. Assim, a equação (2.14) pode ser simplificada como a equação (2.23):

$$P = KR_X[I|c_Y], \quad (2.23)$$

em que R_X é a matriz de rotação da câmera e c_Y é o vetor de translação ao longo da direção Y .

No intuito de reduzir o número de parâmetros de calibração em K , considera-se também o ângulo de inclinação de pixels nulo, a razão de aspecto unitária (1:1) e os pontos principais como $[0,0]^T$. A matriz da câmera pode ser escrita conforme a equação (2.24):

$$P = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -c \\ 0 & 0 & 1 & 0 \end{bmatrix} \rightarrow \quad (2.24)$$

$$P = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f \cos \theta & -f \sin \theta & -fc \cos \theta \\ 0 & \sin \theta & \cos \theta & -c \sin \theta \end{bmatrix},$$

em que f é a distância focal, θ é o ângulo de inclinação e c é a altura da câmera.

Desta forma, o mapeamento de um ponto $(X, Y, Z)^T$ no sistema de coordenadas do mundo para um ponto $(x, y)^T$ no sistema de coordenadas do plano da imagem é dado, em coordenadas homogêneas, pela equação (2.25):

$$\begin{bmatrix} x \\ y \\ \lambda \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} fX & 0 & 0 & 0 \\ 0 & f \cos \theta Y & -f \sin \theta Z & -fc \cos \theta \\ 0 & \sin \theta Y & \cos \theta Z & -c \sin \theta \end{bmatrix}. \quad (2.25)$$

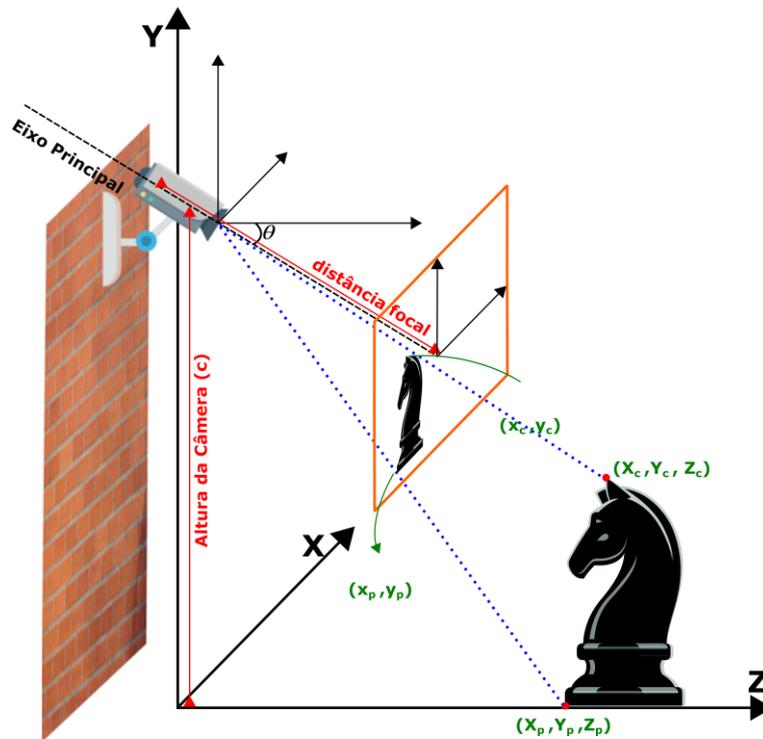


Figura 2.7. Instalação típica de uma câmera de sistema de vigilância e seu respectivo sistema de coordenadas, adaptado de (Li et al., 2015).

Representando a equação (2.25) em coordenadas cartesianas, tem-se a equação (2.26):

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{fX}{\frac{\text{sen } \theta Y + \cos \theta Z - c \text{sen } \theta}{f \cos \theta Y - f \text{sen } \theta Z - fc \cos \theta}} \\ \frac{fX}{\text{sen } \theta Y + \cos \theta Z - c \text{sen } \theta} \end{bmatrix}. \quad (2.26)$$

Uma escala de referência posicionada verticalmente em relação ao piso nas cenas de um vídeo fornece um conjunto de coordenadas de pontos de suas extremidades: superior (topo) e inferior (base). Contudo, por fornecer mais informações que a coordenada x , a coordenada y da imagem pode ser associada com a altura de um objeto de interesse e fornece uma relação básica com as coordenadas Y e Z do mundo (Li et al., 2015).

Tomando-se o termo inferior da equação (2.26), considerando $\cos \theta \neq 0$ e sabendo-se que cada par de coordenadas topo-base, denotados como y_t e y_b , pode ser indicado na imagem, obtém-se o seguinte sistema com três incógnitas, apresentado na equação (2.27):

$$\begin{cases} y_b = \frac{fY_b - f \tan \theta Z_b - fc}{\tan \theta Y_b + Z_b - c \tan \theta} \\ y_t = \frac{fY_t - f \tan \theta Z_t - fc}{\tan \theta Y_t + Z_t - c \tan \theta} \end{cases} \quad (2.27)$$

em que $Y_b = 0$ refere-se à coordenada Y inferior (base) e Y_t é a coordenada Y superior (topo) de uma escala de altura conhecida, e Z_b e Z_t são, respectivamente, coordenadas da base e do topo no eixo Z . Como é mais fácil medir Y do que Z , que necessita de objetos de calibração no piso, elimina-se Z da equação (2.27) pela substituição de Z_t , na equação inferior, por Z_b , da equação superior. A equação resultante depende dos valores de y_b e Y_t e o valor estimado da coordenada y_t , denotado como \hat{y}_t , é dado pela equação (2.28):

$$\hat{y}_t(y_b, Y_t) = \frac{f(-c \tan^2 \theta + Y_t - c)y_b - f^2 \tan \theta Y_t}{\tan \theta Y_t y_b + f(\tan^2 \theta Y_t - c \tan^2 \theta - c)}. \quad (2.28)$$

Considerando a presença de ruídos em dados reais e ε como sendo o erro produzido pelos parâmetros de calibração estimados, y_t pode ser escrito como:

$$y_t = \hat{y}_t(y_b, Y_t) + \varepsilon. \quad (2.29)$$

A equação (2.28) possui uma forma não linear e seus parâmetros podem ser encontrados por meio da seguinte regressão não linear, apresentada pela equação (2.30):

$$\begin{bmatrix} \hat{f} \\ \hat{\theta} \\ \hat{c} \end{bmatrix} = \arg \min_{f, \theta, c} \sum_{i=0}^N (\hat{y}_{t_i} - y_{t_i})^2. \quad (2.30)$$

Valores iniciais de $\hat{\theta}$ e \hat{c} podem ser mensurados no local onde a câmara está instalada e \hat{f} pode ser ajustado com valores que podem variar de 0,5 a 1,5 vezes a altura da imagem, se a unidade de comprimento do mundo estiver em centímetros.

Uma vez estimados os valores \hat{f} , $\hat{\theta}$ e \hat{c} , a altura de uma pessoa (\hat{Y}) pode ser calculada, considerando as coordenadas de seus pés e cabeça marcados nos quadros do vídeo, a partir da equação (2.31), escrita em função de y_b e y_t (Li et al., 2015):

$$\hat{Y}(y_b, y_t) = \frac{-\hat{f}\hat{c}(\tan^2 \hat{\theta} + 1)(y_b - y_t)}{\tan \hat{\theta} y_b y_t - \hat{f} y_b + \hat{f} \tan^2 \hat{\theta} y_t - \hat{f}^2 \tan \hat{\theta}}. \quad (2.31)$$

2.5. TRABALHOS RELACIONADOS

Esta seção apresenta trabalhos que tratam do problema da estimativa de altura humana a partir de registros de imagem. Grande parte das soluções aqui referenciadas foram propostas com a finalidade de uso na área forense e utilizam, como material de análise, vídeos produzidos por sistemas de vigilância.

Criminisi et al. (1999, 2000) apresentaram um algoritmo de estimativa de altura humana em vídeos baseada em geometria projetiva e visão computacional. Esta abordagem de metrologia de visão única não requer informações da geometria relativa entre a câmara e a cena vista. Para calibrar a cena, utiliza um ponto de fuga vertical, a linha de fuga do plano horizontal e a altura de um objeto de referência. O ponto de fuga vertical é estimado com a intersecção de ao menos duas linhas verticais, enquanto que dois conjuntos de linhas horizontais paralelas, com direções diferentes, são necessários para determinar a linha de fuga do plano horizontal. A altura desejada é calculada pela razão entre o alvo e o objeto de referência na cena, de altura conhecida. Os autores também tecem considerações a respeito das variações da altura durante a movimentação de um indivíduo e quanto isso pode influenciar nos resultados de estimativa.

Shao et al. (2010) propuseram uma abordagem de metrologia de vídeo baseada em movimentação de objetos da cena. A partir de informações de movimento de objetos rastreados

nos quadros de um vídeo não calibrado, as propriedades geométricas da cena são adquiridas, incluindo a linha de fuga do plano de solo e a linha de fuga vertical, com o intuito de se obter a calibração mínima da cena. Com as propriedades geométricas obtidas, um modelo probabilístico é construído – utilizando-se o algoritmo de maximização de expectativas (EM, do inglês, *expectation maximization*) – para organizar as trajetórias em grupos e simultaneamente estimar pontos de fuga pertencentes a cada grupo, bem como detectar trajetórias atípicas. Em seguida, o algoritmo de metrologia de visão simples é aplicado em cada quadro do vídeo para estimar a altura de um objeto. As medições altura em cada quadro são combinadas com a regressão da mediana mínima dos quadrados (LMedS, do inglês, *Least Median of Squares*) e a estimativa de altura final é dada pela aproximação estocástica de Robbins-Monro (RMSA, do inglês, *Robbins-Monro Stochastic Approximation*).

Outro método baseado em pontos de fuga e razão cruzada, porém com múltiplas visões, foi apresentado por Jung et al. (2016a). Os autores propuseram uma análise automática de altura humana normalizada por parâmetros de calibração de múltiplas câmeras. O método consiste em estimar pontos de fuga a partir de linhas extraídas de planos de fundo (segmento de gradiente máximo de planos de fundo subtraídos da imagem de entrada) e de pontos dos limites inferiores e superiores de pedestres presentes na cena. A estimativa de pontos de fuga é otimizada com o algoritmo RANSAC (do inglês, *Random Sample Consensus*): o ponto central da distribuição com a densidade máxima torna-se o ponto de fuga. Os parâmetros de calibração das múltiplas cenas são calculados com base nos pontos de fuga estimados. As informações de altura, extraídas também das diferentes cenas, são então normalizadas com estes parâmetros de calibração. A estimativa final de altura humana é dada pela razão cruzada com um objeto de referência na cena.

Uma extensão do algoritmo de estimativa automática de altura humana proposto por Jung et al. (2016a) é apresentada por Jung et al. (2016b) e consiste em: detectar e rastrear pessoas na cena; proceder a calibração automática da câmera; e estimar, com correção de erros a altura humana com base em objetos de referência. O método de classificação proposto na detecção de pessoas utiliza o histograma de gradientes orientados e padrão binário local (HOG-LBP, acrônimo de *Histogram of Oriented Gradients and Local Binary Pattern*) e um método de detecção humana baseado em máquina de vetores de suporte (SVM, acrônimo de *Support Vector Machine*). A precisão da estimativa da altura depende da região humana detectada. Para reduzir erros, o

método proposto acumula as alturas estimadas em cada em cada quadro do vídeo e os corrige usando o algoritmo RANSAC.

A abordagem de metrologia de visão única proposta por Criminisi et al. (1999, 2000) influenciou outros trabalhos que tratam do problema de estimativa de altura. Moon et al. (2013) propuseram um sistema inteligente de vigilância por vídeo que, além de monitorar o ambiente, é capaz de realizar autenticação de indivíduo com o uso de um cartão inteligente ou, na ausência deste, com informações de altura e de cor previamente extraídas dos vídeos das câmeras de vigilância. Moneni et al. (2012) apresentaram uma técnica manual de estimativa de altura de objetos em visão única que requer apenas informações sobre o posicionamento da câmara em relação ao mundo (por exemplo, sua altura e seu ângulo de inclinação em relação ao piso) e um ponto de fuga (sem a necessidade de linhas de fuga verticais, direções de referência e detecção de propriedades 2D e 3D das imagens). Para calcular a altura de um objeto de interesse, com extremidades superior e inferior distinguíveis na cena, é necessário obter sua altura em pixels e relação pixel/unidade de medida presente da cena, considerando a altura conhecida de um objeto 3D, no caso, a câmara. Nguyen e Hartley (2012) mostraram um método de estimativa baseado em razão cruzada, o qual utiliza pontos de referência presentes especialmente numa cena de um ambiente controlado, com o objetivo de determinar os limites de medição de altura por meio de vídeos. Lee (2012) propôs uma abordagem de calibração de câmara para estimar a altura humana formulada como um problema de otimização e baseada em metrologia de visão única. Três pontos de fuga da cena são estimados por meio de um cubo com seis pontos característicos e o algoritmo genético, com o operador de mutação *Cauchy*, é utilizado para o refinamento desta estimativa.

Porém, abordagens que utilizam a geometria projetiva dependem de imagens que possuam informações suficientes para determinar a linha de fuga do plano de referência e um ponto de fuga vertical. No entanto, nem sempre estruturas geométricas estão disponíveis na cena, impossibilitando o uso de métodos deste tipo. Alternativamente, a posição de um objeto sobre uma superfície planar no espaço 3D pode ser calculada simplesmente usando uma homografia planar, como é o caso dos trabalhos de Madden & Piccardi (2005), Jeges et al. (2008) e Lee & Choi (2010) descritos a seguir.

Madden e Piccardi (2005) introduziram um algoritmo que usa biometria baseada em “sessões de vigilância” para rastrear um indivíduo durante o período em que é capturado por câmeras pertencentes a um mesmo sistema de segurança. O método estima a altura de indivíduo pela

determinação da localização da extremidade de sua cabeça numa sequência de imagens, após a compensação de distorções da lente, e projeta essa localização no plano do solo. A altura resultante é a média das alturas calculadas sobre a série de quadros analisados.

A partir de quadros de um vídeo idealizado (com distorções corrigidas), Jeges et al. (2008) utilizaram diferentes análises geométricas no domínio de referência do mundo para estimar a altura humana. Em cada quadro de um vídeo, as diferentes formas humanas são segmentadas (com misturas de gaussianas) do plano de fundo e passam por uma compensação de sombras (a partir de informações de intensidade extraídas diretamente de valores RGB dos pixels). Dois pontos característicos, que representam o pé e a cabeça, são extraídos destas formas no domínio da imagem com distorções. A correção de distorções é feita somente nos pontos característicos citados, que são transformados em coordenadas retificadas no plano de imagem. Do plano de imagem, estes pontos são deslocados para a referência do mundo real a partir dos parâmetros extrínsecos da câmera. Duas linhas são empregadas no plano de referência do mundo para estimar a altura: a linha do pé e a linha da cabeça, que conectam, respectivamente, estes limites com o centro de projeção da câmera, passando pelos pontos correspondentes (pés e cabeça) no plano de imagem. A altura de um objeto posicionado verticalmente em relação ao plano do solo é estimada analisando-se a posição mútua da linha que passa pelo objeto vertical e pela linha da cabeça da forma. A solução depende da determinação adequada dos pontos dos pés nas imagens de vídeo, o que nem sempre acontece devido ao movimento das pernas no processo de caminhada e à compensação imperfeita de sombras no momento da detecção das formas humanas nos quadros.

Lee & Choi (2010) apresentaram uma abordagem que utiliza retroprojeção da imagem 2D em espaço 3D, baseada em visão simples, para estimar, em tempo real, altura e posicionamento humanos em movimento numa cena. No método proposto, as características bidimensionais do alvo são retroprojetadas no espaço tridimensional da cena, em que seu sistema de coordenadas é dado por um marcador retangular, disposto no plano de referência (piso). A posição e a altura do alvo são estimadas no espaço 3D e o erro geométrico associado à imprecisão do mapeamento projetivo é corrigido por meio de restrições geométricas fornecidas pelo marcador triangular. Semelhante ao trabalho de Jeges et al. (2008), os autores realizam extração de formas humanas do plano de fundo dos quadros do vídeo, porém utilizando uma técnica estatística robusta para a presença de sombras. Para garantir a presença de um eixo vertical ao plano de referência (piso), os pontos que representam os pés e a cabeça são obtidos da seguinte forma: calcula-se a

matriz de covariância da região de primeiro plano da imagem e estima-se os dois eixos principais da forma neste plano; um retângulo de contorno da forma no primeiro plano da imagem é detectado; calcula-se as interseções do eixo principal vertical e os limites verticais do forma; as duas interseções são consideradas como as posições aparentes da cabeça e dos pés, que são retroprojetadas para estimar a altura e posição do alvo.

A importância do problema de estimar altura em vídeos é evidenciada pela diversidade de soluções propostas na literatura. Técnicas distintas às apresentadas, que utilizam projeção de régua virtual, informações estatísticas com base em medidas antropométricas, montagem de grades na cena e feixe de laser, são pormenorizadas na sequência.

O trabalho de Lee et al. (2008) sugeriu um método de medição de altura que projeta uma régua virtual 3D sobre quadros extraídos de vídeos câmeras de vigilância. Para garantir que as coordenadas (X, Y) dos pés e da cabeça sejam iguais, uma régua virtual 3D, perpendicular ao plano do piso, é projetada automaticamente a partir de um ponto do pé até um ponto da cabeça, cuja coordenada é resolvida por uma equação não linear relacionada com a coordenada do ponto dos pés. Os autores comparam o método proposto com outros dois métodos fotogramétricos manuais para obter a altura humana: um que resolve equações lineares, a partir de pontos que representam o pé e cabeça de um indivíduo, com o algoritmo de transformação linear direta e sem considerar distorções de lente; e outro que resolve equações não lineares, com o método de Levenberg-Marquardt, para calcular as distorções de lente e retificar os pontos que representam os pés e cabeça do alvo. O algoritmo proposto possui limitações para estimar a altura de pessoas em movimento, devido às variações da altura durante o processo de caminhada.

BenAbdelkader & Yacoob (2008) apresentaram um método estatístico para estimar altura humana em vídeos não calibrados. A solução é direcionada para os casos onde as imagens não possuem informações geométricas de referência e apenas a parte superior do corpo é vista nas cenas. Os autores resolvem o problema da ausência de medidas de referência na cena por meio de relações antropométricas extraídas da imagem e estatísticas antropométricas e associam estas relações/estatísticas com técnicas de metrologia para estimar a altura desejada.

Uma abordagem utilizando a técnica de grades para estimar a estatura humana foi proposta por De Donno et al. (2012). Todavia, para ser considerada útil para fins forenses, o indivíduo deve ser filmado em posição estática e ereta sob uma estrutura de altura conhecida – uma porta, por

exemplo – e a cena deve ser geometricamente estruturada. A técnica baseia-se em criar uma grade a partir de quatro linhas, duas passando pelos limites inferior e superior da estrutura e duas nos limites laterais. Os pontos de interseção das linhas horizontais e verticais criam vértices, cuja união forma os lados da grade onde o sujeito está posicionado. O prolongamento das linhas que passam pelos limites superior e inferior da estrutura formam um ponto de fuga horizontal (F). O ponto de interseção (M) de duas linhas diagonais traçadas a partir dos vértices da grade coincide com o seu centro, sendo M equidistante aos planos do piso e do limite superior da estrutura. Uma linha traçada de M até F divide a grade em duas metades de mesma dimensão. O algoritmo permite divisões subsequentes da grade e a altura de um indivíduo pode ser obtida quando a linha tangencial da subdivisão da grade coincidir com o contorno da cabeça.

Uma solução mais elaborada para estimar altura humana e de restrita aplicação forense foi proposta por Wang & Chen (2015), a qual utiliza técnicas de processamento de imagem e princípios de medidas triangulares. Os autores desenvolveram um método de medida de distância por triangulação utilizando um ponto de laser fixo (FLPTDM, do inglês, *Fixed Laser Point Triangular Distance Measurement*) para construir um esquema de medição de altura humana sem contato. O sistema é posicionado a uma altura determinada e ortogonal ao piso. Um feixe de laser é utilizado para emissão de sinal e uma câmera digital é usada na detecção. Com isso, a distância entre a imagem de projeção do feixe de laser e o centro da imagem pode ser obtida por relações trigonométricas. Parâmetros de calibração da câmera são utilizados para aumentar a precisão das medidas. Contudo, as precisões das medidas variam de acordo com a distância entre o feixe de laser e câmera, sendo esta uma desvantagem do método.

3. METODOLOGIA

A metodologia para estimativa semiautomática de altura humana em vídeos de sistemas de vigilância (CFTV) possui duas fases principais: uma de Aquisição de Dados e outra de Análise de Dados.

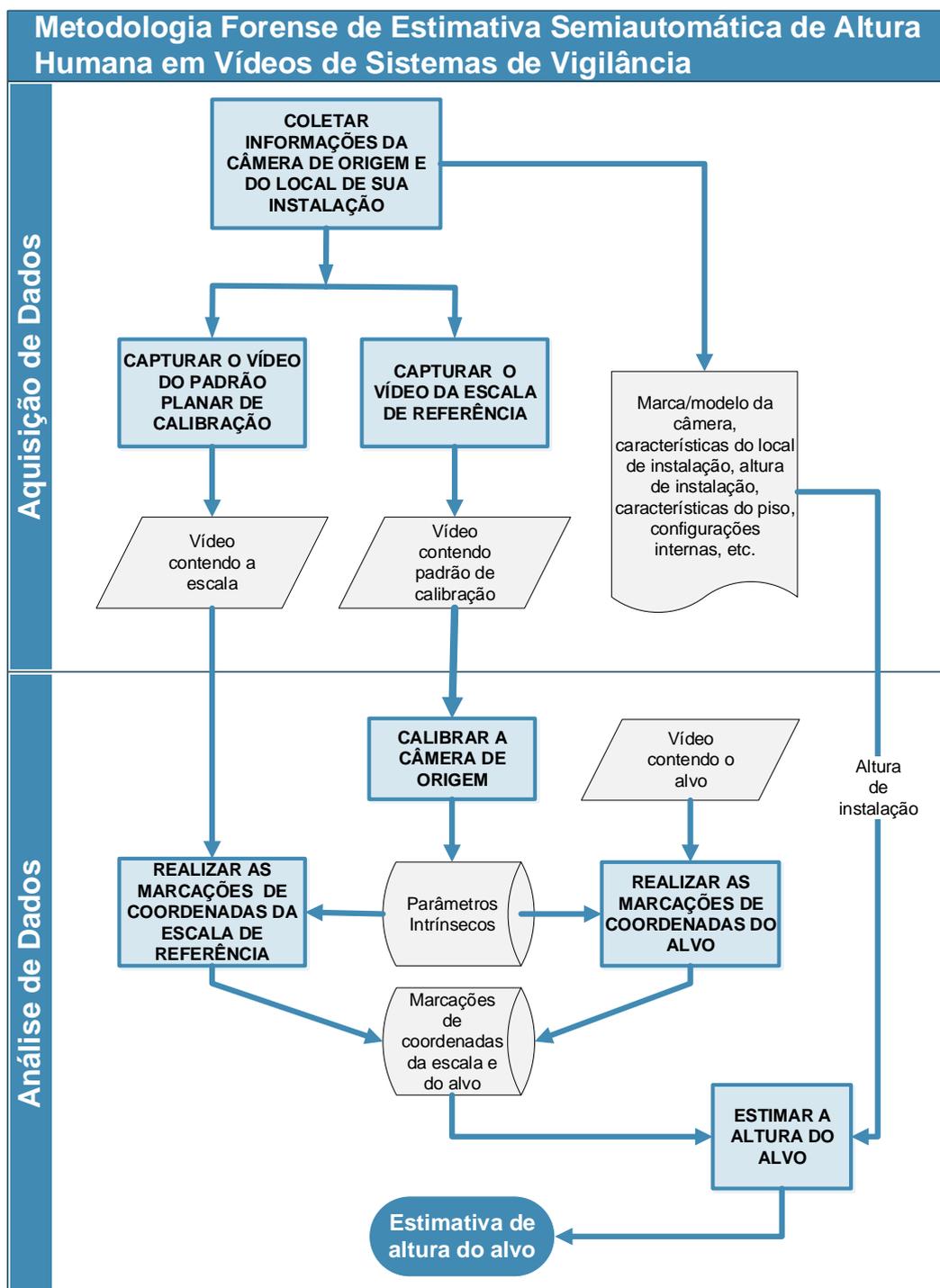


Figura 3.1. Diagrama contendo as etapas da metodologia proposta.

A fase de Aquisição de Dados é composta pela coleta de informações externas e internas da câmera que capturou o vídeo que contém o alvo – denominada câmera de origem – e do local de sua instalação, e; pela produção dos vídeos do padrão planar de calibração e da escala de referência.

Na fase de Análise de Dados, são processadas as informações adquiridas na fase de Aquisição de Dados. Os parâmetros de calibração da câmera de origem são calculados e são utilizados na retificação dos quadros dos vídeos da escala e do alvo, para que suas coordenadas (inferior e superior) sejam marcadas. Obtém-se, a partir das marcações de coordenadas desta fase, uma estimativa de altura do alvo (Figura 3.1).

3.1. AQUISIÇÃO DE DADOS

Na fase de Aquisição de Dados, as informações da câmera de origem e do local de sua instalação são coletadas e os vídeos do padrão planar de calibração e da escala de referência são produzidos.

Nas subseções seguintes, são apresentadas orientações dessas subfases, cujas informações produzidas são entradas para fase de Análise de Dados.

3.1.1. Coletar Informações da Câmera de Origem e do Local de sua Instalação

Nesta subfase, as características externas e internas da câmera de origem e do local de sua instalação são anotadas. Essas informações são úteis na definição da estratégia de captura dos vídeos do padrão de calibração e da escala de referência, bem como para a realização da fase de Análise de Dados.

São exemplos de informações que podem ser adquiridas nesta subfase:

- **Marca/modelo da câmera:** câmeras de mesma marca/modelo podem, em geral, compartilhar os mesmos parâmetros de calibração, desde que estejam utilizando a mesma configuração de resolução para armazenar os vídeos. Em outras palavras, num local em que são utilizadas câmeras de mesma marca/modelo e configuração de resolução de vídeo, bastaria calibrar uma das câmeras e utilizar os parâmetros obtidos para as demais, facilitando os trabalhos;
- **Altura da câmera:** câmeras instaladas em lugares muito altos requerem o uso de escadas ou de outros meios de aproximação do padrão de calibração. Além disso, podem

exigir o uso de uma escala de referência de maiores dimensões, que possibilitem sua completa visualização. Ademais, altura da câmera é uma informação que pode ser utilizada no processo de estimativa de altura, conforme é detalhado na seção 2.4.

- **Características da região onde o alvo foi visto:** as características do piso – se plano, irregular ou em aclave/declive – e a distância entre essa região e a câmera de origem, demandam critérios específicos tanto no posicionamento da escala, durante a gravação de seu vídeo (para reproduzir as posições onde o alvo foi visto), quanto no momento da realização de suas marcações (para possibilitar uma densidade de pontos de referência suficientes nos ajustes da estimativa de altura);
- **Características do local de instalação da câmera:** o tipo de ambiente (se interno ou externo) e suas características de iluminação (incidência de luz solar, presença de sombra, iluminação artificial ou infravermelho, etc) podem dificultar a visualização do padrão de calibração e/ou da escala de referência no momento da produção de seus respectivos vídeos exigindo-se, assim, a adoção de soluções de contorno para as dificuldades porventura encontradas;
- **Características internas da câmera:** anotar informações de configurações, como: resolução utilizada para armazenar os vídeos, opções de detecção de movimento, foco automático e quaisquer outras que possam influenciar nos resultados.

3.1.2. Capturar o Vídeo do Padrão Planar de Calibração

O vídeo contendo o padrão planar de calibração é utilizado como entrada para determinar os parâmetros de calibração da câmera, mais especificamente, os seus parâmetros intrínsecos.

Conforme explica a subseção 2.2.2, os parâmetros intrínsecos são as características óticas e geométricas internas da câmera. Em muitos casos, a precisão do processo de calibração da câmera influencia fortemente o desempenho global de sistemas de visão computacional (Heikkila & Silven, 1997). Deste modo, para que seja introduzido o mínimo de erros nesse processo, são necessários alguns cuidados especiais com a preparação do padrão planar de calibração e com a sua posterior aquisição pela câmera de origem.

3.1.2.1.Preparação do padrão planar de calibração

- O suporte do padrão de calibração deve ser feito com material de superfície plana e resistente a deformações (umidade do ambiente) e a dilatações (temperaturas excessivas), visando garantir a manutenção de suas dimensões (Callari, 2010);
- O padrão planar deve ser impresso nas cores branca e preta. Serviços profissionais de impressão oferecem, além de impressões⁵ de alta qualidade, tipos variados de materiais (papéis com diferentes gramaturas e vinil com adesivo, por exemplo) (Callari, 2010);
- Antes de ser fixado, o impresso do padrão deve estar bem alinhado às extremidades do suporte. Uma fixação cuidadosa previne a ocorrência de bolhas, amassados ou dobras (Callari, 2010).

3.1.2.2.Aquisição do vídeo do padrão planar de calibração

- É necessário que a câmera de origem esteja configurada com a mesma resolução à época que capturou o alvo (requisito de parte dos parâmetros intrínsecos, conforme explica a subseção 2.2.2);
- Caso a câmera de origem possua o recurso de foco automático, este deve ser desabilitado. Com esse recurso ativado, as peças que compõem a lente da câmera movem-se fisicamente e isso influencia na determinação dos parâmetros intrínsecos do equipamento (Callari, 2010);
- O suporte deve ser posicionado em frente à câmera de origem a uma distância mínima que garanta a visualização das bordas e dos detalhes do padrão de calibração (Callari, 2010);
- É preciso movimentar o suporte em frente à câmera, a fim de que o padrão de calibração seja capturado pelas extremidades da lente, onde as distorções são mais evidentes (a subseção 2.2.4 detalha as distorções de lentes). Além disso, o suporte deve ser rotacionado e inclinado, não bruscamente, e mantido estável por um determinado instante antes de ter sua posição alterada.

⁵ Vale lembrar que impressões em grandes formatos normalmente utilizam sistema de cores CMYK. Caso opte em fazer a arte gráfica do padrão planar, utilize o CMYK como padrão de cores, caso contrário, a tonalidade da cor preta da impressão poderá ficar diferente daquela vista na tela do computador.

3.1.3. Capturar o Vídeo do da Escala de Referência

O vídeo contendo a escala de referência é utilizado como entrada para marcar as coordenadas – inferior e superior – dessa escala. Uma vez marcadas as coordenadas da escala, estas são utilizadas na estimativa de altura do alvo (conforme mostra a seção 2.4).

O processo de aquisição do vídeo da escala de referência também requer cuidados especiais para que não sejam introduzidos erros na metodologia.

Assim como na captura do padrão de calibração, é requerido que a configuração da resolução da câmera de origem seja a mesma à época que capturou o alvo. Além de estar instalada no mesmo local – preferencialmente, na mesma altura e com o mesmo campo de visão.

A escala de referência, de altura conhecida, deve ser posicionada, perpendicularmente, no piso e deslocada pela região, e suas adjacências, da cena onde o alvo foi visto. É essencial que as extremidades superior e inferior da escala sejam visíveis nos quadros do vídeo, caso contrário, não será possível realizar as marcações de suas coordenadas.

3.2. ANÁLISE DE DADOS

Na fase de Análise de Dados, as informações adquiridas na fase anterior são processadas com o objetivo de estimar a altura do alvo.

Nesta fase, são calculados os parâmetros de calibração da câmera, com vistas a retificar os vídeos do alvo e da escala; são marcadas as coordenadas das extremidades inferior e superior, tanto da escala (base e topo) quanto do alvo (pés e cabeça) em seus respectivos vídeos já retificados; são estimados, por meio de uma regressão não linear e utilizando-se os registros de marcações da escala, a distância focal, a altura e o ângulo de inclinação da câmera de origem. Com esses três parâmetros da câmera e com as marcações de coordenadas do alvo, é possível estimar a altura desejada (alvo).

Cada subfase da Análise de Dados é a seguir apresentada.

3.2.1. Calibrar a Câmera de Origem

Uma câmera, formada por um plano da imagem (sensor) e uma lente, é capaz de transformar um objeto do espaço tridimensional (3D) em uma imagem bidimensional (2D). Porém, as distorções existentes entre os pontos desse objeto do espaço e suas respectivas localizações na

imagem (pixel), impedem que essa transformação seja perfeitamente traduzida por meio de transformações de perspectivas. Por se tratar de aproximações do mundo real, a modelagem das distorções depende da eficácia do modelo utilizado e quão bem estimados são seus parâmetros. A modelagem de distorções está relacionada com o processo de calibração de câmeras, o qual determina as suas características geométricas e óticas internas (parâmetros intrínsecos) – e/ou sua posição e orientação tridimensional em relação a um determinado sistema de coordenadas do mundo (parâmetros extrínsecos) (Clarke & Fryer, 1998).

Segundo Zhang (1998), a calibração de câmera é um passo necessário na visão computacional 3D, pois visa extrair informações métricas de imagens 2D. Como os vídeos produzidos por sistemas de vigilância, em geral, possuem significativa distorção em seus quadros (Jeges et al., 2008), é notória a necessidade prévia de se calcular os parâmetros de calibração geométrica das câmeras que os originaram, compensar os efeitos da distorção e, posteriormente, extrair quaisquer informações métricas desses vídeos.

O módulo de calibração de câmera proposto por este trabalho é conciso: um vídeo contendo o padrão planar de calibração é fornecido como entrada e os parâmetros intrínsecos e extrínsecos otimizados da câmera são dados como saída. Não há necessidade de intervenções manuais no processo, como ocorre na solução apresentada por Bouguet (2004), que requer a indicação manual dos cantos mais extremos do padrão de calibração tabuleiro de xadrez nas imagens de entrada.

Desenvolvido na linguagem de programação Python, o módulo de calibração proposto utiliza os recursos da biblioteca de *software* OpenCV, cujas funções de calibração de câmera fundamentam-se no modelo de câmera estenopeica (*pinhole*) e introduz nesse modelo as distorções radiais e tangenciais de lentes (Wang et al., 2010).

O módulo do processo de calibração de câmera possui as seguintes etapas:

A. *Detecção do padrão planar de calibração nos vídeos*

Este procedimento tem como objetivo a detecção do padrão planar de calibração – os pontos do objeto – nos quadros dos vídeos – pontos de imagem – para a extração suas respectivas de coordenadas. Os pontos de objeto e os pontos de imagem que são empregados na estimativa dos parâmetros intrínsecos e extrínsecos da câmera.

As seguintes funções da biblioteca OpenCV, pormenorizadas na seção 2.3, são utilizadas: *cv2.findChessboardCorners()* – com as opções *cv2.CALIB_CB_NORMALIZE_IMAGE*, *cv2.CALIB_CB_ADAPTIVE_THRESH*, *cv2.CALIB_CB_FAST_CHECK*, *cv2.CALIB_CB_FILTER_QUADS* habilitadas – para a detecção das coordenadas dos pontos de imagem; *cv2.cornerSubPix()* para refinar estas coordenadas; *cv2.drawChessboardCorners()* para desenhar os pontos de imagens detectados, a partir de suas coordenadas, nos quadros do vídeo, conforme ilustra a Figura 3.2.



Figura 3.2. Exemplo de detecção dos cantos internos do padrão de calibração.

É importante destacar que um conjunto de pontos de imagem similares resultam em equações similares e isso pode gerar erros no processo de calibração. Desta forma, os pontos do objeto – pontos do padrão de calibração representados no seu próprio espaço de coordenadas – e seus respectivos pontos de imagens são selecionados conforme critério mostrado no **Algoritmo 1**, que considera um limite máximo, informado, de pontos detectados.

Algoritmo 1: Seleção de intervalo de pontos detectados

entrada: [objpts]: pontos do objeto
[imgpts]: pontos de imagem
limite: valor limite de pontos
saída: [n_objpts]: pontos selecionados de objeto
[n_imgpts]: pontos selecionados de imagem

```
1 passo ← inteiro(tamanho([objpts]) / limite)
2 para i no intervalo(0, tamanho([objpts]), passo) faça:
3     insere [objpts]{i} em [n_objpts]
4     insere [imgpts]{i} em [n_imgpts]
5 fim
```

B. Cálculo dos parâmetros de calibração da câmera

O cálculo dos parâmetros de calibração é realizado pela função `cv2.CalibrateCamera()`, que recebe os pontos do objeto e os pontos de imagem obtidos, segundo explicações da alínea anterior, e retorna os parâmetros de calibração da câmera (matriz da câmera, coeficientes de distorção, vetores de rotação e vetores de translação) e a raiz do erro médio quadrático do processo.

C. Eliminação de pontos com erros de reprojeção atípicos

O limite máximo de erros de reprojeção neste trabalho é fixado em 0,5 pixel, embora quaisquer valores menores que 1 pixel são aceitáveis, conforme é visto na seção 2.3.

A eliminação de pontos com erros de reprojeção atípicos é feita conforme o **Algoritmo 2**.

Algoritmo 2: Eliminação de pontos com erros de reprojeção atípicos

entrada: [objpts]: pontos do objeto
[imgpts]: pontos de imagem
[std_eros]{x,y}: desvio padrão de erros de reprojeção em cada um dos eixos x e y.
saída: [n_objpts]: pontos típicos de objeto
[n_imgpts]: pontos típicos de imagem

```
1 percentual ← 1
2 limite ← 0,5
3 qtdtotpts ← tamanho([objpts])
4 qtdminpts ← ((qtdtotpts) - resto(qtdtotpts / 5) / 5)
```

```

5 enquanto percentual ≤ 100 faça:
6   típico_x ← mínimo([std_erro] {x}) * (1+percentual/100):
7   típico_y ← mínimo([std_erro] {y}) * (1+percentual/100)
8   limite_x ← limite se típico_x > limite, senão típico_x
9   limite_y ← limite se típico_y > limite, senão típico_y
10  para i no intervalo(0, qtdtotpts) faça:
11    se [std_erro] {x} ≤ limite_x e [std_erro] {y} ≤ limite_y então:
12      insere [objpts] {i} em [n_objpts]
13      insere [imgpts] {i} em [n_imgpts]
14    fim
15  fim
16  se tamanho([n_objpts]) < qtdminpts, então
17    percentual ← percentual + 1
18  senão:
19    pausa
20 fim

```

D. Recálculo de parâmetros de calibração

Após a eliminação de pontos com erros de reprojeção atípicos, um novo cálculo de parâmetros de calibração é executado, com o objetivo de obter maior acurácia de seus valores.

A Figura 3.3 mostra um gráfico dos erros de reprojeção de pontos. À esquerda na figura, são os erros de reprojeção obtidos no primeiro cálculo dos parâmetros de calibração. À direita, são erros de reprojeção do segundo cálculo de parâmetros de calibração, após a eliminação dos pontos cujos erros foram considerados atípicos no primeiro cálculo.

A Figura 3.4 mostra em detalhes os parâmetros de calibração encontrados. Os valores do erro de reprojeção são desvios padrões dos erros de reprojeção em cada um dos eixos x e y (Bouguet, 2004).

E. Gravação dos parâmetros de calibração

Os parâmetros de calibração calculados são armazenados para serem utilizados na retificação dos vídeos nas etapas de marcação de coordenadas da escala de referência e do alvo, descritas a seguir.

Erros de reprojeção (em pixels)

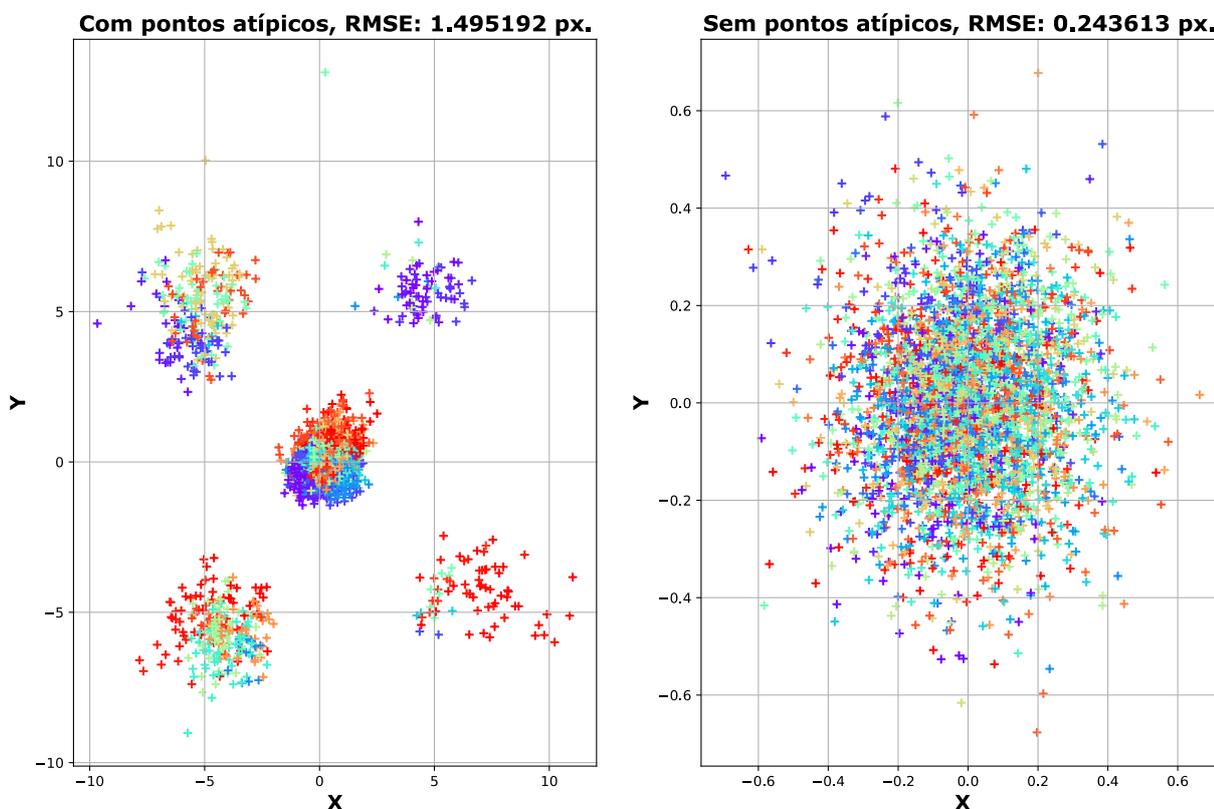


Figura 3.3. Erros de reprojeção e as respectivas raízes dos erros médios quadráticos (RMSE) antes e após eliminação de pontos com erros de reprojeção atípicos. O RMSE apresentado refere-se ao retornado pela função `cv2.CalibrateCamera()`.

```

Resultado da calibração sobre 178 padrões encontrados.:
Distância Focal.....: f = [ 1093.544495 1055.494821 ]
Ponto principal.....: c = [ 673.839380 331.624534 ]
Inclinação.....: gamma = [ 0.000000 ]
Coeficientes de distorção: k = [ -0.404280 0.119363 0.001188 -0.004354 0.037585 ]
Erro de reprojeção.....: [ 1.045681 1.068714 ]

Resultado da calibração da câmera sobre 35 padrões encontrados, após remoção de pontos atípicos:
Distância Focal.....: f = [ 1105.658246 1066.430420 ]
Ponto principal.....: c = [ 672.925508 329.289903 ]
Inclinação.....: gamma = [ 0.000000 ]
Coeficientes de distorção: k = [ -0.400574 0.092124 0.000637 -0.002353 0.166008 ]
Erro de reprojeção.....: [ 0.176400 0.168019 ]
    
```

Figura 3.4. Exemplo do resultado do cálculo dos parâmetros de calibração. No destaque, erros de reprojeção encontrados em cada eixo, antes e após remoção de pontos atípicos.

3.2.2. Realizar as Marcações de Coordenadas da Escala de Referência

Nesta etapa, as coordenadas inferior e superior (base e topo) da escala de referência são marcadas no vídeo que a contém. As coordenadas da escala são utilizadas num modelo de regressão não linear – com ajustes baseados no método de mínimos quadrados – para estimar a distância focal, altura e ângulo de inclinação da câmera, conforme é explicado na seção 2.4.

As coordenadas da escala de referência, de altura conhecida, são marcadas, de forma manual, diretamente nos quadros do vídeo. É necessário que sejam escolhidos os quadros nos quais há total visibilidade da escala (base e topo). A Figura 3.5 mostra o resultado das marcações de escala nos quadros de um vídeo.

As coordenadas indicadas nas marcações são fornecidas ao processo de regressão não linear para estimar parâmetros da câmera, que são parte dos requisitos para estimar a altura do alvo. Desta forma, quanto maior a densidade de pontos marcados para a escala, maior a probabilidade de a escala reproduzir as movimentações do alvo e mais precisos tendem a ser os cálculos de estimativa de altura.



Figura 3.5. Resultado das marcações das coordenadas da escala em um vídeo.

3.2.3. Realizar as Marcações de Coordenadas do Alvo

As marcações de coordenadas do alvo são também feitas de forma manual. Todavia, é necessário eleger os quadros do vídeo que representam a altura estática de um indivíduo, de modo a reduzir a introdução de erros nesse processo e aumentar a confiabilidade dos resultados de estimativa de altura.

Em consonância com o trabalho de Criminisi et al. (1999), durante o ciclo de marcha, a altura humana não corresponde necessariamente à sua altura estática, podendo variar em torno de 6 cm. Contudo, Ramstrand et al. (2011) relatam que, em pesquisas biomecânicas sobre o ciclo de marcha humana, essas variações de altura são ligeiramente menores durante o deslocamento vertical do corpo.

Uma breve introdução sobre ciclo de marcha e a simulação três situações cotidianas de movimentação que podem influenciar na variação de altura humana são apresentadas nas subseções a seguir.

Um estudo mais detalhado sobre a influência da postura e o deslocamento do corpo humano na estimativa de altura em vídeos foi descrito por Ramstrand et al. (2011).

3.2.3.1. Ciclo de Marcha

Define-se marcha como a maneira e o estilo de caminhada de uma pessoa. Assim, o ciclo de marcha é o intervalo de tempo entre duas ocorrências sucessivas de um dos eventos repetitivos de caminhada (Whittle, 2007).

O ciclo de marcha possui sete eventos – contato inicial, dedo oposto fora, calcanhar elevado, contato inicial oposto, dedo fora, pés adjacentes e verticalização da tíbia –, os quais dividem o ciclo em sete períodos: quatro agrupados na fase de apoio, quando o pé está em contato com o piso e sustenta o peso do corpo (60% do ciclo), e; três na fase de balanço, quando o pé está elevado e movimentando-se para frente (40% do ciclo) (Figura 3.6).

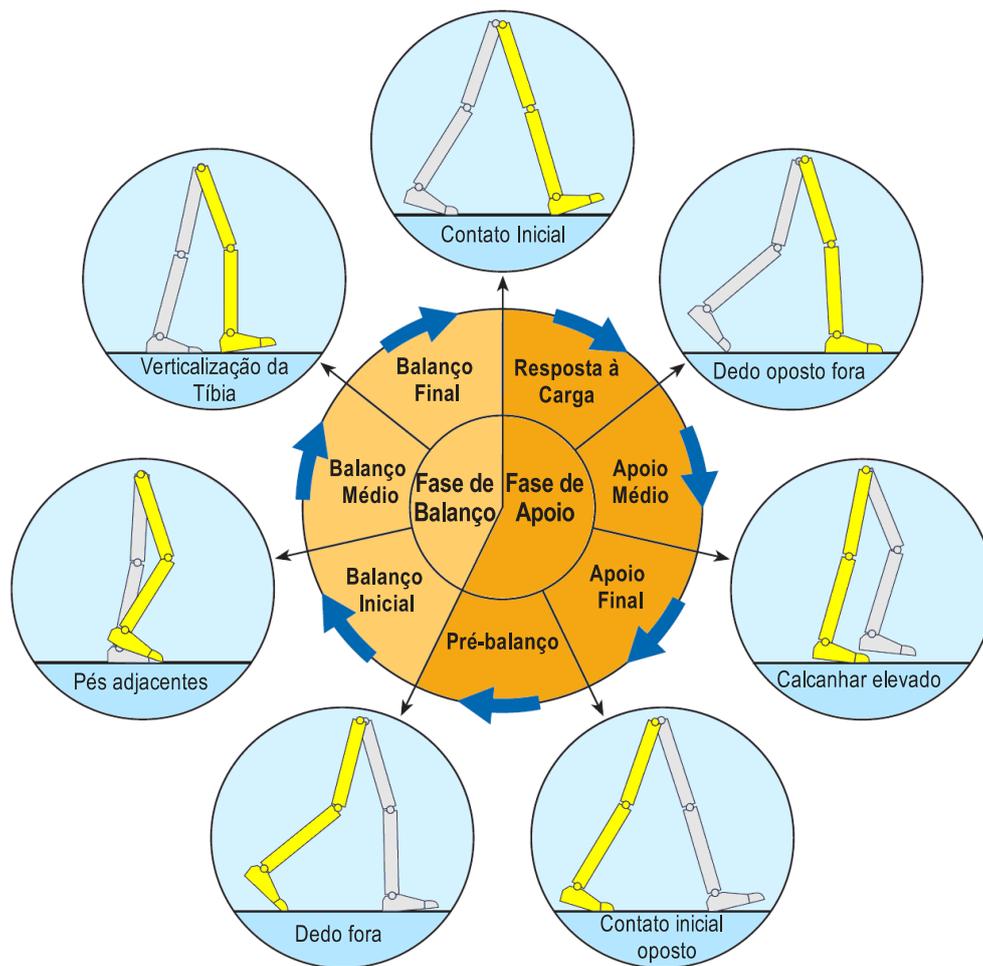


Figura 3.6. Terminologia do ciclo de marcha e posições das pernas durante o ciclo, considerando a perna direita, adaptado de Whittle (2007).

Embora qualquer evento possa ser escolhido para definir o ciclo de marcha, é conveniente utilizar o instante em que o pé escolhido como referência toca o piso, ou seja, no contato inicial. O ciclo se encerra no próximo contato inicial do pé de referência.

O **apoio médio** é o período da fase de apoio, entre os eventos “dedo oposto fora” e “calcanhar elevado”, onde a perna oposta (da fase de balanço) passa a perna de referência (da fase de apoio), a qual está apoiada no piso. Neste período, o tronco eleva-se para o seu ponto mais alto e a velocidade de deslocamento para a frente diminui, sendo convertida para a energia potencial de altura. No final do **apoio médio**, próximo ao evento “calcanhar elevado”, o joelho tem seu pico de extensão.

O **apoio médio** corresponde ao evento **pés adjacentes** da fase de balanço. Neste evento, o tronco também atinge sua posição mais alta e é deslocado de forma máxima sobre a perna de referência (da fase de apoio).

Desta forma, tanto o período de **apoio médio** (fase de apoio) quanto o evento **pés adjacentes** (fase de balanço) são os que melhor representam a altura estática num processo normal de caminhada.

3.2.3.2. Escolhendo os Melhores Quadros

Com vistas a auxiliar na escolha dos quadros do vídeo que melhor representam a altura estática de um indivíduo, é apresentada simulação de três situações cotidianas, que podem ser registradas em vídeos de vigilância: caminhada normal, caminhada simultânea com o uso de um *smartphone* e corrida, todas com calçados.

A altura estática foi mensurada com o tronco ereto, a cabeça voltada para frente e em duas situações: sem calçados, 1,880 m; e com calçados 1,895 m.

A Figura 3.7 mostra a simulação da caminhada com o tronco ereto e a cabeça voltada para frente. Nesta situação, a altura máxima foi alcançada durante o período **apoio médio** ou o evento **pés adjacentes** (dependendo do pé escolhido como referência para iniciar o ciclo de marcha), conforme explicado na subseção anterior, e foi próxima à altura estática mensurada com calçados. A altura mínima ficou em torno de 1,775 m.

A Figura 3.8 mostra a simulação da caminhada com uso simultâneo de um *smartphone*. Nesta situação, há inclinação da cabeça e com isso a altura máxima atingiu 1,790 m, mesmo no período de **apoio médio** ou evento **pés adjacentes**, ou seja, 10 cm menor que a altura estática. A altura mínima oscilou em torno de 1,750 m.

A simulação de corrida é mostrada na Figura 3.9. Durante a corrida, há a maior variação de altura, sendo a máxima alcançada de 1,765 m e a mínima de 1,620 m.

Considerando as simulações realizadas, os quadros do vídeo onde o alvo é visto caminhando com o tronco ereto, com a cabeça voltada para frente e durante o período de **apoio médio** ou evento **pés adjacentes** são elegíveis para se realizar marcações de coordenadas. As demais situações inviabilizam uma estimativa de altura.

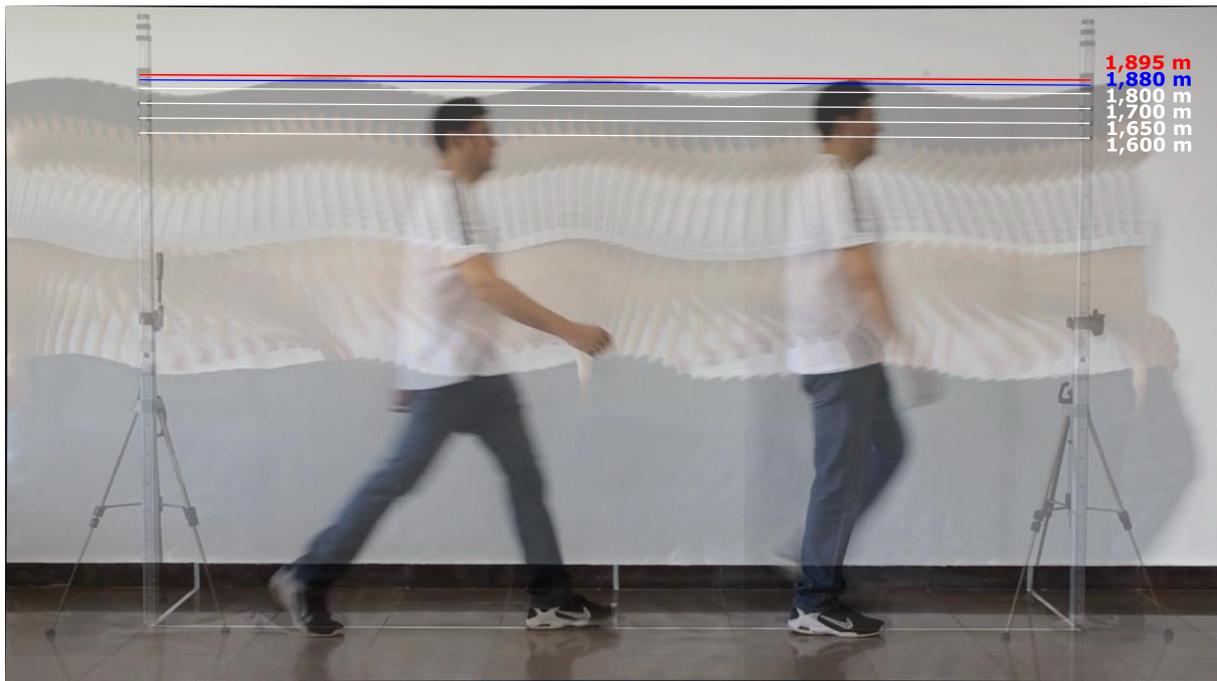


Figura 3.7. Variação da altura durante processo de caminhada normal.

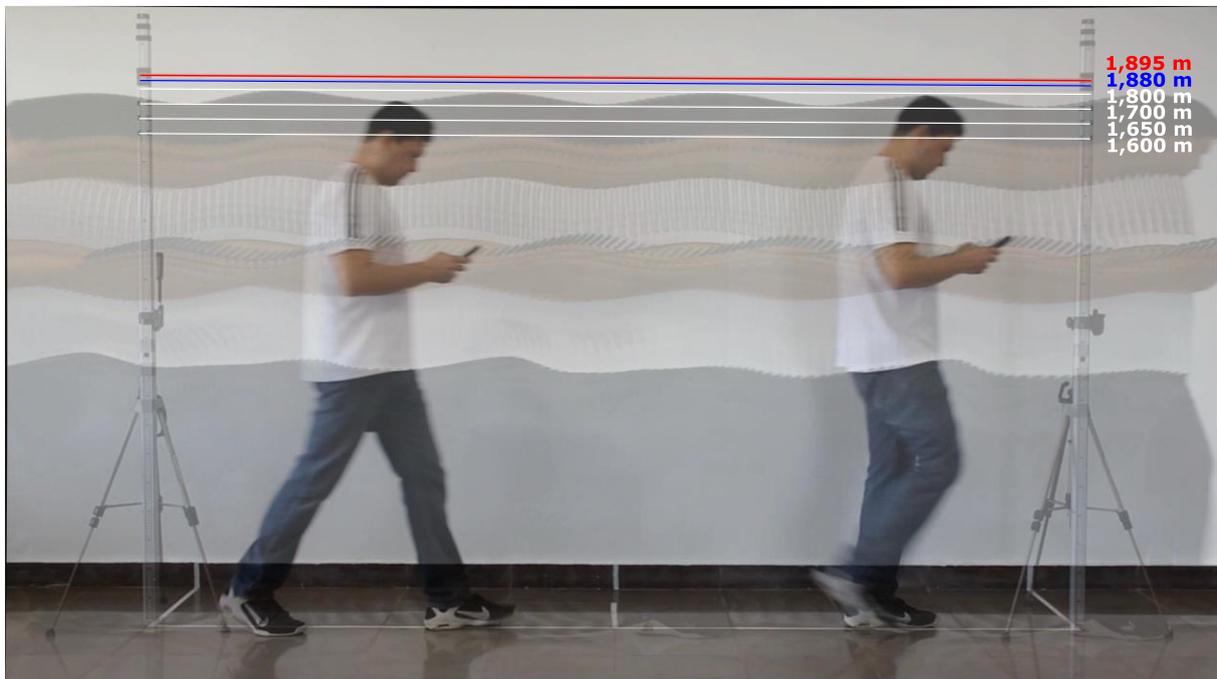


Figura 3.8. Variação da altura durante processo de caminhada simultânea com o uso de um *smartphone*.



Figura 3.9. Variação da altura durante processo de corrida.

3.2.3.3. Marcando as Coordenadas do Alvo

Em termos práticos, é comum encontrar ocorrências não favoráveis para se marcar as coordenadas inferior e superior (pés e cabeça) do alvo em quadros de vídeo – como exemplo, alvos muito distantes da câmera –, ou ainda, haver incertezas em qual o melhor ponto para indicar a cabeça e/ou os pés.

Diante dessas ocorrências, recomenda-se definir o início do ciclo de marcha do alvo a partir do membro inferior cujo calcanhar é melhor visualizado no quadro escolhido; marcar o ponto de contato do pé com o piso, e; marcar a extremidade da cabeça. Repetir esse procedimento em todos os ciclos de marcha, sempre respeitando o membro inferior escolhido como referência (perna direita, por exemplo).

A Figura 3.10 mostra um exemplo de marcações de coordenadas do alvo.

Uma ferramenta de acessibilidade que auxilia a marcação de coordenadas do alvo (e da escala) é a lupa, que está presente em alguns sistemas operacionais.

A Figura 3.11 mostra um exemplo de marcação utilizando a lupa dos sistemas operacionais da Microsoft (2016).

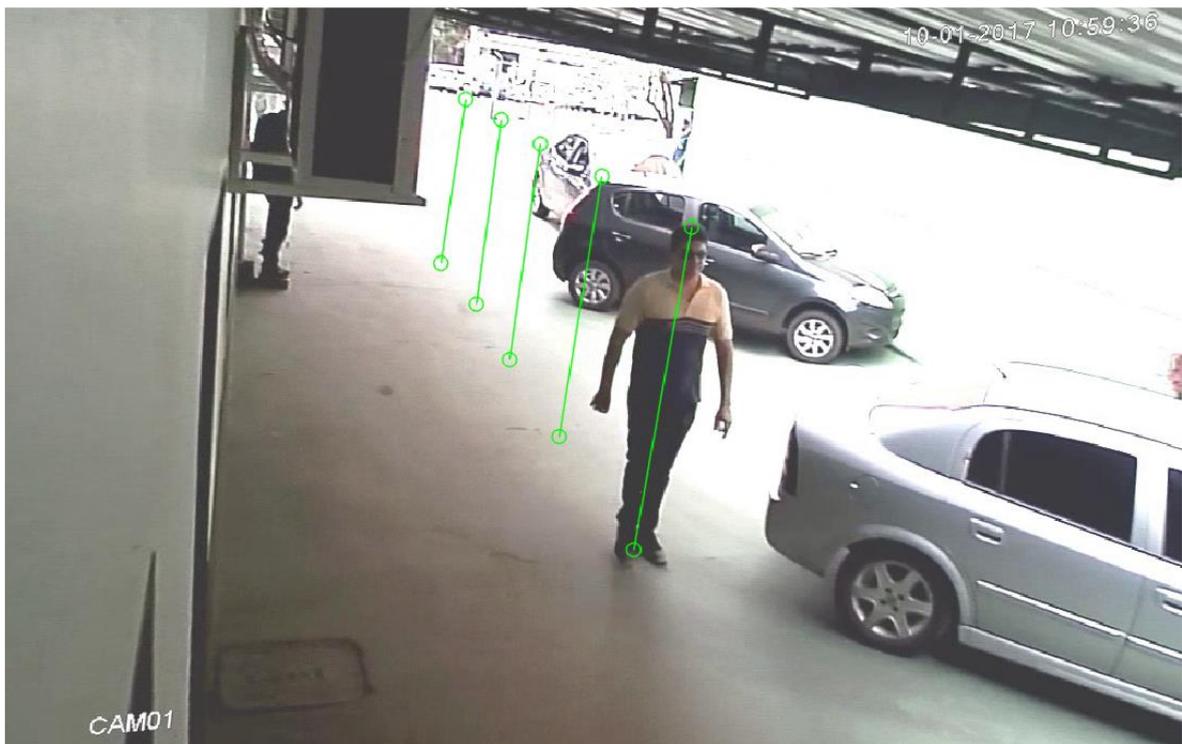


Figura 3.10. Resultado de marcações de coordenadas do alvo em quadros de um vídeo.



Figura 3.11. Exemplo de marcação de coordenadas de alvo utilizando a ferramenta de acessibilidade “lupa” dos sistemas operacionais da Microsoft (2016).

3.2.4. Estimar a Altura do Alvo

A estimativa de altura humana utilizada neste trabalho é fundamentada no método de Li et al. (2015), apresentado na seção 2.4.

No entanto, para garantir maior precisão nas estimativas, foram utilizadas também as coordenadas x das marcações do topo da escala de referência.

Assim, tomando-se o termo superior da equação (2.26) definido como x_t , a equação que determina o valor estimado x_t , denotado como \hat{x}_t , depende dos valores de x_b , y_b e Y_t e é dada pela equação (3.1):

$$\hat{x}_t(x_b, y_b, Y_t) = \frac{f c x_b - c x_b y_b \tan \theta}{(Y_t - c) \tan \theta y_b + f(Y_t \tan^2 \theta + c)}. \quad (3.1)$$

Considerando a presença de ruídos em dados reais e ε como sendo o erro produzido pelos parâmetros de calibração estimados, x_t pode ser escrito conforme a equação (3.2):

$$x_t = \hat{x}_t(x_b, y_b, Y_t) + \varepsilon. \quad (3.2)$$

Assim como a equação (2.28), a equação (3.1) também possui forma não linear. Desta forma, os parâmetros da câmera – distância focal (\hat{f}), ângulo de inclinação ($\hat{\theta}$) e a altura (\hat{c}) – podem ser obtidos por meio da equação não linear (3.3), que considera tanto as coordenadas y quando as coordenadas x do topo da escala de referência:

$$\begin{bmatrix} \hat{f} \\ \hat{\theta} \\ \hat{c} \end{bmatrix} = \arg \min_{f, \theta, c} \sum_{i=0}^N ([\hat{x}_{t_i}, \hat{y}_{t_i}] - [x_{t_i}, y_{t_i}])^2. \quad (3.3)$$

Estimados os valores \hat{f} , $\hat{\theta}$ e \hat{c} , a altura de uma pessoa (\hat{Y}) pode ser calculada pela equação (2.31).

4. RESULTADOS EXPERIMENTAIS

No intuito de verificar a confiabilidade da metodologia em casos concretos, uma base de dados de vídeos foi criada utilizando-se 15 câmeras de um sistema de vigilância de um local real.

Seis pessoas orientadas a caminhar de forma constante, olhando adiante e mantendo a postura ereta, participaram como alvos. Antes das gravações dos vídeos, a altura de cada alvo foi mensurada, considerando seus respectivos calçados.

Os experimentos seguiram as fases propostas pela metodologia, desde a Aquisição de Dados até a Análise de Dados, com a demonstração dos resultados obtidos.

4.1. COLETA DE INFORMAÇÕES DA CÂMERA DE ORIGEM E DO LOCAL DE SUA INSTALAÇÃO

O levantamento das características externas e internas das câmeras e de seus respectivos locais de instalação são apresentados na Tabela 4.1 e na Tabela 4.2.

4.2. CAPTURA DO VÍDEO DO PADRÃO PLANAR DE CALIBRAÇÃO

Para a captura do padrão planar de calibração foi criado um modelo do tipo tabuleiro com as seguintes características: dimensões de 11 x 11 quadrados, cada um com 50 mm de lado, borda de 25 mm na cor branca, paleta de cores CMYK e resolução de 300 dpi. O padrão foi impresso por um estabelecimento especializado em material vinil adesivo de cor branca e aplicado em uma placa de vidro de 5 mm de espessura.

A aproximação do padrão de calibração até a Cam10, a câmera mais alta, foi realizada por um caminhão *munck*. Para as câmeras Cam02, Cam04, Cam05, Cam07, Cam09, Cam11, Cam12 e Cam16 foi utilizada uma empilhadeira. Para as demais, utilizou-se, quando necessário, uma escada.

TABELA 4.1 – CARACTERÍSTICAS EXTERNAS DAS CÂMERAS⁶ E DE SEUS RESPECTIVOS LOCAIS DE INSTALAÇÃO.

Id. CÂMERA	LOCALIZAÇÃO	AMBIENTE	PISO	ALTURA (m)	OBSERVAÇÕES
Cam01	Pátio anterior	Externo	Active/declive	2,60	Câmera instalada em amplo local coberto, porém, com presença de luz solar no ambiente.
Cam02	Fundos lateral	Externo	Irregular ⁷	2,90	Local possui maior parte do piso irregular ⁷ , o restante é pavimentado e plano.
Cam03	Serralheria	Externo	Plano	2,50	Local com presença de materiais que podem obstruir a visualização da escala de referência.
Cam04	Vidraçaria	Interno	Active/declive	3,90	Câmera voltada para local com pouco espaço para circulação de pessoas. Necessário dispositivo especial para aproximação da câmera
Cam05	Marcenaria	Interno	Active/declive	3,20	Necessário dispositivo especial para aproximação da câmera.
Cam07	Lateral direito do galpão (frente)	Externo	Plano	3,30	Câmera instalada em local coberto reduzido, com presença de luz solar no ambiente. A cor do piso do local possui tonalidade clara.
Cam08	Almoxarifado	Interno	Plano	2,70	Ambiente com iluminação deficiente e o piso do local é de cor escura.
Cam09	Estacionamento (frente)	Externo	Plano	4,80	Presença de luz solar no ambiente. A cor do piso do local possui tonalidade clara. Necessário dispositivo especial para aproximação da câmera.
Cam10	Fundos	Externo	Irregular	7,10	Presença de grande quantidade de material em frente à câmera. Necessário dispositivo especial para aproximação.
Cam11	Estacionamento (fundos)	Externo	Plano	4,30	Câmera instalada em amplo local coberto, porém, com presença de luz solar no ambiente. A cor do piso do local possui tonalidade escura. Necessário dispositivo especial para aproximação.
Cam12	Lateral direito do galpão (fundos)	Externo	Plano	3,30	Câmera instalada em local coberto reduzido, com presença de luz solar no ambiente. A cor do piso possui tonalidade clara.
Cam13	Corredor anterior	Interno	Plano	2,90	Câmera instalada em corredor, com acentuado ângulo de inclinação. A cor do piso do local possui tonalidade clara.
Cam14	Estoque almoxarifado	Interno	Plano	3,50	Ambiente com iluminação muito deficiente e o piso do local é de cor escura. Presença de materiais que podem obstruir a visualização da escala de referência.
Cam15	Corredor posterior	Interno	Active/declive	2,60	Câmera instalada em corredor, com acentuado ângulo de inclinação. Piso de cor clara.
Cam16	Galpão	Interno	Plano	3,60	Ambiente com iluminação deficiente. A cor do piso do local possui tonalidade escura. A câmera é voltada para um dos acessos do local, com presença de luz solar na cena.

⁶ Informações de marca/modelo foram suprimidas.

⁷ Piso não pavimentado, com pequenas depressões e elevações no terreno.

TABELA 4.2 – CARACTERÍSTICAS INTERNAS DAS CÂMERAS (CONFIGURAÇÕES).

ID. CÂMERA	CONT.	FORMATO/ CODEC	RESOLUÇÃO EM PIXELS (LARGURA X ALTURA)	TAXA DE QUADROS (FPS)	RAZÃO DE ASPECTO	OBSERVAÇÕES
Cam01	AVI	AVC/ H264	1280 x 720	15	16:9	Presença de luz solar ao fundo da cena, dificultando visualização dos detalhes de objetos na área de sombra.
Cam02	AVI	AVC/ H264	1280 x 720	15	16:9	Campo de visão da câmera voltado, em grande parte, para piso irregular (não pavimentado).
Cam03	AVI	AVC/ H264	1280 x 720	15	16:9	Presença de ondulações (distorções) nos quadros do vídeo, causadas, provavelmente por interferência da rede elétrica ou por problemas nos cabos e/ou conexões do equipamento.
Cam04	AVI	AVC/ H264	1280 x 720	15	16:9	Campo de visão útil pequeno, localizado na porção inferior esquerda da cena.
Cam05	AVI	AVC/ H264	1280 x 720	15	16:9	
Cam07	AVI	AVC/ H264	928 x 480	15	16:9	Iluminação inadequada da cena devido à forte presença de luz solar.
Cam08	AVI	AVC/ H264	928 x 480	15	16:9	A cor do piso do local dificulta a visualização das extremidades inferiores dos objetos na cena.
Cam09	AVI	AVC/ H264	928 x 480	15	16:9	Presença de luz solar na cena. A cor do piso do local possui tonalidade clara. Necessário dispositivo especial para aproximação da câmera.
Cam10	AVI	AVC/ H264	928 x 480	15	16:9	Objetos aparecem distantes na cena devido à elevada altura da câmera. Campo de visão útil pequeno, localizado na porção superior esquerda da cena.
Cam11	AVI	AVC/ H264	928 x 480	15	16:9	Presença de luz solar à esquerda na cena, dificultando visualização dos detalhes de objetos na área de sombra. A cor do piso do local dificulta a visualização das extremidades inferiores dos objetos na cena.
Cam12	AVI	AVC/ H264	928 x 480	15	16:9	Presença de ondulações nos quadros do vídeo, causadas, provavelmente por interferência da rede elétrica ou por problemas nos cabos e/ou conexões do equipamento. Iluminação inadequada da cena devido à forte presença de luz solar.
Cam13	AVI	AVC/ H264	928 x 480	15	16:9	Objetos de interesse são vistos de cima para baixo e próximos da câmera.
Cam14	AVI	AVC/ H264	928 x 480	15	16:9	Vídeo com iluminação infravermelho, grande dificuldade de visualização das extremidades dos objetos de interesse.
Cam15	AVI	AVC/ H264	928 x 480	15	16:9	Vídeo com iluminação infravermelho. Objetos de interesse são vistos de cima para baixo e próximos da câmera.
Cam16	AVI	AVC/ H264	928 x 480	15	16:9	Presença de luz solar ao fundo da cena, dificultando visualização dos detalhes de objetos na área de sombra.

4.3. CAPTURA DO VÍDEO DA ESCALA DE REFERÊNCIA

Uma escala semelhante à utilizada por Li et al. (2015) foi empregada na captura do vídeo da escala de referência. Contudo, suas dimensões não permitiram uma perfeita visualização de suas extremidades nos quadros dos vídeos (quase sempre se confundiam ou com o piso ou com fundo da cena, quando provida de iluminação excessiva).

A fim de contornar as variações dos ambientes encontradas no local das gravações, foi proposta uma nova escala de referência, conforme mostrado na Figura 4.1.

Além de fornecer maior visibilidade de suas extremidades, a escala proposta possibilitou, a partir de um único posicionamento na cena, realizar mais de uma marcação de suas coordenadas (de suas laterais, por exemplo) e, com isso, obteve-se uma maior densidade de pontos de referência.

4.4. CALIBRAÇÃO DAS CÂMERAS DE ORIGEM

Os vídeos do padrão planar de calibração produzidos pelas câmeras do local foram submetidos ao processo de calibração proposto.

Os parâmetros de calibração foram calculados com sucesso para todas as câmeras, exceto para a Cam03, que apresentou erro de reprojeção de, aproximadamente, 2,077 pixels na coordenada x , superior ao limite definido no **Algoritmo 2**. Esse erro de reprojeção justifica-se pelas distorções observadas nos quadros de seus vídeos, conforme anotado na Tabela 4.2 (coluna Observações).

Assim, para a câmera Cam03, foram utilizados os parâmetros de calibração calculados para a Cam01, de mesma marca/modelo e resolução de gravação de vídeo.

Para fins de avaliação, os parâmetros de calibração da Cam14 foram copiados para a Cam09 e Cam12, também compatíveis entre si.

Os parâmetros da calibração obtidos e utilizados nas subfases de marcações de escala e alvo são apresentados na Tabela 4.3.

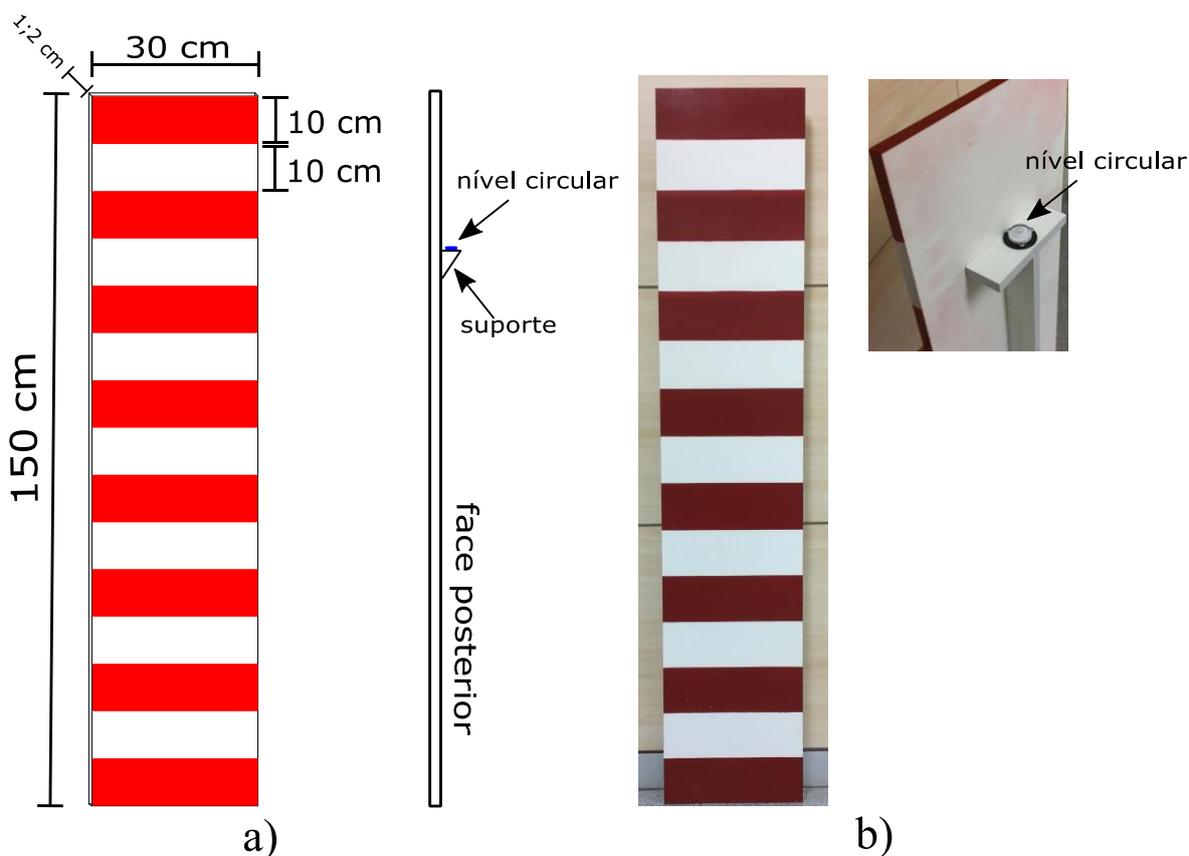


Figura 4.1. Escala de referência: a) projeto da escala utilizada na metodologia; b) escala pronta para o uso.

TABELA 4.3 – MATRIZES DAS CÂMERAS E COEFICIENTES DE DISTORÇÃO COM RESPECTIVOS ERROS DE REPROJEÇÃO.

Id.	DISTÂNCIA FOCAL		PONTO PRINCIPAL		COEFICIENTES DE DISTORÇÃO					ERRO DE REPROJEÇÃO	
	x	y	x	y	k_1	k_2	k_3	k_4	k_5	x	y
Cam01	1105,658	1066,43	672,9255	329,2899	-0,40057	0,092124	0,000637	-0,00235	0,166008	0,1764	0,168019
Cam02	1105,924	1059,822	607,2078	471,1077	-0,40958	0,167129	0,001005	0,001714	-0,02386	0,201871	0,151988
Cam03	1105,658	1066,43	672,9255	329,2899	-0,40057	0,092124	0,000637	-0,00235	0,166008	0,1764	0,168019
Cam04	1093,184	1057,004	713,8915	346,0867	-0,40915	0,176762	-0,00096	-0,00095	-0,04544	0,166156	0,139925
Cam05	1100,768	1056,672	683,0358	466,0166	-0,40631	0,19143	0,00164	0,000397	-0,04971	0,153495	0,160675
Cam07	846,1331	560,3164	547,3199	226,2291	-0,41592	0,250099	0,000123	-0,00293	-0,10268	0,288561	0,258335
Cam08	845,8404	559,6703	571,7214	207,9829	-0,4151	0,221297	-0,00066	-0,00318	-0,06805	0,258663	0,23765
Cam09	832,6206	557,7119	486,3956	209,3066	-0,42402	0,284738	-0,00283	-0,00063	-0,10266	0,246524	0,260674
Cam10	2515,809	1702,158	468,6611	261,0056	-0,18649	10,03937	0,001004	-0,00557	-171,845	0,201969	0,205891
Cam11	832,6206	557,7119	486,3956	209,3066	-0,42402	0,284738	-0,00283	-0,00063	-0,10266	0,246524	0,260674
Cam12	838,6405	554,708	504,569	189,5044	-0,43572	0,391261	0,002952	-0,00045	-0,34473	0,43392	0,244501
Cam13	834,7008	553,6176	487,6803	227,3729	-0,40523	0,252688	0,001364	-0,00022	-0,13234	0,244417	0,239397
Cam14	832,6206	557,7119	486,3956	209,3066	-0,42402	0,284738	-0,00283	-0,00063	-0,10266	0,246524	0,260674
Cam15	838,1058	555,5813	565,2658	206,4078	-0,40776	0,20805	0,002056	-0,0022	-0,05759	0,256599	0,318057
Cam16	827,5115	547,8291	475,3819	243,5458	-0,39225	0,134332	0,00129	-0,0063	0,191208	0,280825	0,238453

TABELA 4.4 – ESTIMATIVAS DE ALTURA DA ESCALA01.

ID. CÂMERA	AMOSTRAS	VLR. VERD. (A)	MÉDIA EST. (B)	DESV. PAD. EXP.	DESV. PAD. EXP. MÉDIA	MÍN.	MÁX.	(A-B)	RESULTADO CONFIANÇA DE 99%
Cam01	64	1,500	1,500	0,049	0,006	1,394	1,582	0,000	(1,50 ± 0,02) m (k=2,66)
Cam02	42	1,500	1,500	0,014	0,002	1,462	1,525	0,000	(1,50 ± 0,01) m (k=2,70)
Cam03	77	1,500	1,499	0,047	0,005	1,405	1,604	0,001	(1,50 ± 0,01) m (k=2,64)
Cam04	28	1,500	1,500	0,032	0,006	1,442	1,550	0,000	(1,50 ± 0,02) m (k=2,77)
Cam05	72	1,500	1,500	0,050	0,006	1,421	1,606	0,000	(1,50 ± 0,02) m (k=2,65)
Cam07	42	1,500	1,501	0,015	0,002	1,464	1,539	0,001	(1,50 ± 0,01) m (k=2,70)
Cam08	24	1,500	1,500	0,044	0,009	1,424	1,580	0,000	(1,50 ± 0,02) m (k=2,81)
Cam09	90	1,500	1,500	0,037	0,004	1,424	1,587	0,000	(1,50 ± 0,01) m (k=2,63)
Cam10	34	1,500	1,498	0,029	0,005	1,428	1,574	0,002	(1,50 ± 0,01) m (k=2,73)
Cam11	92	1,500	1,500	0,042	0,004	1,400	1,598	0,000	(1,50 ± 0,01) m (k=2,63)
Cam12	72	1,500	1,500	0,022	0,003	1,450	1,537	0,000	(1,50 ± 0,01) m (k=2,65)
Cam13	97	1,500	1,499	0,028	0,003	1,437	1,544	0,001	(1,50 ± 0,01) m (k=2,63)
Cam14	62	1,500	1,500	0,058	0,007	1,395	1,603	0,000	(1,50 ± 0,02) m (k=2,66)
Cam15	80	1,500	1,499	0,036	0,004	1,425	1,573	0,001	(1,50 ± 0,01) m (k=2,64)
Cam16	43	1,500	1,495	0,044	0,007	1,405	1,572	0,005	(1,50 ± 0,02) m (k=2,70)

4.5. MARCAÇÕES DE COORDENADAS DA ESCALA DE REFERÊNCIA

As marcações de coordenadas da escala realizadas neste experimento procuraram não somente reproduzir as trajetórias dos alvos, mas também fornecer elevada quantidade de pontos de referência para os ajustes da regressão não linear utilizada na metodologia

A coluna “Amostras” da Tabela 4.4 mostra a quantidade de marcações realizadas em cada uma das câmeras para escala de referência identificada como ESCALA01.

4.6. MARCAÇÕES DE COORDENADAS DOS ALVOS

Na maior parte dos vídeos, as marcações dos alvos seguiram os passos sugeridos na subseção 3.2.3. As exceções ocorreram com as marcações realizadas nos vídeos produzidos pelas câmeras Cam13 e Cam15, instaladas de frente para um corredor, onde os alvos eram vistos superiormente (de cima para baixo) e muito próximos às câmeras.

Na Cam13, os alvos foram capturados afastando-se da câmera. Nessa posição, apesar da maior facilidade de marcação de suas coordenadas inferiores – calcanhares –, houve dificuldades de marcação das coordenadas superiores – ausência de referências na região da cabeça.

Na Cam15, ao contrário da Cam13, os alvos foram gravados aproximando-se da câmera. Nessa situação, as respectivas faces dos alvos foram utilizadas como referência para marcar os seus limites superiores, por outro lado, houve dificuldades na determinação dos seus limites inferiores – ponto em que o pé toca o piso.

Exemplos de marcações de coordenadas dos alvos, respectivamente, nas câmeras Cam13 e Cam15. Em ambas, foi considerada como limite superior a porção superior da cabeça e como limite inferior o ponto de contato de ambos os pés com o piso, respeitado o ciclo de marcha, conforme indicado na Figura 4.2.



Figura 4.2. Exemplo de marcação de coordenadas de alvos: a) quadro da câmera Cam13; b) quadro da câmera Cam15.

4.7. ESTIMATIVAS DE ALTURA

Dois resultados de estimativa de altura são apresentados: um para a escala, com vista a verificar e validar as marcações de referência, e outra para os alvos. Os resultados são exibidos com intervalos estimados – ou incertezas expandidas (U) – para terem um nível de confiança de 99%. A incerteza padrão combinada (u_c) é composta somente pelo desvio-padrão experimental da média (uma vez que a altura é uma grandeza única e normalmente distribuída) e o fator de abrangência (k) é avaliado com base na distribuição- t (distribuição de *Student*) (JCGM, 2008).

A altura de cada uma das câmeras, obtida na fase de Aquisição de Dados, foi utilizada como entrada nos ajustes da equação (3.3), realizados com o método de Levenberg-Marquardt (mínimos quadrados).

Os resultados de validação da ESCALA01 são mostrados na Tabela 4.4. A distribuição dos erros é mostrada na Figura 4.3.

O erro médio, o desvio padrão experimental e o erro médio absoluto foram de 0,001 m e o erro absoluto máximo foi de 0,005 m.

Os resultados dos alvos são apresentados na Tabela 4.5. A distribuição dos erros é mostrada na Figura 4.4. O erro médio, o desvio padrão experimental, o erro médio absoluto e o erro absoluto máximo foram, respectivamente, de 0,008 m, 0,014 m, 0,013 m e 0,054 m.

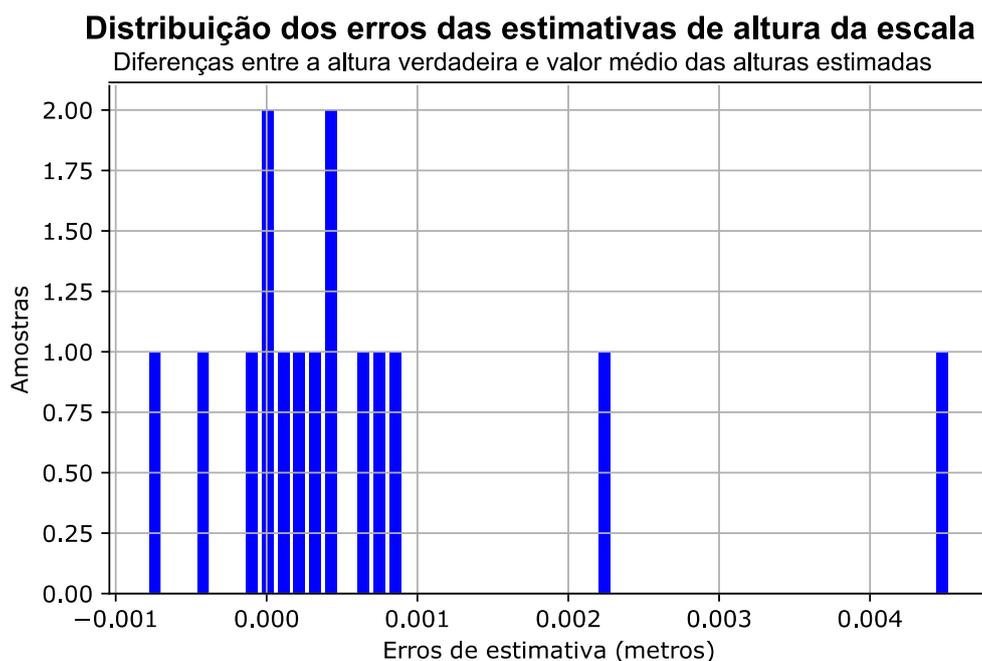


Figura 4.3. Histograma dos erros de estimativa de altura da escala.

Distribuição dos erros das estimativas de altura dos alvos

Diferenças entre a altura verdadeira e valor médio das alturas estimadas

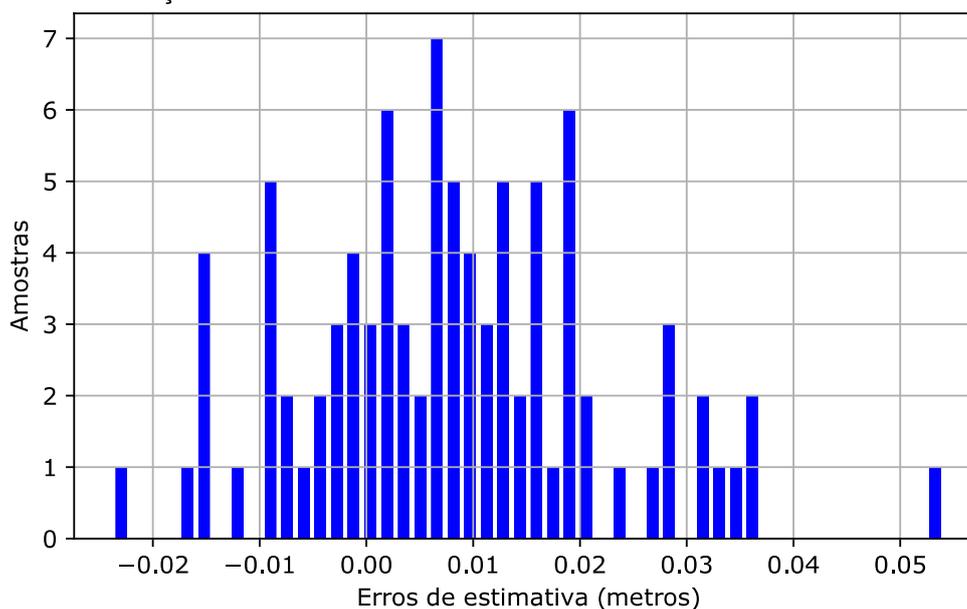


Figura 4.4. Histograma dos erros de estimativa de altura dos alvos.

TABELA 4.5 – ESTIMATIVAS DE ALTURA DOS ALVOS.

Id. CÂMERA	Id. ALVO	AMOSTRAS	VLR. VERD. (A)	MÉDIA EST. (B)	DESV. PAD. EXP.	DESV. PAD. EXP. MÉDIA	MÍN.	MÁX.	(A-B)	RESULTADO CONFIANÇA DE 99%
Cam01	alvo01	20	1,895	1,881	0,023	0,005	1,850	1,935	0,014	(1,88 ± 0,01) m (k=2,86)
Cam01	alvo02	18	1,780	1,772	0,025	0,006	1,717	1,804	0,008	(1,77 ± 0,02) m (k=2,90)
Cam01	alvo03	20	1,825	1,840	0,037	0,008	1,766	1,895	0,015	(1,84 ± 0,02) m (k=2,86)
Cam01	alvo04	21	1,735	1,728	0,042	0,009	1,634	1,790	0,007	(1,73 ± 0,03) m (k=2,85)
Cam01	alvo05	22	1,610	1,579	0,035	0,008	1,529	1,642	0,031	(1,58 ± 0,02) m (k=2,83)
Cam01	alvo06	17	1,805	1,803	0,021	0,005	1,764	1,832	0,002	(1,80 ± 0,01) m (k=2,92)
Cam02	alvo01	16	1,895	1,877	0,012	0,003	1,855	1,892	0,018	(1,88 ± 0,01) m (k=2,95)
Cam02	alvo02	12	1,780	1,768	0,007	0,002	1,761	1,782	0,012	(1,77 ± 0,01) m (k=3,11)
Cam02	alvo03	11	1,825	1,822	0,007	0,002	1,812	1,832	0,003	(1,82 ± 0,01) m (k=3,17)
Cam02	alvo04	13	1,735	1,732	0,018	0,005	1,697	1,759	0,003	(1,73 ± 0,01) m (k=3,05)
Cam02	alvo05	11	1,610	1,586	0,014	0,004	1,569	1,608	0,024	(1,59 ± 0,01) m (k=3,17)
Cam02	alvo06	12	1,805	1,795	0,015	0,004	1,773	1,820	0,010	(1,79 ± 0,01) m (k=3,11)
Cam03	alvo01	15	1,895	1,894	0,045	0,012	1,814	1,937	0,001	(1,89 ± 0,03) m (k=2,98)
Cam03	alvo02	17	1,780	1,783	0,060	0,015	1,707	1,871	0,003	(1,78 ± 0,04) m (k=2,92)
Cam03	alvo03	15	1,825	1,813	0,052	0,014	1,753	1,911	0,012	(1,81 ± 0,04) m (k=2,98)
Cam03	alvo04	15	1,735	1,727	0,064	0,017	1,659	1,832	0,008	(1,73 ± 0,05) m (k=2,98)
Cam03	alvo05	16	1,610	1,605	0,053	0,013	1,532	1,676	0,005	(1,60 ± 0,04) m (k=2,95)
Cam03	alvo06	16	1,805	1,777	0,041	0,010	1,718	1,835	0,028	(1,78 ± 0,03) m (k=2,95)
Cam04	alvo01	15	1,895	1,879	0,031	0,008	1,840	1,921	0,016	(1,88 ± 0,02) m (k=2,98)
Cam04	alvo02	14	1,780	1,792	0,025	0,007	1,769	1,837	0,012	(1,79 ± 0,02) m (k=3,01)
Cam04	alvo03	12	1,825	1,823	0,022	0,006	1,795	1,858	0,002	(1,82 ± 0,02) m (k=3,11)
Cam04	alvo04	14	1,735	1,745	0,018	0,005	1,721	1,765	0,010	(1,74 ± 0,01) m (k=3,01)

Cam04	alvo05	16	1,610	1,597	0,026	0,007	1,555	1,624	0,013	(1,60 ± 0,02) m (k=2,95)
Cam04	alvo06	11	1,805	1,799	0,035	0,011	1,743	1,842	0,006	(1,80 ± 0,03) m (k=3,17)
Cam05	alvo01	15	1,895	1,897	0,010	0,003	1,886	1,922	0,002	(1,90 ± 0,01) m (k=2,98)
Cam05	alvo02	19	1,780	1,783	0,017	0,004	1,753	1,810	0,003	(1,78 ± 0,01) m (k=2,88)
Cam05	alvo03	18	1,825	1,840	0,016	0,004	1,806	1,867	0,015	(1,84 ± 0,01) m (k=2,90)
Cam05	alvo04	16	1,735	1,759	0,013	0,003	1,734	1,783	0,024	(1,76 ± 0,01) m (k=2,95)
Cam05	alvo05	16	1,610	1,598	0,011	0,003	1,579	1,620	0,012	(1,60 ± 0,01) m (k=2,95)
Cam05	alvo06	16	1,805	1,812	0,019	0,005	1,783	1,845	0,007	(1,81 ± 0,01) m (k=2,95)
Cam07	alvo01	15	1,895	1,876	0,016	0,004	1,841	1,894	0,019	(1,88 ± 0,01) m (k=2,98)
Cam07	alvo02	19	1,780	1,752	0,021	0,005	1,719	1,784	0,028	(1,75 ± 0,01) m (k=2,88)
Cam07	alvo03	17	1,825	1,816	0,020	0,005	1,781	1,855	0,009	(1,82 ± 0,01) m (k=2,92)
Cam07	alvo04	16	1,735	1,736	0,030	0,007	1,670	1,779	0,001	(1,74 ± 0,02) m (k=2,95)
Cam07	alvo05	19	1,610	1,583	0,021	0,005	1,546	1,610	0,027	(1,58 ± 0,01) m (k=2,88)
Cam07	alvo06	15	1,805	1,790	0,021	0,005	1,766	1,834	0,015	(1,79 ± 0,02) m (k=2,98)
Cam08	alvo01	10	1,895	1,889	0,028	0,009	1,862	1,927	0,006	(1,89 ± 0,03) m (k=3,25)
Cam08	alvo02	12	1,780	1,778	0,020	0,006	1,755	1,801	0,002	(1,78 ± 0,02) m (k=3,11)
Cam08	alvo03	10	1,825	1,829	0,009	0,003	1,820	1,842	0,004	(1,83 ± 0,01) m (k=3,25)
Cam08	alvo04	13	1,735	1,735	0,039	0,011	1,682	1,780	0,000	(1,74 ± 0,03) m (k=3,05)
Cam08	alvo05	11	1,610	1,620	0,029	0,009	1,591	1,678	0,010	(1,62 ± 0,03) m (k=3,17)
Cam08	alvo06	11	1,805	1,801	0,038	0,011	1,769	1,862	0,004	(1,80 ± 0,04) m (k=3,17)
Cam09	alvo01	13	1,895	1,879	0,015	0,004	1,856	1,903	0,016	(1,88 ± 0,01) m (k=3,05)
Cam09	alvo02	14	1,780	1,767	0,015	0,004	1,741	1,797	0,013	(1,77 ± 0,01) m (k=3,01)
Cam09	alvo03	13	1,825	1,817	0,023	0,006	1,768	1,845	0,008	(1,82 ± 0,02) m (k=3,05)
Cam09	alvo04	14	1,735	1,743	0,015	0,004	1,715	1,759	0,008	(1,74 ± 0,01) m (k=3,01)
Cam09	alvo05	14	1,610	1,603	0,018	0,005	1,564	1,625	0,007	(1,60 ± 0,01) m (k=3,01)
Cam09	alvo06	14	1,805	1,796	0,015	0,004	1,772	1,822	0,009	(1,80 ± 0,01) m (k=3,01)
Cam10	alvo01	18	1,895	1,841	0,020	0,005	1,795	1,867	0,054	(1,84 ± 0,01) m (k=2,90)
Cam10	alvo02	18	1,780	1,743	0,025	0,006	1,705	1,808	0,037	(1,74 ± 0,02) m (k=2,90)
Cam10	alvo03	18	1,825	1,791	0,032	0,008	1,747	1,846	0,034	(1,79 ± 0,02) m (k=2,90)
Cam10	alvo04	18	1,735	1,706	0,021	0,005	1,655	1,737	0,029	(1,71 ± 0,01) m (k=2,90)
Cam10	alvo05	20	1,610	1,576	0,022	0,005	1,537	1,623	0,034	(1,58 ± 0,01) m (k=2,86)
Cam10	alvo06	22	1,805	1,768	0,025	0,005	1,736	1,828	0,037	(1,77 ± 0,02) m (k=2,83)
Cam11	alvo01	16	1,895	1,879	0,042	0,010	1,789	1,938	0,016	(1,88 ± 0,03) m (k=2,95)
Cam11	alvo02	17	1,780	1,778	0,054	0,013	1,695	1,854	0,002	(1,78 ± 0,04) m (k=2,92)
Cam11	alvo03	17	1,825	1,834	0,047	0,011	1,748	1,895	0,009	(1,83 ± 0,03) m (k=2,92)
Cam11	alvo04	22	1,735	1,745	0,033	0,007	1,674	1,816	0,010	(1,74 ± 0,02) m (k=2,83)
Cam11	alvo05	16	1,610	1,614	0,028	0,007	1,574	1,659	0,004	(1,61 ± 0,02) m (k=2,95)
Cam11	alvo06	17	1,805	1,821	0,036	0,009	1,765	1,876	0,016	(1,82 ± 0,03) m (k=2,92)
Cam12	alvo01	18	1,895	1,877	0,029	0,007	1,837	1,932	0,018	(1,88 ± 0,02) m (k=2,90)
Cam12	alvo02	19	1,780	1,760	0,030	0,007	1,712	1,816	0,020	(1,76 ± 0,02) m (k=2,88)
Cam12	alvo03	16	1,825	1,815	0,029	0,007	1,747	1,871	0,010	(1,81 ± 0,02) m (k=2,95)
Cam12	alvo04	12	1,735	1,744	0,021	0,006	1,723	1,774	0,009	(1,74 ± 0,02) m (k=3,11)
Cam12	alvo05	11	1,610	1,604	0,026	0,008	1,565	1,648	0,006	(1,60 ± 0,02) m (k=3,17)
Cam12	alvo06	12	1,805	1,798	0,031	0,009	1,759	1,842	0,007	(1,80 ± 0,03) m (k=3,11)
Cam13	alvo01	11	1,895	1,891	0,018	0,005	1,875	1,924	0,004	(1,89 ± 0,02) m (k=3,17)
Cam13	alvo02	12	1,780	1,795	0,018	0,005	1,767	1,829	0,015	(1,79 ± 0,02) m (k=3,11)
Cam13	alvo03	12	1,825	1,842	0,012	0,003	1,823	1,858	0,017	(1,84 ± 0,01) m (k=3,11)

Cam13	alvo04	11	1,735	1,727	0,018	0,005	1,692	1,744	0,008	(1,73 ± 0,02) m (k=3,17)
Cam13	alvo05	12	1,610	1,598	0,016	0,005	1,564	1,622	0,012	(1,60 ± 0,01) m (k=3,11)
Cam13	alvo06	11	1,805	1,811	0,024	0,007	1,783	1,848	0,006	(1,81 ± 0,02) m (k=3,17)
Cam14	alvo01	17	1,895	1,863	0,029	0,007	1,819	1,909	0,032	(1,86 ± 0,02) m (k=2,92)
Cam14	alvo02	19	1,780	1,769	0,055	0,013	1,690	1,865	0,011	(1,77 ± 0,04) m (k=2,88)
Cam14	alvo03	18	1,825	1,804	0,041	0,010	1,708	1,857	0,021	(1,80 ± 0,03) m (k=2,90)
Cam14	alvo04	19	1,735	1,729	0,030	0,007	1,687	1,782	0,006	(1,73 ± 0,02) m (k=2,88)
Cam14	alvo05	20	1,610	1,592	0,046	0,010	1,507	1,656	0,018	(1,59 ± 0,03) m (k=2,86)
Cam14	alvo06	21	1,805	1,795	0,037	0,008	1,733	1,856	0,010	(1,80 ± 0,02) m (k=2,85)
Cam15	alvo01	12	1,895	1,890	0,020	0,006	1,856	1,912	0,005	(1,89 ± 0,02) m (k=3,11)
Cam15	alvo02	15	1,780	1,778	0,015	0,004	1,755	1,802	0,002	(1,78 ± 0,01) m (k=2,98)
Cam15	alvo03	15	1,825	1,826	0,013	0,003	1,806	1,851	0,001	(1,83 ± 0,01) m (k=2,98)
Cam15	alvo04	16	1,735	1,736	0,011	0,003	1,716	1,757	0,001	(1,74 ± 0,01) m (k=2,95)
Cam15	alvo05	12	1,610	1,598	0,045	0,013	1,536	1,653	0,012	(1,60 ± 0,04) m (k=3,11)
Cam15	alvo06	13	1,805	1,805	0,018	0,005	1,778	1,834	0,000	(1,81 ± 0,01) m (k=3,05)
Cam16	alvo01	13	1,895	1,876	0,034	0,009	1,825	1,916	0,019	(1,88 ± 0,03) m (k=3,05)
Cam16	alvo02	15	1,780	1,765	0,023	0,006	1,730	1,796	0,015	(1,76 ± 0,02) m (k=2,98)
Cam16	alvo03	13	1,825	1,806	0,034	0,010	1,741	1,847	0,019	(1,81 ± 0,03) m (k=3,05)
Cam16	alvo04	14	1,735	1,736	0,024	0,007	1,692	1,771	0,001	(1,74 ± 0,02) m (k=3,01)
Cam16	alvo05	18	1,610	1,591	0,016	0,004	1,570	1,632	0,019	(1,59 ± 0,01) m (k=2,90)
Cam16	alvo06	11	1,805	1,790	0,014	0,004	1,772	1,809	0,015	(1,79 ± 0,01) m (k=3,17)

A Figura 4.5 mostra as correlações entre os erros de estimativa de altura com as respectivas coordenadas das marcações dos alvos.

Os valores dos coeficientes de Pearson (r)⁸ – respectivamente de 0,10 (erros com coordenadas superiores de x), 0,09 (erros com coordenadas inferiores de x), -0,19 (erros com coordenadas superiores de y) e -0,09 (erros com coordenadas inferiores de y) – mostram que as correlações entre os erros de estimativa de altura e as coordenadas x e y das marcações dos alvos são desprezíveis.

⁸ Correlação desprezível: de $\pm 0,00$ até $\pm 0,19$; correlação fraca: de $\pm 0,20$ até $\pm 0,39$; correlação moderada: de $\pm 0,40$ até $\pm 0,69$; correlação forte: de $\pm 0,70$ até $\pm 0,89$; correlação muito forte: de $\pm 0,90$ até $\pm 1,00$.

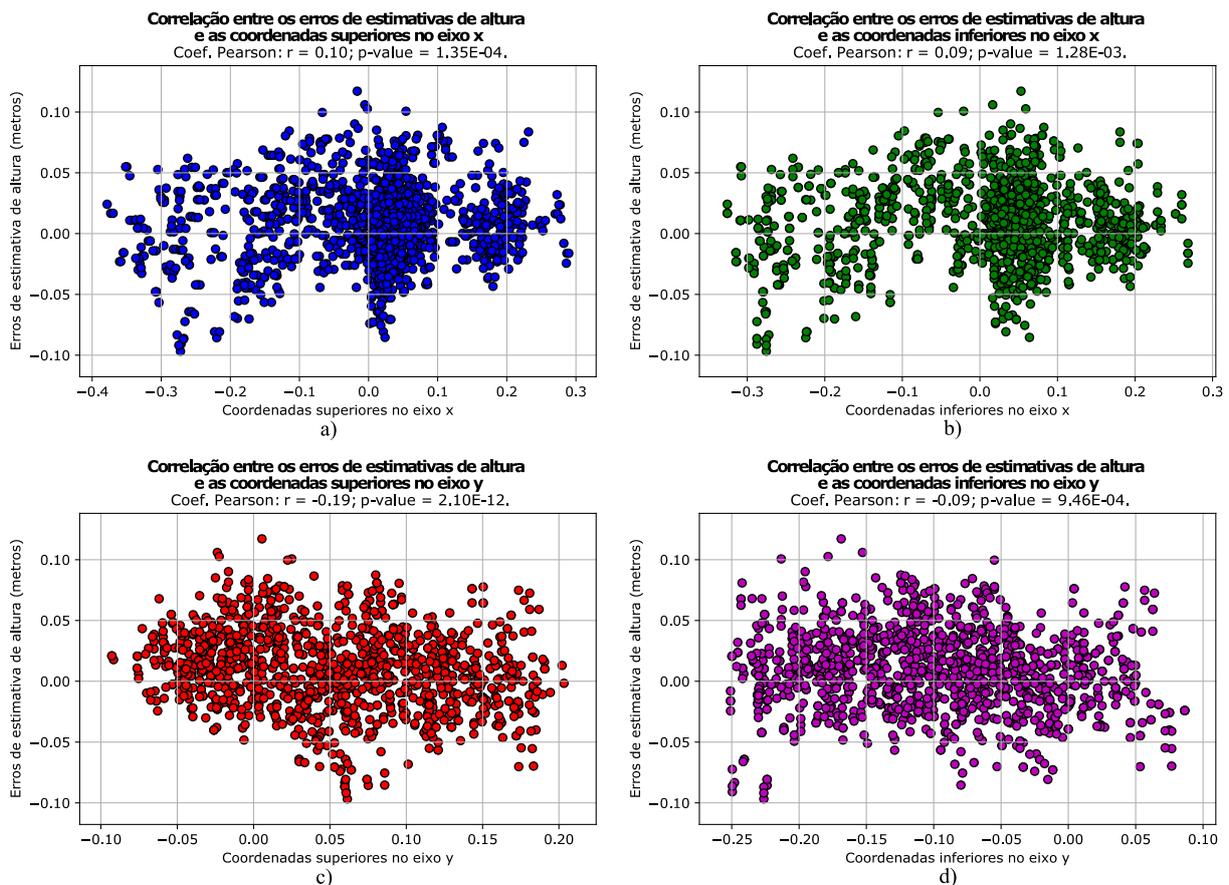


Figura 4.5. Gráfico de correlações: a) erros de estimativa de altura e coordenadas superiores do eixo x; b) erros de estimativa de altura e coordenadas inferiores do eixo x; c) erros de estimativa de altura e coordenadas superiores do eixo y; b) erros de estimativa de altura e coordenadas inferiores do eixo y.

4.8. ALTERAÇÕES NO CAMPO DE VISÃO DA CÂMERA

Alterações do campo de visão de uma câmera alteram as características de perspectiva da cena capturada. Com o intuito de se verificar o impacto dessas alterações nos resultados de estimativa de altura da metodologia proposta, foi realizada uma simulação com cinco diferentes cenários – utilizando-se uma câmera, identificada como Cam99, instalada a 2,53 m de altura e calibrada conforme módulo de calibração de câmera proposto; um alvo de 1,895 m de altura, com calçados; a escala de referência com altura de 1,4 m – da seguinte forma: A) o alvo e a escala de referência, denominada de ESCALA01, foram capturados com câmera na mesma posição; B) a altura da câmera foi alterada para 2,41 m e seu ângulo de inclinação mantido, nesta nova posição, capturou-se a escala como ESCALA02; C) a câmera foi elevada para a sua altura inicial (2,53 m) e seu ângulo de inclinação elevado, capturou-se, então, a escala como

ESCALA03; D) a câmera foi mantida na altura inicial e o ângulo de inclinação novamente elevado, assim, a escala foi capturada como ESCALA04; E) a câmera teve sua altura modificada para 2,49 m e seu ângulo de inclinação ajustado para uma posição aproximada da inicial e, por fim, capturou-se a escala como ESCALA05.

A comparação entre o campo de visão do cenário A e os campos de visão dos cenários B, C, D e E são, respectivamente, mostrados na Figura 4.6, Figura 4.7, Figura 4.8 e Figura 4.9

As alturas estimadas do alvo, considerando suas marcações de coordenadas realizadas no vídeo do cenário A e as marcações de coordenadas da escala de referência realizadas em cada um dos cinco cenários, são apresentadas na Tabela 4.6. Com exceção do cenário A, os demais apresentam valores de erro absoluto elevados – mínimo de 0,086 m e máximo de 0,531 m.

Com a finalidade de contornar as situações de alteração do campo de visão, os vídeos dos cenários B, C, D e E foram submetidos a ajustes de perspectiva, com auxílio de objetos e estruturas da cena e baseados na perspectiva do cenário A (Figura 4.10). As coordenadas da escala de referência foram novamente marcadas nos vídeos ajustados e os valores estimados de altura do alvo são apresentados na Tabela 4.7. Observa-se que os erros absolutos obtidos com os ajustes de perspectiva possuem valores aceitáveis – mínimo de 0,000 m e máximo de 0,014 m.

Portanto, desde que seja possível ajustar a perspectiva do vídeo que contém a escala de referência, a partir de objetos e estruturas na cena e da perspectiva da cena original (no momento em que o alvo foi capturado), a metodologia pode ser aplicada mesmo havendo alterações do campo de visão da câmera.



Figura 4.6. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário B, onde a escala de referência foi capturada como ESCALA02.



Figura 4.7. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário C, onde a escala de referência foi capturada como ESCALA03.



Figura 4.8. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário D, onde a escala de referência foi capturada como ESCALA04.



Figura 4.9. Comparação entre o campo de visão do cenário A, onde foram capturados o alvo e a escala de referência como ESCALA01, e o campo de visão do cenário E, onde a escala de referência foi capturada como ESCALA05.



Figura 4.10. Marcações de coordenadas da escala de referência, denominada, respectivamente, de ESCALA02, ESCALA03, ESCALA04 e ESCALA05 nos vídeos dos cenários B, C, D e E, após ajustes de perspectiva em cada um deles.

TABELA 4.6 – ESTIMATIVAS DE ALTURA DO ALVO CONSIDERANDO OS CENÁRIOS DE POSICIONAMENTO DA CÂMERA CAM99.

ID. ESCALA	AMOSTRAS	VLR. VERD. (A)	MÉDIA EST. (B)	DESV. PAD. EXP.	DESV. PAD. EXP. MÉDIA	MÍN.	MÁX.	(A-B)	RESULTADO CONFIANÇA DE 99%
ESCALA01	20	1,895	1,893	0,018	0,004	1,851	1,925	0,002	(1,89 ± 0,01) m (k=2,86)
ESCALA02	20	1,895	1,996	0,086	0,019	1,863	2,141	0,101	(2,00 ± 0,05) m (k=2,86)
ESCALA03	20	1,895	2,050	0,054	0,012	1,949	2,138	0,155	(2,05 ± 0,03) m (k=2,86)
ESCALA04	20	1,895	2,426	0,207	0,046	2,161	2,784	0,531	(2,43 ± 0,13) m (k=2,86)
ESCALA05	20	1,895	1,809	0,040	0,009	1,743	1,877	0,086	(1,81 ± 0,03) m (k=2,86)

TABELA 4.7 – ESTIMATIVAS DE ALTURA DO ALVO CONSIDERANDO OS CENÁRIOS DE POSICIONAMENTO DA CÂMERA CAM99 APÓS AJUSTES DE PERSPECTIVA.

ID. ESCALA	AMOSTRAS	VLR. VERD. (A)	MÉDIA EST. (B)	DESV. PAD. EXP.	DESV. PAD. EXP. MÉDIA	MÍN.	MÁX.	(A-B)	RESULTADO CONFIANÇA DE 99%
ESCALA02	20	1,895	1,890	0,019	0,004	1,838	1,918	0,005	(1,89 ± 0,01) m (k=2,86)
ESCALA03	20	1,895	1,909	0,021	0,005	1,852	1,942	0,014	(1,91 ± 0,01) m (k=2,86)
ESCALA04	20	1,895	1,883	0,033	0,007	1,829	1,943	0,012	(1,88 ± 0,02) m (k=2,86)
ESCALA05	20	1,895	1,888	0,020	0,005	1,837	1,920	0,007	(1,89 ± 0,01) m (k=2,86)

4.9. DISCUSSÃO

A grande distância de um alvo da câmera de origem (por exemplo, nos vídeos da câmera Cam10), bem como a iluminação inadequada da cena (por exemplo, nos vídeos das câmeras Cam07, Cam14 e Cam16), dificultam a visualização das extremidades dos objetos de interesse (alvos) durante as marcações de coordenadas. Consequentemente, marcações incorretas geram erros nos resultados de estimativa de altura. Isso pode ser notado nos erros médios de estimativa de altura de cada câmera, sendo a Cam10 a que apresentou maior erro, 0,032 m, seguida pelas Cam14 e Cam07, ambas com 0,014 m, e a Cam16 com 0,013 m de erro médio. A Figura 4.11 mostra como o alvo é visto nos vídeos da câmera Cam10.

As variações de nível dos pisos e as irregularidades dos terrenos não produziram erros perceptíveis nos resultados. A densidade de pontos das marcações de escala de referência, bem como seu correto posicionamento na cena (locais onde os alvos foram vistos) reduzem a influência dos diferentes tipos de piso/terreno nas estimativas de altura.

O compartilhamento de parâmetros de calibração entre câmeras compatíveis mostrou-se efetivo, como pode ser observado nos resultados de estimativa de altura apresentados nas câmeras Cam03 (com parâmetros copiados da Cam01), Cam09 e Cam12 (com parâmetros copiados da Cam14).

A distorção nos quadros dos vídeos das câmeras Cam03 e Cam12 (consoante à coluna Observações da Tabela 4.2), perceptíveis mesmo após a retificação, não influenciou nos resultados dos cálculos de altura dos alvos. Diferentemente das distorções radiais, que são mais expressivas nas bordas dos quadros, a distorção vista nos vídeos da câmera Cam12 e, especialmente, nos da câmera Cam13, apresentou-se de forma linear (coordenadas do eixo x), tanto que, nesta última, não foi possível calcular seus parâmetros de calibração, conforme explicado na seção 4.4. Uma ilustração de um quadro de vídeo produzido pela Cam03 é apresentada na Figura 4.12.

Verificou-se também, por meio de um estudo de caso, que a metodologia pode ser aplicada quando há alterações do campo de visão da câmera, desde que seja possível ajustar a perspectiva do vídeo com objetos e estruturas de referência na cena, bem como com sua perspectiva original.



Figura 4.11. Quadro de vídeo produzido pela câmera Cam10. No destaque, exemplo de como um alvo é visto no vídeo.



Figura 4.12. À esquerda, um quadro de vídeo produzido pela câmera Cam13. À direita, e um recorte ampliado do mesmo quadro. Observa-se a continuidade das faixas da escala no quadro.

5. CONCLUSÕES E TRABALHOS FUTUROS

Motivado pelo crescente uso de sistemas de CFTV – como forma de coibir, monitorar, reunir provas e identificar criminosos –, pelo aumento de solicitações de exames periciais de determinação de fonte a partir de vídeos capturados por esses sistemas e pelo fato de as características biométricas tradicionais (como as estruturas faciais) nem sempre estão disponíveis neste tipo de material, este trabalho tratou do problema da estimativa da altura humana em vídeos produzidos por câmeras de sistemas de vigilância, visando ao uso desta biometria branda em exames periciais, em conjunto ou em substituição às características biométricas tradicionais.

Como as soluções empregadas na Criminalística para estimar altura humana em vídeos são, em geral, restritas a técnicas de projeção reversa e de fotogrametria em apenas uma imagem, o presente trabalho apresentou uma metodologia semiautomática para se obter a altura humana em vídeos, mesmo que estes tenham sido capturados em condições inadequadas. Com a solução é possível estimar a altura de um alvo a partir de múltiplos quadros de um vídeo e não apenas de uma imagem estática.

No intuito de promover o uso da solução proposta, foi detalhada a fase de Aquisição de Dados – composta pela coleta de informações de câmeras de origem e de seus respectivos locais de instalação e pela produção de vídeos do padrão planar de calibração e da escala de referência – e implementada, em Python com OpenCV, a fase de Análise de Dados – constituída por cálculo dos parâmetros calibração de câmera (para compensar distorções nos vídeos), marcações manuais de coordenadas (do alvo e da escala) nos vídeos e estimativa semiautomática de altura em vídeos, com apresentação de resultados com incertezas de medição, bem como análises e gráficos estatísticos.

A metodologia foi submetida a experimentos em condições reais, com câmeras de diferentes marcas, modelos e configurações. Foram consideradas situações adversas, como: grande distância do alvo em relação à câmera, baixa qualidade dos vídeos (iluminação, contraste e/ou resolução espacial inadequados), pisos não planos ou irregulares e presença de ruído nos quadros dos vídeos. Os resultados de estimativa de alturas dos alvos obtidos nos experimentos apresentaram erro médio absoluto de 0,013 m em relação aos dados de validação, praticamente os mesmos apresentados por Li et al. (2015). Contudo, em Li et al. (2015), as condições de

validação foram muito limitadas se comparadas com os experimentos do presente trabalho, cujos resultados enaltecem o emprego da metodologia proposta em fins forenses.

Os experimentos evidenciaram também a efetividade do módulo de calibração de câmera proposto na metodologia, o qual apresentou resultados com erros de reprojeção inferiores a 0,5 pixel. Além disso, este módulo destaca-se pela simplicidade operacional, pois não há necessidade de intervenções manuais no processo, como ocorre em soluções como a de Bouguet (2004), que requer a indicação manual dos cantos extremos do padrão de calibração nas imagens de entrada.

Foi realizada também uma simulação com cinco cenários distintos para verificar o impacto das mudanças de perspectiva da cena capturada nos resultados de estimativa de altura. Verificou-se que, desde que seja possível ajustar a perspectiva de um vídeo, a partir de objetos e estruturas na cena e da perspectiva da cena original (no momento em que o alvo foi capturado), a metodologia pode ser aplicada mesmo havendo alterações do campo de visão da câmera.

Considerando a robustez e a confiabilidade evidenciadas nos resultados experimentais, a solução proposta por este trabalho foi aplicada em Laudo de Perícia Criminal (Gonçalves Júnior & Costa, 2017b), cuja natureza da perícia era Exame de Unicidade por Comparação Facial. No caso concreto, foi realizada uma estimativa de altura do autor do delito em estudo, com vistas a suprir a ausência de dados biométricos tradicionais no material (vídeo) analisado.

O estudo de caso da aplicação da metodologia no Laudo produzido por Gonçalves Júnior & Costa (2017b) promoveu o resumo denominado “A Estatura como Traço Biométrico Auxiliar na Identificação de Autores de Delito – Aplicação de Metodologia Forense de Estimativa de Altura Humana em Vídeos de Sistema de Vigilância”, aceito no XXIV Congresso Nacional de Criminalística, Florianópolis-SC, outubro de 2017.

Como contribuições à comunidade científica e à Criminalística, este trabalho produziu duas bases de dados de vídeos, que podem ser empregadas tanto para validar a metodologia proposta, quanto para fomentar futuras pesquisas nesta área, e dezoito módulos de códigos-fonte, totalizando, aproximadamente, três mil e oitocentas linhas de código, disponíveis para aplicação imediata da solução.

Pesquisas futuras incluem estudos mais detalhados sobre o impacto das mudanças de perspectiva da cena capturada, causadas pelas alterações do campo de visão das câmeras, e a aplicação da metodologia em locais dotados de câmeras do tipo PTZ (do inglês, *pan-tilt-zoom*).

REFERÊNCIAS BIBLIOGRÁFICAS

ANP, Academia Nacional de Polícia (2016). Curso de Fotogrametria Aplicada a Perícias em Imagens, de 07 a 11 de novembro de 2016. Diretoria de Gestão de Pessoal, Polícia Federal, Brasília-DF. Brasil.

Arruda, G. H. M., & Morrison, A. L. C. (2012). Exames de Reconhecimento Facial na Polícia Federal. *Anais da Conferência Internacional de Ciências Forenses em Multimídia e Segurança Eletrônica - ICMedia*, p. 78-85.

BenAbdelkader, C., & Yacoob, Y. (2008). Statistical Body Height Estimation from a Single Image. *IEEE 8th International Conference on Automatic Face and Gesture Recognition*, p. 1-7.

Bouguet, J.Y. (2004). “Camera Calibration Toolbox for Matlab®”. Disponível em: <http://www.vision.caltech.edu/bouguetj/calib_doc/index.html>, acesso em março de 2017.

Callari, F. (2010). How to Verify the Correctness of Calibration of a Webcam? StackOverflow. Disponível em: <<http://stackoverflow.com/questions/12794876/how-to-verify-the-correctness-of-calibration-of-a-webcam/12821056#12821056>>, acesso em fevereiro de 2017.

Clarke, T. A., & Fryer, J. G. (1998). The Development of Camera Calibration Methods and Models. *The Photogrammetric Record*, v. 16, n. 91, p. 51-66.

Criminisi, A., Zisserman, A., Van Gool, L. J., Bramble, S. K., & Compton, D. (1999). New Approach to Obtain Height Measurements from Video. In *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3576, p. 227-238.

Criminisi, A., Reid, I., & Zisserman, A. (2000). Single View Metrology. *International Journal of Computer Vision*, vol. 40, n. 2, p. 123-148.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 1, p. 886-893.

De Donno, A., Valeria, S., Bruno, M., Francesco, I., Fulvio, L., & Peter, L. (2012). A Photogrammetric Method for Stature Estimation: Reliability and Validity. *Research Journal of Applied Sciences, Engineering and Technology*, v. 4, n. 22, p. 4928-4933.

FISWG, Facial Identification Scientific Working Group (2012). Guidelines for Facial Comparison Methods, versão 1.0. Aprovada em 02/02/2012 e publicada em 20/04/2012. Disponível em <<https://fiswg.org/documents.html>>, acesso em fevereiro de 2017.

Gonçalves Júnior, A., & Costa, N. R. S. (2017). Laudo de Perícia Criminal nº 2.605/2017 – Exame de Unicidade por Comparação Facial. Seção de Perícias de Biometria Forense e Audiovisuais, Instituto de Criminalística, Departamento de Polícia Técnica, Polícia Civil do Distrito Federal, Brasília-DF.

Gonçalves Júnior, A., & Costa, N. R. S. (2017). Laudo de Perícia Criminal nº 7.972/2017 – Exame de Unicidade por Comparação Facial. Seção de Perícias de Biometria Forense e Audiovisuais, Instituto de Criminalística, Departamento de Polícia Técnica, Polícia Civil do Distrito Federal, Brasília-DF. Brasil.

Hartley, R., & Zisserman, A. (2003). Multiple View Geometry in Computer Vision. *Cambridge University Press*, ed. 2, cap. 6, p. 153-177.

Heikkila, J., & Silven, O. (1997). A Four-step Camera Calibration Procedure with Implicit Image Correction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 1106-1112.

HOSDB, Home Office Scientific Development Branch (2007). Video Processing and Analysis Manual, cap. 4. UK Government. Disponível em: <http://tna.europarchive.org/20100413151426/http://scienceandresearch.homeoffice.gov.uk/hosdb/publications/cctv-publications/VP_A_Manual_-_Analysis_-_Si19056.html?view=Standard&pubID=512762>, acesso em fevereiro de 2017.

Jain, A. K., Nandakumar, K., Lu, X., & Park, U. (2004). Integrating Faces, Fingerprints, and Soft Biometric Traits for User Recognition. In: *ECCV International Workshop on Biometric Authentication*, p. 259-269.

Jain, A. K., Ross A., & Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 14, n. 1, p. 4-20.

Jain, A. K., Bolle, R., & Pankanti, S. (Ed.). (2006). Biometrics: Personal Identification in Networked Society. *Springer Science & Business Media*, p.16.

Jain, A. K., Flynn, P., & Ross, A. (Ed.). (2007). Handbook of Biometrics. *Springer Science & Business Media*, p. 1-23.

JCGM, Joint Committee for Guides in Metrology (2008). Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement – JCGM 100:2008. Disponível em <<http://www.iso.org/sites/JCGM/GUM-JCGM100.htm>>, acesso em março de 2017.

Jeges, E., Kispal, I., & Hornak, Z. (2008). Measuring Human Height Using Calibrated Cameras. *IEEE Conference on Human System Interactions*, p. 755-760.

Jung, J., Kim, H., Yoon, I., & Paik, J. (2016). Human Height Analysis Using Multiple Uncalibrated Cameras. In *IEEE International Conference on Consumer Electronics*, p. 213-214.

Jung, J., Yoon, I., Lee, S., & Paik, J. (2016). Object Detection and Tracking-Based Camera Calibration for Normalized Human Height Estimation. *Journal of Sensors*, v. 2016.

Lee, J., Lee, E. D., Tark, H. O., Hwang, J. W., & Yoon, D. Y. (2008). Efficient Height Measurement Method of Surveillance Camera Image. *Forensic Science International*, v. 177, n. 1, p. 17-23.

Lee, K. Z. (2012). A Simple Calibration Approach to Single View Height Estimation. In *IEEE Ninth Conference on Computer and Robot Vision*, p. 161-166.

Lee, S. H., & Choi, J. S. (2010). A Single-View Based Framework for Robust Estimation of Heights and Positions of Moving People. In *IEEE Digest of Technical Papers International Conference on Consumer Electronics*, p. 503-504.

Li, S., Nguyen, V. H., Ma, M., Jin, C. B., Do, T. D., & Kim, H. (2015). A Simplified Nonlinear Regression Method for Human Height Estimation in Video Surveillance. *EURASIP Journal on Image and Video Processing*, v. 2015, n. 1, p. 32.

Madden, C. S., & Piccardi, M. (2005). Height Measurement as a Session-Based Biometric for People Matching Across Disjoint Camera Views. *IEEE Conference on Image and Vision Computing*, p. 282-286.

MathWorks, The MathWorks, Inc ©, Documentation. (2017). Single Camera Calibration App. Disponível em: <<https://www.mathworks.com/help/vision/ug/single-camera-calibrator-app.html>>, acesso em julho de 2017.

MathWorks, The MathWorks, Inc ©, Documentation. (2017). What Is Camera Calibration? Disponível em: <<https://www.mathworks.com/help/vision/ug/camera-calibration.html>>, acesso em fevereiro de 2017.

Meuwly, D. (2006). Forensic Individualization from Biometric Data. *Science & Justice*, v. 46, n. 4, p. 205-213.

Microsoft © (2016). Usar a Lupa para Ver Itens na Tela. Aplicável aos sistemas operacionais Windows 10, Windows 8.1 e Windows 7. Artigo 11542, rev. 3. Disponível em: <<https://support.microsoft.com/pt-br/help/11542/windows-use-magnifier>>, acesso em março de 2017.

Momeni-K, M., Diamantas, S. C., Ruggiero, F., & Siciliano, B. (2012). Height Estimation from a Single Camera View. *International Conference on Computer Vision Theory and Applications*, p. 358-364.

Moon, H. M., Chae, S. H., Moon, D., Chung, Y., & Pan, S. B. (2013). Intelligent Video Surveillance System Using Two-factor *Human Information*. *Telecommunication Systems*, v. 52, p. 2249-2257.

Morvan, Y. Y. (2009). Acquisition, Compression and Rendering of Depth and Texture for Multi-view Vídeo. Tese de Doutorado, Universidade de Tecnologia de Eindhoven, Eindhoven, Holanda.

Nguyen, N. H., & Hartley, R. (2012). Height Measurement for Humans in Motion Using a Camera: a Comparison of Different Methods. In *IEEE International Conference on Digital Image Computing Techniques and Applications*, p. 1-8.

OpenCV, Open Source Computer Vision Library. (2017). About. Developers Team, Itseez. Disponível em <<http://opencv.org/about.html>>, acesso em fevereiro de 2017.

OpenCV, Open Source Computer Vision Library (2017). Camera Calibration and 3D Reconstruction. Developers Team, Itseez. Disponível em <http://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html>, acesso em fevereiro de 2017.

OpenCV, Open Source Computer Vision Library. (2017). Camera Calibration with OpenCV. Developers Team, Itseez. Disponível em <http://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration/camera_calibration.html>, acesso em março de 2017.

Python. About. (2017). Python Software Foundation. Disponível em <<https://www.python.org/about/>>, acesso em fevereiro de 2017.

Ramstrand, N., Ramstrand, S., Brolund, P., Norell, K., & Bergström, P. (2011). Relative Effects of Posture and Activity on Human Height Estimation from Surveillance Footage. *Elsevier Forensic Science International*, v. 212, n. 1, p. 27-31.

Shao, J., Zhou, S. K., & Chellappa, R. (2010). Robust Height Estimation of Moving Objects from Uncalibrated Videos. *IEEE Transactions on Image Processing*, v. 19, n. 8, p. 2221-2232.

Silva Filho, L. O., & Melo, D. F. O. (2015). IV Curso de Introdução às Técnicas de Fotogrametria em Perícias de Local, Notas de Aula. Instituto Nacional de Criminalística, Polícia Federal, Brasília-DF, p. 31. Brasil.

Tsai, R. (1987). A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-The-Shelf TV Cameras and Lenses. *IEEE Journal on Robotics and Automation*, v. 3, n. 4, p. 323-344.

Valente, C. R. (2016). Curso de Inferência Lógica na Criminalística, Notas de Aula. Instituto de Criminalística, Departamento de Polícia Técnica, Polícia Civil do Distrito Federal, Brasília-DF, p.54. Brasil.

Valente, C. R. (2012). Perspectivas da Fonética Forense num Cenário de Quebra do Dogma da Unicidade. *Anais da Conferência Internacional de Ciências Forenses em Multimídia e Segurança Eletrônica - ICMedia*, p.7-27.

Wang, C. M., & Chen, W. Y. (2015). The Human-Height Measurement Scheme by Using Image Processing Techniques. *International Journal of Computer, Consumer and Control*, v. 4, n. 3, p. 42-50.

Wang, J., Shi, F., Zhang, J., & Liu, Y. (2008). A New Calibration Model of Camera Lens Distortion. *Elsevier Pattern Recognition*, v. 41, n. 2, p. 607-615.

Wang, Y. M., Li, Y., & Zheng, J. B. (2010). A Camera Calibration Technique Based on OpenCV. In *3rd IEEE International Conference on Information Sciences and Interaction Sciences*, p. 403-406.

Whittle, M. W. (2007). *Gait Analysis: An Introduction*. Elsevier Ltd., ed. 4, p. 52-80.

Zhang, Z. (1998). A Flexible New Technique for Camera Calibration. *Technical Report MSR-TR-98-71, Microsoft Research*. Disponível em <<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr98-71.pdf>>, acesso em fevereiro de 2017.

Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, v. 22, n. 11, p. 1330-1334.