



**UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"**

Henrique Leal Tavares

**Identificação de Pessoas Baseada em
Características Antropométricas e de Marcha
Extraídas de Poses 2D**

Bauru
2021

Henrique Leal Tavares

Identificação de Pessoas Baseada em Características Antropométricas e de Marcha Extraídas de Poses 2D

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Estadual Paulista "Júlio de Mesquita Filho".

Financiadora: PETROBRAS

Orientador: Prof. Associado Aparecido Nilceu Marana

Coorientador: Prof. Dr. João Baptista Cardia Neto

Bauru
2021

Leal Tavares, Henrique.

Identificação de Pessoas Baseada em Características Antropométricas e de Marcha Extraídas de Poses 2D / Henrique Leal Tavares. Bauru, 2021

80 f. : il.,

Orientador: Aparecido Nilceu Marana

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho". Faculdade de Ciências, Bauru, 2021.

1. Identificação de Pessoas. 2. Biometria. 3. Poses 2D. 4. Mapas de Calor das Articulações 5. Medidas Antropométricas 6. Reconhecimento de Marcha I. Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências. II. Título.

ATA DA DEFESA EM SESSÃO FECHADA DA DISSERTAÇÃO DE Mestrado DE HENRIQUE LEAL TAVARES, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.

Aos 30 dias do mês de julho do ano de 2021, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE Mestrado de HENRIQUE LEAL TAVARES, intitulada **Identificação de Pessoas Baseada em Características Antropométricas e de Marcha Extraídas de Poses 2D**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Associado APARECIDO NILCEU MARANA (Orientador - Participação Virtual) do Departamento de Computação/UNESP/Campus de Bauru, Prof. Associado ANTONIO CARLOS SEMENTILLE (Participação Virtual) do Departamento de Computação/UNESP/Campus de Bauru, Profa. Dra. ALESSANDRA APARECIDA PAULINO (Participação Virtual) da Faculdade de Computação / Universidade Federal de Uberlândia. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final **APROVADO**. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo Presidente da Comissão Examinadora.



Prof. Associado APARECIDO NILCEU MARANA

Comissão Examinadora

Prof. Associado Aparecido Nilceu Marana

UNESP - Câmpus de Bauru

Orientador

Profa. Dra. Alessandra Aparecida Paulino

Universidade Federal de Uberlândia

Prof. Associado Antonio Carlos Sementille

UNESP - Câmpus de Bauru

Agradecimentos

Agradeço ao Prof. Associado Aparecido Nilceu Marana, pela orientação deste trabalho.

Agradeço ao Prof. Dr. João Baptista Cardia Neto, pela coorientação deste trabalho.

Agradeço ao Daniel Ricardo dos Santos Jangua, pela colaboração em um dos experimentos.

Agradeço ao Laboratório Recogna, da Faculdade de Ciências, UNESP, Bauru, pela infraestrutura fornecida para a realização deste trabalho.

Agradeço ao Programa de Pós-Graduação em Ciência da Computação da UNESP, pela oportunidade de realizar esta dissertação de mestrado.

Agradeço à Faculdade de Ciências de Bauru e à UNESP, pela infraestrutura e pelo suporte em minha especialização acadêmica.

Agradeço à PETROBRAS, pelo financiamento deste projeto (Processo Fundunesp no. 2662/2017) e pela bolsa concedida.

Resumo

De acordo com evidências atuais, em determinados cenários e ambientes, as características biométricas tradicionais, tais como impressões digitais, íris e faces, podem ter suas eficácias diminuídas na identificação de pessoas devido às condições heterogêneas ou precárias de iluminação dos ambientes, à distância das pessoas aos sensores de captura, ao uso de sensores de baixa resolução, à necessidade de cooperação das pessoas com o processo de identificação, dentre outros. Tendo em vista esta problemática, este trabalho propõe um método de identificação de pessoas baseado na fusão de características antropométricas e de marcha, classificadas como características biométricas *soft* e comportamentais, respectivamente, como uma forma de eludir-se destes obstáculos. O método proposto baseia-se em medidas de partes do esqueleto humano, como braços e pernas, e em mapas térmicos das articulações dos esqueletos, obtidos de poses 2D estimadas nas imagens por meio de algoritmos estado da arte propostos para esta finalidade. O método proposto foi avaliado na CASIA Gait Database-A e CASIA Gait Database-B, duas bases de dados de acesso público comumente utilizadas em pesquisas sobre identificação biométrica baseada em reconhecimento de padrões de marcha humana, e também em uma base de dados privada. Os resultados experimentais obtidos mostram que as características biométricas propostas, obtidas de poses 2D, podem ser utilizadas com êxito para a identificação humana, alcançando taxas de acurácia similares ou superiores às obtidas por métodos estado da arte.

Palavras-chave: Identificação de Pessoas, Biometria, Poses 2D, Mapas de Calor das Articulações, Medidas Antropométricas, Reconhecimento de Marcha.

Abstract

According to current evidence, in certain scenarios and environments, traditional biometric characteristics, such as fingerprints, irises and faces, may have their effectiveness reduced in identifying people due to heterogeneous or poor lighting conditions of environments, distance from people to capture sensors, use of low-resolution sensors, need for people to cooperate with the identification process, among others. In view of this issue, this work proposes a method of identifying people based on biometric characteristics classified as *soft biometrics*, such as anthropometric measures, and behavioral measures, such as gait, as a way to overcome such obstacles. The proposed method is based on measurements of parts of the human skeleton, such as arms and legs, and on thermal maps of skeletal joints, obtained from 2D poses estimated in the images using state-of-the-art algorithms proposed for this purpose. The proposed method was evaluated in the CASIA Gait Database-A and CASIA Gait Database-B, two publicly available databases commonly used in research on biometric identification based on recognition of human gait, and also in a private dataset. The results obtained show that the proposed biometric characteristics, obtained from 2D poses, can be successfully used for human identification, reaching accuracy rates similar or superior to those obtained by state-of-the-art methods.

Keywords: People Identification, Biometric, Soft Biometrics, 2D Poses, Joints Heat-maps, Anthropometric Measurements, Gait Recognition.

Lista de ilustrações

Figura 1 – Exemplo de modelo de pose humana 2D que representa as partes principais do esqueleto humano. Os rótulos referem-se às 12 principais articulações do esqueleto.	18
Figura 2 – Exemplo de pose humana 2D estimada pelo método PifPaf (KREISS; BERTONI; ALAHI, 2019a).	19
Figura 3 – Exemplos de características biométricas. a) DNA; b) Orelha; c) Face; d) Termograma Facial; e) Termograma da Mão; f) Veias da Mão; g) Impressão Digital; h) Caminhar; i) Geometria da Mão; j) Íris; k) Impressão da Palma; l) Retina; m) Assinatura; n) Voz	25
Figura 4 – Exemplos de características biométricas da categoria <i>soft biometrics</i>	27
Figura 5 – Representação dos períodos de apoio e de balanço no ciclo da marcha.	28
Figura 6 – Sistema biométrico bimodal mostrando os três níveis de fusão: características, pontuação e decisão.	31
Figura 7 – Estimativas de poses 2D utilizando o método <i>PifPaf</i> (KREISS; BERTONI; ALAHI, 2019b).	33
Figura 8 – Estimativas de poses 2D utilizando o método <i>OpenPose</i> (CAO et al., 2018a).	33
Figura 9 – Diagrama do método <i>PifPaf</i>	35
Figura 10 – Componentes PIF para o ombro esquerdo.	35
Figura 11 – PAF que associa ombros esquerdos com respectivos quadris esquerdos das pessoas presentes na imagem.	35
Figura 12 – Diagrama do pipeline do método <i>OpenPose</i>	36
Figura 13 – Pipeline geral do método <i>OpenPose</i> . (a) Imagem de entrada; (b) Mapas de confiança para detecção das articulações; (c) PAFs para associação parcial; (d) Conjunto de correspondência para associar as articulações que formam os membros; (e) As poses 2D de corpo inteiro construídas para todas as pessoas encontradas na imagem.	37
Figura 14 – Taxas de erro do sistema biométrico: FMR (Taxa de correspondência falsa) e FNMR (Taxa de não correspondência falsa)	38
Figura 15 – Um exemplo de curva CMC, em que no eixo das abscissas tem-se as posições do ranque e no eixo das ordenadas tem-se a acurácia obtida pelo classificador.	39
Figura 16 – Diagrama de blocos do método proposto para identificação de pessoas baseada em medidas antropométricas (características <i>Soft Biometrics</i>) e características de marcha, obtidas de poses 2D.	47
Figura 17 – Exemplo da aplicação do pré-processamento em um <i>frame</i> do vídeo de entrada.	48
Figura 18 – Exemplo de pose 2D estimada a partir de uma imagem pré-processada.	49

Figura 19 – Medidas antropométricas obtidas das poses 2D dos indivíduos, utilizadas pelo método proposto.	50
Figura 20 – Vetor de características composto pela média dos comprimentos de cada parte do esqueleto humano: Úmero Direito (UmD), Úmero Esquerdo (UmE), Ulna Direita (UnD), Ulna Esquerda (UnE), Fêmur Direito (FD), Fêmur Esquerdo (FE), Tíbia Direita (TD) e Tíbia Esquerda (TE), em cada um dos n quadros do vídeo.	51
Figura 21 – Vetor de características composto pela concatenação dos vetores contendo os comprimentos de cada parte do esqueleto humano: Úmero Direito (UmD), Úmero Esquerdo (UmE), Ulna Direita (UnD), Ulna Esquerda (UnE), Fêmur Direito (FD), Fêmur Esquerdo (FE), Tíbia Direita (TD) e Tíbia Esquerda (TE), em cada um dos n quadros do vídeo.	52
Figura 22 – Vetor de características composto pela concatenação dos histogramas dos comprimentos das partes do esqueleto humano: Úmero Direito (UmD), Úmero Esquerdo (UmE), Ulna Direita (UnD), Ulna Esquerda (UnE), Fêmur Direito (FD), Fêmur Esquerdo (FE), Tíbia Direita (TD) e Tíbia Esquerda (TE).	53
Figura 23 – Diagrama do método proposto para a extração das características de marcha humana baseado nos mapas de calor das articulações gerados na etapa de estimação de poses 2D. O mapa de calor médio é utilizado como o descritor de marcha do vídeo.	53
Figura 24 – Imagens do banco de dados CASIA Gait Dataset-A de pessoas caminhando nas direções paralela (imagem da esquerda), perpendicular (imagem central) e diagonal (imagem da direita) com relação ao plano da câmera.	56
Figura 25 – Exemplos de imagens do banco de dados CASIA Gait Dataset-B; (a) Múltiplas visões da marcha, capturadas por 11 câmeras; (b) Marcha normal do indivíduo; (c) Marcha com o indivíduo vestindo casaco; (d) Marcha com o indivíduo carregando uma mochila.	57
Figura 26 – Quadro de um dos vídeos da base de dados privada, disponibilizada pela Petrobras, na qual é possível observar funcionários em sua rotina de trabalho. As características do ambiente, as oclusões, a distância da câmera e a utilização de equipamentos de proteção pelos funcionários dificultam sobremaneira o processo de identificação.	59
Figura 27 – Exemplos da aplicação do algoritmo Deep SORT no processo de rastreamento e anotação de indivíduos nos vídeos da base de dados privada. . . .	59

Figura 28 – Curvas CMC obtidas utilizando-se os descritores antropométricos (estratégia da concatenação) e os descritores de marcha obtidos de poses 2D estimadas por três métodos de estimativa de pose: PifPaf, OpenPose e Tf-Pose-Estimation, no subconjunto de dados CASIA Gait Dataset-B com vista lateral - 90°, utilizando-se um classificador RNA para os descritores de marcha e o classificador 1-NN, com as distâncias euclidiana e <i>city-block</i> , para os descritores antropométricos.	62
Figura 29 – Curvas CMC obtidas utilizando-se os descritores antropométricos (estratégia da média) e os descritores de marcha obtidos de poses 2D estimadas por três métodos de estimativa de pose: PifPaf, OpenPose e TF-Pose-Estimation, no subconjunto de dados CASIA Gait Dataset-B com vista lateral - 90°, utilizando-se um classificador RNA para os descritores de marcha e o classificador 1-NN, com as distâncias euclidiana e <i>city-block</i> , para os descritores antropométricos.	63
Figura 30 – Curvas CMC obtidas na base dados CASIA Gait Dataset-A, utilizando-se os métodos PifPaf e OpenPose, com os descritores antropométricos (estratégia de concatenação) e o classificador 1-NN, com as funções de distância euclidiana e <i>city block</i>	64
Figura 31 – Métodos de estimação de pose PifPaf e OpenPose aplicados em um mesmo <i>frame</i> de um vídeo da base de dados privada. Pode-se observar que dos cinco indivíduos na cena, o PifPaf (a) detectou quatro deles, enquanto que o OpenPose (b) detectou apenas dois, sendo que os membros inferiores de um dos indivíduos foram estimados erroneamente.	65
Figura 32 – Exemplo de uma imagem em que o método PifPaf detectou e estimou corretamente a pose 2D de um trabalhador (a), enquanto que o método OpenPose falhou (b).	66
Figura 33 – Curvas CMC obtidas utilizando-se os descritores antropométricos obtidos de poses 2D estimadas pelos métodos PifPaf e OpenPose, no subconjunto de dados CASIA Gait Dataset-B com vista lateral - 90°, utilizando-se o classificador 1-NN, com as distâncias euclidiana e <i>city-block</i> , sem aplicar os preprocessamentos.	67
Figura 34 – Curvas CMC obtidas usando o descritor de marcha baseado nos mapas de calor das articulações de poses estimadas pelos métodos PifPaf, OpenPose e TF-Pose-Estimation, para um conjunto de 10 indivíduos da base de dados CASIA Gait Dataset B.	68
Figura 35 – Curvas CMC obtidas com os métodos baseados nos descritores antropométricos (estratégia da concatenação) e de marcha (média dos mapas de calor) individualmente e também aplicando-se a fusão de ambos, com pesos de 70% para os descritores antropométricos e 30% para os descritores de marcha.	70

Figura 36 – Curvas CMC obtidas utilizando-se os descritores antropométricos e os descritores de marcha obtidos de poses 2D estimadas pelo método PifPaf, na base de dados privada, utilizando-se um classificador RNA para os descritores de marcha e o classificador 1-NN, com a distância euclidiana para os descritores antropométricos. 73

Lista de tabelas

Tabela 1 – Comparação entre características biométricas. As pontuações alta, média e baixa são denotadas por A, M e B, respectivamente.	25
Tabela 2 – Propriedades de algumas características biométricas da categoria <i>soft biometrics</i>	27
Tabela 3 – Exemplo de informações geradas por algoritmos de estimativa de pose. . .	34
Tabela 4 – Valores da AUC referentes às curvas CMC apresentadas na Figura 28. . . .	63
Tabela 5 – Medidas de precisão, revocação e F1-score utilizando-se os descritores de marcha extraídos dos mapas de calor das juntas das poses estimadas pelo método PifPaf em um conjunto de 10 indivíduos da base de dados CASIA Gait Dataset-B.	68
Tabela 6 – Medidas de Precisão, Revocação e F1-Score obtidas na base de dados CASIA Gait Dataset-B utilizando-se os descritores de marcha extraídos dos mapas de calor das juntas das poses estimadas pelo método PifPaf para todos os indivíduos da base.	69
Tabela 7 – Acurácia Rank-1 - CASIA Gait Dataset-A.	71
Tabela 8 – Acurácia Rank-1 - CASIA Gait Dataset-B.	72

Lista de abreviaturas e siglas

ANN	- <i>Artificial Neural Network</i>
AUC	- <i>Area Under the Curve</i>
CFTV	- <i>Circuito Fechado de Televisão</i>
CMC	- <i>Cumulative Match Curve</i>
CMDA	- <i>Coupled Multi-Linear Discriminant Analysis</i>
CMLP	- <i>Coupled Multi-Linear Locality-Preserved</i>
CMMF	- <i>Coupled Multi-Linear Marginal</i>
CNN	- <i>Convolutional Neural Network</i>
CPA	- <i>Couple Patch Alignment</i>
EER	- <i>Equal Error Rate</i>
EPI	- <i>Equipamento de Proteção Individual</i>
FAR	- <i>False Acceptance Rate</i>
FN	- <i>False Negative</i>
FNMR	- <i>False Non-Match Rate</i>
FNR	- <i>False Negative Rate</i>
FP	- <i>False Positive</i>
FRR	- <i>False Rejection Rate</i>
GAN	- <i>Generative Adversarial Networks</i>
IA	- <i>Inteligencia Artificial</i>
KNN	- <i>K-nearest Neighbors Algorithm</i>
NN	- <i>Nearest Neighbors</i>
ONG	- <i>Organização Não Governamental</i>
PAF	- <i>Part Affinity Fields</i>

PIF - *Part Intensity Field*

RNA - *Rede Neural Artificial*

ROC - *Receiver Operating Characteristic*

SORT - *Simple Online and Realtime Tracking*

SVM - *Support Vector Machine*

TN - *True Negative*

TP - *True Positive*

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	19
1.2	Objetivos	20
1.2.1	Objetivo Geral	20
1.2.2	Objetivos Específicos	20
1.3	Estrutura da Dissertação	20
2	FUNDAMENTAÇÃO TEÓRICA	22
2.1	Biometria	22
2.1.1	Sistemas Biométricos	23
2.1.2	Requisitos das Características Biométricas	24
2.2	<i>Soft Biometrics</i>	25
2.3	Reconhecimento da Marcha Humana	26
2.3.1	Ciclo da Marcha	28
2.3.2	A Marcha Humana no Contexto da Biometria	29
2.4	Fusão de Características Biométricas	29
2.4.1	Fusão de Características <i>Soft Biometrics</i>	31
2.5	Estimativa de Pose Humana 2D	32
2.5.1	PifPaf	33
2.5.2	OpenPose	34
2.5.3	Tf-Pose-Estimation	36
2.6	Métricas	37
2.6.1	Curvas ROC e CMC	37
2.6.2	Precisão, Revocação e Medida F1	40
2.7	Funções de Distância	40
2.8	Classificadores	41
2.8.1	K-Vizinhos Mais Próximos - KNN	41
2.8.2	Rede Neural Artificial - RNA	41
2.8.2.1	MLP - <i>Multilayer Perceptron</i>	42
3	TRABALHOS CORRELATOS	43
3.1	Métodos Baseados em Medidas Antropométricas	43
3.1.1	Método Baseado em Poses	43
3.1.2	Métodos Não Baseados em Poses	43
3.2	Métodos Baseados em Marcha Humana	44
3.2.1	Métodos Baseados em Poses	44

3.2.2	Métodos Não Baseados em Poses	45
3.3	Métodos Baseados na Fusão de Características Antropométricas e de Marcha	45
4	MÉTODO PROPOSTO	47
4.1	Pré-Processamento	47
4.2	Estimativas de Pose	48
4.3	Extração de Características	49
4.3.1	Medidas Antropométricas	49
4.3.1.1	Média das Medidas Antropométricas	50
4.3.1.2	Concatenação das Medidas Antropométricas	50
4.3.1.3	Histogramas das Medidas Antropométricas	51
4.3.2	Características de Marcha	51
4.4	Reconhecimento	53
4.4.1	Medidas Antropométricas	54
4.4.1.1	Média das Medidas Antropométricas	54
4.4.1.2	Concatenação das Medidas Antropométricas	54
4.4.1.3	Histogramas das Medidas Antropométricas	55
4.4.2	Características de Marcha	55
5	BASES DE DADOS	56
5.1	CASIA Gait Dataset-A	56
5.2	CASIA Gait Dataset-B	56
5.3	Base de Dados Privada	58
5.3.1	Deep SORT	60
6	RESULTADOS EXPERIMENTAIS	61
6.1	Avaliação dos Métodos de Estimação de Pose	61
6.2	Avaliação dos Preprocessamentos	66
6.3	Avaliação do Impacto do Número de Indivíduos	66
6.4	Avaliação da Fusão no Nível de Pontuação	69
6.5	Avaliação da Fusão no Nível de Características	69
6.6	Avaliação em uma Aplicação Real	71
7	CONCLUSÃO E TRABALHOS FUTUROS	74
7.1	Trabalhos Futuros	75
7.2	Publicações Realizadas	75
	REFERÊNCIAS	76

1 Introdução

Biometria é a ciência que estabelece a identidade de uma pessoa com base em seus atributos físicos ou comportamentais (??). Como os métodos biométricos são mais confiáveis e difíceis de serem fraudados do que os métodos tradicionais de identificação baseados em posse (cartões, documentos, etc.) e em conhecimento (senhas, códigos, etc.), eles têm se tornado cada vez mais usados para identificação humana em diferentes aplicações.

Um desafio importante na área de Biometria é a identificação automática realizada em tempo real, em ambientes não controlados, por meio de câmeras de baixa resolução, como câmeras de CFTV (Circuito Fechado de TV), instaladas à distância em posições nem sempre favoráveis para a utilização das técnicas biométricas convencionais. Nesses tipos de cenários, o uso de características biométricas tradicionais, como íris, impressão digital ou face, classificadas como *hard biometrics*, pode ser inviável. Nesses casos, o uso de características biométricas classificadas como *soft biometrics* pode ser uma melhor opção ou, às vezes, a única opção.

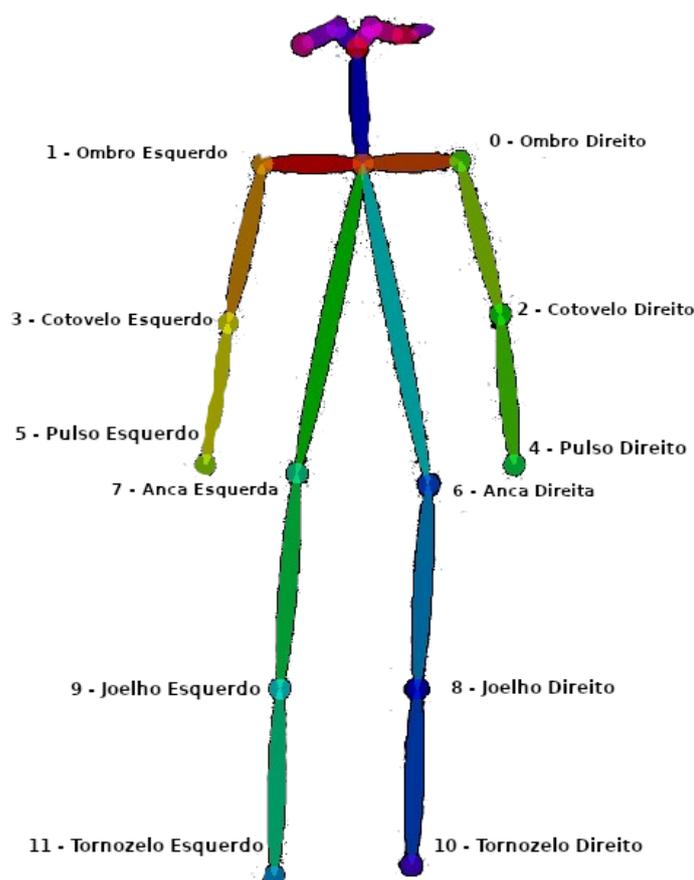
Soft biometrics consiste em um conjunto de características, tais como altura, peso, gênero, cor da pele, cor do cabelo, cor dos olhos, dentre outras. Essas características fornecem alguma informação sobre o indivíduo, mas não apresentam capacidades de distinção e permanência suficientes para diferenciar quaisquer dois indivíduos (JAIN; ROSS; NANDAKUMAR, 2011). Embora tais características não permitam a identificação unívoca das pessoas quando utilizadas individualmente (TOME et al., 2014), elas podem ser muito eficazes na identificação de pessoas quando utilizadas em conjunto. Além disso, elas podem contribuir sobremaneira no processo de identificação quando combinadas com outras características biométricas, como, por exemplo, o padrão da marcha.

O reconhecimento do padrão da marcha humana, conhecido como *gait recognition*, é uma tarefa complexa e desafiadora que ganhou interesse especialmente porque requer pouca interação do usuário e, normalmente, as imagens necessárias para esse tipo de identificação podem ser de baixa resolução. Por outro lado, a identificação da marcha depende do movimento, o que constitui uma limitação importante (BEN et al., 2019a; CHAO et al., 2018; TIAN et al., 2019; TONG; FU; LING, 2019; WANG et al., 2019). Além disso, por ser uma característica comportamental, o padrão de marcha pode ser facilmente imitado ou alterado por indivíduos que desejam burlar o sistema biométrico.

Características biométricas, como o padrão da marcha, podem ser obtidas de poses humanas 2D, estimadas em imagens ou vídeos (LIMA; SCHWARTZ, 2019; JANGUA; MARANA, 2020). A detecção de pose humana 2D é definida como o problema de localização das articulações do corpo humano (cotovelos, ombros, quadris, pulsos, joelhos, tornozelos, etc.). A Figura 1 mostra um modelo de pose humana 2D que pode ser estimado de imagens ou quadros

de vídeos. Este modelo representa as partes principais do esqueleto humano.

Figura 1 – Exemplo de modelo de pose humana 2D que representa as partes principais do esqueleto humano. Os rótulos referem-se às 12 principais articulações do esqueleto.



Fonte: Elaborada pelo Autor.

Recentemente, foram propostos dois métodos muito eficazes e eficientes para estimativa de pose humana 2D a partir de imagens de vídeos: OpenPose (CAO et al., 2018b) e PifPaf (KREISS; BERTONI; ALAHI, 2019a). O método OpenPose utiliza *Part Affinity Fields* (PAFs) para aprender como associar partes do corpo a indivíduos que são detectados em uma imagem. O método PifPaf (KREISS; BERTONI; ALAHI, 2019a) utiliza o campo de *Part Intensity Field* (PIF) para detectar a parte do corpo de um indivíduo e um campo de *Part Affinity Fields* (PAF) para associar partes do corpo e construir um corpo humano completo. A Figura 2 mostra um exemplo de uma pose humana 2D estimada pelo método PifPaf. Pode-se observar que o método foi capaz de detectar todos os 12 pontos articulares do esqueleto humano e, conseqüentemente, suas partes correspondentes, como braços, pernas e tronco. Ambos os métodos, OpenPose e PifPaf, são capazes de estimar em tempo real poses humanas 2D a partir de vídeo e, portanto, podem ser usados para a identificação biométrica de pessoas em

tempo real, por meio do padrão da marcha, de medidas antropométricas (um exemplo de *soft biometrics*), ou pela combinação de ambos.

Figura 2 – Exemplo de pose humana 2D estimada pelo método PifPaf (KREISS; BERTONI; ALAHI, 2019a).



Fonte: Elaborada pelo Autor.

1.1 Motivação

A necessidade de implementação de sistemas biométricos em que haja pouca ou nenhuma cooperação dos usuários, que devam ser implantados em ambientes pouco controlados, e cuja identificação dos indivíduos deva ser realizada por meio de câmeras de baixo custo e, portanto, com imagens de média ou baixa resolução, impulsiona a busca por novos métodos de reconhecimento.

O uso de *soft biometrics* é uma abordagem relativamente nova quando se considera os sistemas automáticos de identificação biométrica e que está em constante estudo e desenvolvimento. A motivação para isto deve-se a alguns aspectos, tais como:

- A identificação pode ser realizada sem a colaboração do usuário;
- Variações ou deficiências na iluminação, presença de ruídos e uso de sensores de baixa resolução não afetam tanto o desempenho em comparação a outros tipos de biometrias;
- Os indivíduos podem se posicionar diante do sensor de uma forma mais natural, sem a necessidade de estar olhando para uma câmera ou tocando nos sensores;

- Em um sistema multi-biométrico, pode-se utilizar os mesmos sensores sendo utilizados para se obter as características biométricas primárias, aumentando as taxas de acurácia da identificação das pessoas sem aumentar os custos.

Atualmente, não existem muitos trabalhos na literatura que utilizam a estimativa de pose humana como forma de extrair características biométricas para a identificação de pessoas. Com a escassez de trabalhos nesta área e os ótimos resultados apresentados pelos algoritmos de estimativa de pose, surgiu o incentivo para a realização deste trabalho de dissertação de mestrado.

1.2 Objetivos

Nesta seção serão apresentados os objetivos gerais e específicos deste trabalho.

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é propor um método de identificação de pessoas baseado em características antropométricas (classificadas como *soft biometrics*) e características de marcha obtidas de poses 2D estimadas de videos por meio de algoritmos estado da arte.

1.2.2 Objetivos Específicos

- Avaliar e comparar os algoritmos do estado da arte referentes à estimativa de pose 2D;
- Utilizar as poses 2D para extrair características antropométricas, classificadas como *soft biometrics*;
- Utilizar as poses 2D para extrair características de marcha humana;
- Avaliar as melhores técnicas de fusão para serem utilizadas na integração das características antropométricas e de marcha em um único método multibiométrico de identificação humana.

1.3 Estrutura da Dissertação

Esta dissertação está organizada nos seguintes capítulos.

No Capítulo 1 apresentam-se a introdução ao tema desta dissertação de mestrado e os seus objetivos.

No Capítulo 2 apresenta-se a fundamentação teórica do trabalho, com ênfase nos conceitos relacionados à Biometria, estimativa de poses humanas e análise de padrões de marcha humana.

No Capítulo 3 apresenta-se uma revisão da literatura sobre métodos estado da arte que são correlatos ao método proposto neste trabalho.

No Capítulo 4 apresenta-se o método proposto para se promover o reconhecimento de pessoas utilizando-se características *soft*.

No Capítulo 5 apresentam-se as bases de dados utilizadas nos experimentos para avaliação do método proposto neste trabalho.

No Capítulo 6 apresentam-se os resultados obtidos nos experimentos.

No Capítulo 7 apresentam-se as conclusões, os trabalhos futuros e as publicações derivadas deste trabalho.

2 Fundamentação Teórica

Neste capítulo apresentam-se inicialmente os aspectos e métodos relacionados à biometria, analisando-se seus pontos positivos em relação às abordagens tradicionais de identificação humana, com ênfase em *soft biometrics* como uma alternativa para melhorar a acurácia da biometria.

Após, apresenta-se o problema de estimativa de pose humanas 2D, analisando-se as possíveis dificuldades com essa abordagem.

Por fim, apresentam-se os conceitos relacionados ao reconhecimento de marcha humana (*gait recognition*), uma modalidade biométrica não invasiva, discreta e que facilita o reconhecimento de pessoas à distância pela forma como elas caminham.

2.1 Biometria

Os constantes avanços tecnológicos somados ao aumento dos processos computacionais em larga escala nos mais variados tipos de problemas, evidenciam cada vez mais o tema Biometria, que se refere ao estudo estatístico das características físicas e comportamentais dos seres vivos (JAIN; KUMAR, 2012). Isto porque o seu uso propõe uma alternativa para identificação automática de pessoas, em relação a métodos tradicionais baseados em posse ou conhecimento.

Segundo Jain (2007), existem basicamente duas categorias de métodos tradicionais para identificação de pessoas:

- **Baseados em posse:** São os métodos que requerem algo que a pessoa possui, como, por exemplo, um documento, um cartão de identificação, uma chave, etc;
- **Baseados em conhecimento:** São os métodos que requerem alguma informação pré cadastrada que a pessoa precisa conhecer, como, por exemplo, uma senha de acesso, os dados pessoais, etc.

O emprego desses métodos tradicionais pode acarretar alguns problemas, tanto para o usuário, quanto para o sistema em si, pois uma posse pode ser perdida ou clonada e uma senha pode ser esquecida ou violada.

Os sistemas de segurança baseados em biometria são uma boa alternativa para contornar alguns problemas enfrentados pelos sistemas tradicionais, pois as características biométricas (biológicas ou comportamental) tendem a ser distintas para indivíduos distintos, são mais difíceis de serem fraudadas, e não podem ser perdidas ou esquecidas.

Apesar do tema Biometria ter ganho bastante atenção no final da década de 1990 na área de acesso primário à informação, os primeiros estudos envolvendo desenvolvimento de sistemas biométricos surgiram em meados do século XIX (MAGUIRE, 2009).

A publicação de Mitchell Trauring, em 1963, é considerada o estudo pioneiro (JAIN; NANDAKUMAR; ROSS, 2016). Além disso, os estudos de Henry Faulds, William Herschel e Francis Galton, contribuíram para estabelecer uma forma de apontar singularidades em impressões digitais, identificando padrões, como por exemplo, pontos de minúcia.

No início, os estudos relacionados à sistemas biométricos tiveram seu foco em aplicações forenses. Com o passar dos anos esses estudos foram se expandindo para tratar também de problemas envolvendo acesso restrito de pessoas a lugares e informações. Com o aprimoramento das técnicas de reconhecimento, a Biometria passou a ser utilizada de forma ampla, sendo possível encontrar atualmente aplicações para identificação biométrica em celulares e dispositivos portáteis em geral, visando não somente à segurança, mas também à praticidade e agilidade.

2.1.1 Sistemas Biométricos

Os sistemas biométricos são um conjunto de diversas tecnologias com a função de realizar tarefas de extrair, processar, armazenar, recuperar e comparar dados biológicos, com a finalidade de identificação e autenticação de indivíduos (MARCONDES, 2020).

Dependendo de sua aplicação, um sistema biométrico pode, basicamente, ter finalidade de identificação ou verificação. Um sistema que busca a identificação de indivíduos opera de forma a encontrar a identidade do usuário com base apenas em sua entrada biométrica no sistema, fazendo uma leitura em todo o banco de dados. Por outro lado, a abordagem de verificação refere-se ao problema de confirmar a identidade do usuário com base em uma entrada prévia, onde, sabendo qual indivíduo deseja-se analisar, a comparação é feita apenas com o dado biométrico cadastrado para aquele indivíduo.

A abordagem escolhida para compor um sistema biométrico dependerá muito da aplicação que se deseja desenvolver. Atualmente, é bastante comum utilizar-se tecnologias biométricas para validações de clientes em caixas eletrônicos, nos quais a pessoa, após inserir seu cartão de identificação, necessita confirmar a identidade por meio da leitura de impressão digital.

Neste caso, a impressão digital é utilizada para validar se a tentativa de acesso à conta está realmente sendo realizada pelo dono do cartão e, portanto, proprietário da conta. Já nos sistemas de acesso às academias de esportes, por exemplo, tem sido bastante comum a utilização de sistemas biométricos no modo de identificação, nos quais a leitura biométrica do usuário é realizada e toda a base de dados de clientes é comparada com aquela biometria sendo apresentada para se determinar a identidade do indivíduo.

2.1.2 Requisitos das Características Biométricas

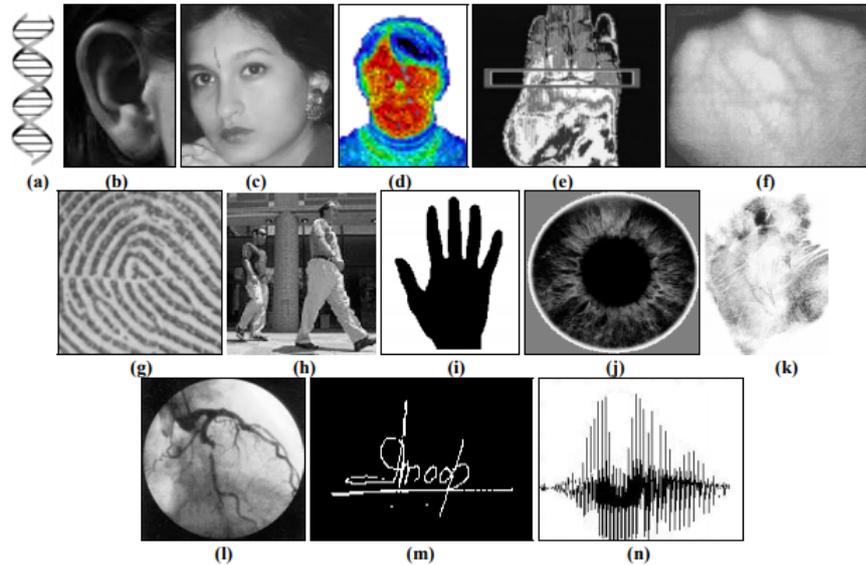
Conforme definido anteriormente, um sistema biométrico realiza a análise de características biológicas ou comportamentais dos seres vivos. Entretanto, para garantir que o sistema tenha um bom desempenho, isto é, consiga diferenciar indivíduos com uma margem de erro aceitável e estabeleça o reconhecimento correto de dados de entrada genuínos, é ideal que tais características analisadas satisfaçam aos seguintes requisitos (JAIN; ROSS; PRABHAKAR, 2004):

- **Universalidade:** Todas as pessoas do público alvo devem ter a característica;
- **Unicidade:** A característica deve ser suficientemente distinguível entre indivíduos distintos;
- **Permanência:** A característica deve ser pouco mutável com o passar do tempo;
- **Coletabilidade:** A característica deve ser mensurável quantitativamente;
- **Performance:** A característica deve propiciar precisão e velocidade de reconhecimento do sistema, incluindo recursos necessários, operações e fatores ambientais que venham a influenciar o desempenho;
- **Aceitabilidade:** A identificação por meio da característica deve contar com a aceitação das pessoas do público alvo;
- **Segurança:** A característica deve apresentar robustez quanto às possibilidades de ataques.

A Figura 3 apresenta algumas características biométricas que apresentam bons resultados levando em consideração os requisitos anteriormente descritos e que, por isso, são geralmente utilizadas para compor diversos tipos de sistemas biométricos.

A Tabela 1 mostra uma comparação apresentada por Jain, Ross e Prabhakar (2004) entre algumas características biométricas, levando-se em consideração os requisitos apresentados. Nota-se nesta Tabela que as características que têm as pontuações mais elevadas são: DNA, termograma facial e íris. As características mais utilizadas, face e impressões digitais, apresentam pontuações elevadas em algum quesito mas não se saem tão bem em outros. A única característica comportamental apresentada na Tabela 1, marcha, tem pontuação elevada em coletabilidade e aceitabilidade, por serem não invasivas, por poderem ser capturadas à distância e sem exigir colaboração do usuário. Por outro lado, nos critérios de unicidade, permanência e performance, as pontuações são baixas. Esta característica deve ser utilizada, portanto, em aplicações onde deseja-se uma identificação discreta, à distância, sem colaboração do usuário e deve, se possível, ser combinada com outras características de tal modo a contrapor os seus pontos fracos.

Figura 3 – Exemplos de características biométricas. a) DNA; b) Orelha; c) Face; d) Termograma Facial; e) Termograma da Mão; f) Veias da Mão; g) Impressão Digital; h) Caminhar; i) Geometria da Mão; j) Íris; k) Impressão da Palma; l) Retina; m) Assinatura; n) Voz



Fonte: Jain, Ross e Prabhakar (2004).

Tabela 1 – Comparação entre características biométricas. As pontuações alta, média e baixa são denotadas por A, M e B, respectivamente.

Característica Biométrica	Universalidade	Unicidade	Permanência	Coletabilidade	Performance	Aceitabilidade	Segurança
DNA	A	A	A	B	A	B	A
Face	A	B	M	A	B	A	B
Termograma facial	A	A	B	A	M	A	A
Impressão digital	M	A	A	M	A	M	M
Geometria da mão	M	M	M	A	M	M	M
Veias da mão	M	M	M	M	M	M	A
Íris	A	A	A	M	A	B	A
Marcha	M	B	B	A	B	A	M

Fonte: Adaptado de Jain, Ross e Prabhakar (2004).

2.2 *Soft Biometrics*

Analisando o contexto histórico, os primeiros sistemas biométricos desenvolvidos para identificação de criminosos eram baseados em três diferentes conjuntos de características (JAIN; DASS; NANDAKUMAR, 2004):

- Medidas antropométricas, como a altura, o peso, o comprimento dos braços, etc;
- Descrições morfológicas da aparência, como cor dos olhos, cor da pele, anomalias específicas, etc;
- Presença de marcas peculiares no corpo, como cicatrizes e tatuagens.

É importante notar que para a época, meados do século XIX, tais sistemas eram importantes e foram muito úteis para se promover a identificação de criminosos. Porém, as taxas de universalidade e permanência presentes nestes sistemas eram baixas e seriam inaceitáveis para os padrões atuais, isto porque as características utilizadas naquela época no reconhecimento de pessoas não apresentavam bons resultados.

Mesmo com baixo poder de distinção de tais características, comparado aos métodos mais tradicionais, como impressão digital e íris, o seu uso não deve ser descartado. Em outras palavras, essas características não possuem a unicidade e a permanência desejáveis para identificar de forma exclusiva um indivíduo durante um período e, portanto, são chamadas de *soft biometrics* (JAIN; DASS; NANDAKUMAR, 2004).

Apesar das limitações, o uso de *soft biometrics* tem ganhado destaque e despertado interesse na área de Biometria. Isso se deve à possibilidade de integração de tais características à sistemas biométricos tradicionais com o intuito de melhorar suas acurácias (DANTCHEVA et al., 2011). Na Figura 4 pode-se observar algumas características que se enquadram na categoria de *soft biometrics*.

Existem diversos tipos de características *soft biometrics* localizadas em diferentes regiões do corpo humano. Na Tabela 2 observam-se exemplos de algumas características que se enquadram nesta categoria, relacionadas quanto à região do corpo, à natureza dos valores que são obtidos, à permanência com o passar dos anos e ao poder de distinção entre indivíduos.

2.3 Reconhecimento da Marcha Humana

Um dos primeiros trabalhos que estudaram a marcha humana foi o de Johansson (JOHANSSON, 1975), que mostrou que as pessoas são capazes de reconhecer o modo de caminhar das outras pessoas e identificar familiares desta maneira.

Pode-se definir *Gait*, ou marcha, como sendo comportamentos motores compostos por movimentos integrados do corpo humano. Trata-se de um padrão cíclico de movimentos corporais que se repetem indefinidamente a cada ciclo (ARANTES, 2010a).

A análise da marcha humana é muito importante na área médica, por exemplo, para detecção de tratamento de distúrbios locomotores. Alterações básicas no padrão da marcha servem como indicadores do início de doenças, tais como: Mal de Parkinson, Esclerose Múltipla e Hidrocefalia (ARANTES, 2010a).

Figura 4 – Exemplos de características biométricas da categoria *soft biometrics*.

Fonte: Jain, Dass e Nandakumar (2004).

Tabela 2 – Propriedades de algumas características biométricas da categoria *soft biometrics*.

Soft Biometric	Região	Natureza	Permanência	Distinção
Cor da pele	Face/Corpo	Contínua	Média	Baixa
Cor do cabelo	Face	Contínua	Média	Média
Cor dos olhos	Face	Contínua	Alta	Média
Barba	Face	Binária	Baixa/Média	Baixa
Bigode	Face	Binária	Baixa/Média	Baixa
Características Faciais	Face	Contínua	Alta	Alta
Gênero	Face/Corpo	Binária	Alta	Baixa
Idade	Face/Corpo	Contínua	Baixa/Média	Média
Peso	Corpo	Contínua	Média/Alta	Média
Altura	Corpo	Contínua	Baixa/Média	Média

Fonte: Adaptado de Dantcheva et al. (2011).

Durante muito tempo a análise da marcha ficou restrita à área médica, mas agora está difundindo-se para outras áreas, como, por exemplo, a biometria.

Pesquisas científicas comprovaram que os humanos têm formas distintas e especiais de caminhar (WINTER, 1991; SARKAR et al., 2005; HAVASI; SZLAVIK; SZIRANYI, 2007; BOULGOURIS; CHI, 2007). A partir desta premissa, a marcha humana pode ser entendida como uma importante característica biométrica.

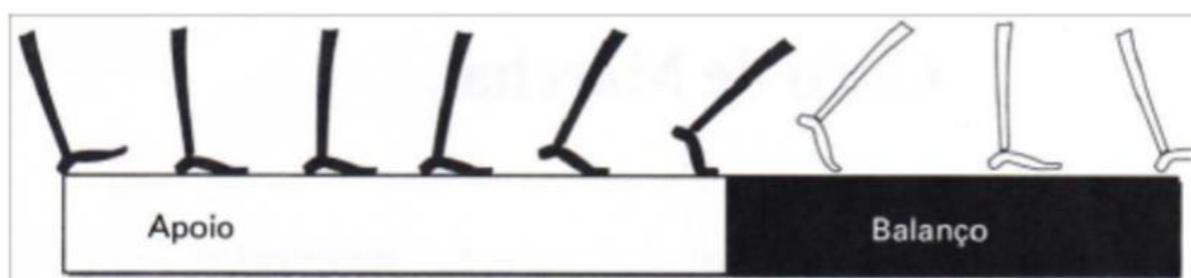
Dentre diversos tipos de sistemas biométricos existentes, o grande desafio está em fazer a identificação de pessoas em imagens de vídeo, em tempo real e de forma não invasiva. Em muitas situações, técnicas biométricas convencionais, como reconhecimento pela íris, face e impressão digital, não são possíveis de serem utilizadas, por falta de cooperação do indivíduo sendo identificado ou devido à distância em que o mesmo se encontra. Por isto, há a necessidade de se avaliar características biométricas que possam ser obtidas de forma não invasiva, como, por exemplo, a marcha humana.

O reconhecimento da marcha humana apresenta muitos desafios, pois a correta identificação do indivíduo depende de uma segmentação precisa da imagem, onde diversos fatores podem influir no resultado, tais como a roupa do indivíduo, as alterações de estado de humor, a velocidade do caminhar, o tipo de calçado, a superfície da caminhada (ex: concreto, grama) as variações do ângulo da câmera ou, até mesmo, o fato da pessoa estar carregando algum objeto. Outro grande desafio também está na extração das características que possam identificar a marcha de forma única.

2.3.1 Ciclo da Marcha

A marcha utiliza uma sequência de repetições de movimento para mover o corpo para a frente enquanto, simultaneamente, mantém a postura estável. Conforme o corpo move-se para frente, um membro serve como fonte móvel de apoio, enquanto outro membro avança para a nova posição. Em seguida, os membros invertem seus papéis. Essa sequência de movimentos é denominada de ciclo de marcha (PERRY, 2005), ou seja, é o período que decorre entre o contato do calcanhar de um pé, e o contato seguinte do mesmo calcanhar. Este ciclo divide-se em duas fases - apoio e balanço - frequentemente chamados de fases da marcha. Quando o pé está tocando a superfície é o período de apoio. Este período começa com o contato inicial do pé na superfície, conforme ilustrado na Figura 5.

Figura 5 – Representação dos períodos de apoio e de balanço no ciclo da marcha.



Fonte: Adaptado de (ARANTES, 2010b).

2.3.2 A Marcha Humana no Contexto da Biometria

Os primeiros estudos sobre a marcha humana foram apresentados por (JOHANSSON, 1975). Ele apresentou a um grupo de observadores, imagens reduzidas a um ponto de luz, na forma de *display*. Seus experimentos sugerem que os seres humanos têm alguma noção implícita do movimento humano e pode-se reconhecer dados temporais dentro deste contexto. Estudos posteriores demonstraram que não somente a forma humana pode ser distinguida a partir de pontos de luz, mas também movimentos, como: saltar, caminhar e inclinar (DITTRICH, 1993). Em um dos estudos e experimentos conduzidos por Dittrich, concluiu-se que o gênero do indivíduo pode ser reconhecido a partir dos movimentos humanos e suas características inerentes. Johansson demonstrou também que observadores poderiam reconhecer-se com segurança, assim como seus amigos, a partir de um *display* de pontos de luz dinâmicos.

Basicamente, as técnicas propostas para o reconhecimento da marcha são divididas em dois grupos principais: baseadas em silhueta (*Silhouette-Based*) (LAM; CHEUNG; LIU, 2011) ou baseadas em modelo (ou na dinâmica) (*Model-Based*) (BENABDELKADER; DAVIS; CUTLER, 2002; NIXON et al., 2009). Na abordagem baseada na silhueta, a forma desta e seus atributos são as características mais importantes. Assim sendo, o reconhecimento do indivíduo é realizado com medições que refletem a sua forma. As abordagens baseadas no modelo têm por objetivo obter o movimento do tronco e/ou das pernas. Ao contrário de abordagens baseadas em silhuetas, esta geralmente se concentra na dinâmica do movimento, omitindo sua forma corporal.

Técnicas baseadas no modelo podem ser divididas em duas classes principais (BENABDELKADER; DAVIS; CUTLER, 2002): baseadas no espaço ou baseadas no espaço-tempo. As técnicas baseadas no espaço consideram o movimento da marcha como sendo composto por uma sequência de poses estáticas do corpo e o reconhecimento é realizado por meio de observações de variações temporais com respeito a estas poses. As técnicas baseadas no espaço-tempo visam capturar os padrões espaço-temporais discriminantes em uma sequência de marcha para identificação humana.

2.4 Fusão de Características Biométricas

Um sistema biométrico baseado em duas ou mais características biométricas é chamado de sistema biométrico multimodal (PANCHAL; SINGH, 2013). O uso de sistemas biométricos multimodais pode ser uma alternativa para melhorar o reconhecimento de pessoas em cenários em que os dados de entradas podem sofrer variações ou ruídos.

Outro aspecto importante do uso de sistemas biométricos multimodais é o aumento na robustez contra fraudes. Isto se dá pelo fato de que a imposição de duas ou mais características biométricas para realizar a identificação dificulta o emprego de técnicas fraudulentas, pois, caso algum impostor decida burlar o sistema, será necessário que o mesmo falsifique todas

as características de um determinado indivíduo sendo utilizadas pelo sistema. Dessa forma, o sistema apresenta uma maior taxa de precisão e uma maior proteção contra fraudes (PANCHAL; SINGH, 2013).

Segundo Ross e Jain (2003), a etapa de fusão de dois tipos de biometrias distintas pode ocorrer em três níveis diferentes do processo de reconhecimento de um sistema biométrico:

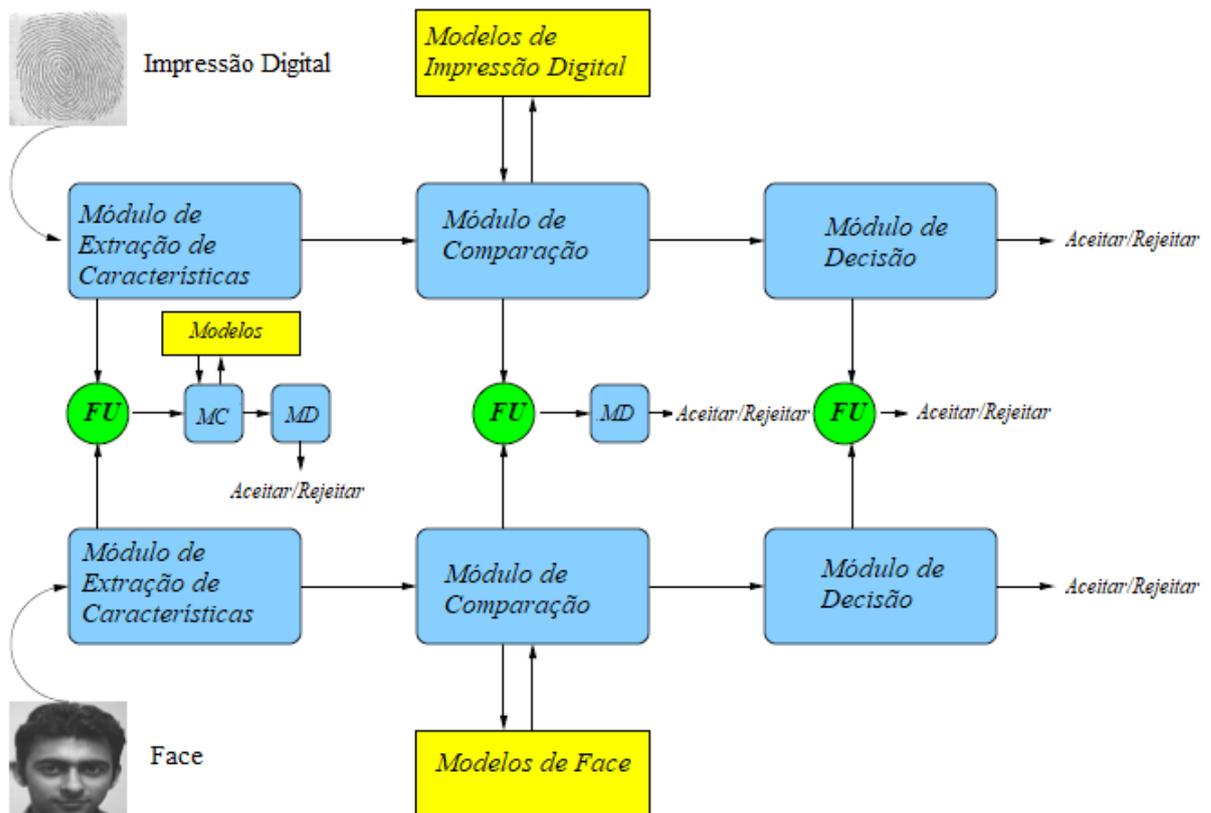
1. **Nível de características:** As informações obtidas por meio dos dados capturados por diferentes sensores são transformadas em vetores de características que são posteriormente combinados, como por exemplo, por meio de uma concatenação. Um novo vetor de características com maior dimensão é, portanto, obtido para representar o indivíduo. Um aspecto negativo da concatenação é o aumento da dimensionalidade dos dados, nesse caso é possível aplicar técnicas de redução de dimensionalidade e gerar uma representação mais compacta;
2. **Nível de pontuação:** Os sistemas biométricos baseados nas características sendo combinadas calculam as respectivas pontuações de similaridade entre os descritores analisados, sendo essas pontuações combinadas posteriormente, por meio de operações de soma, máximo, multiplicação, por exemplo, gerando uma única pontuação de similaridade a fim de estabelecer ou autenticar a identidade do indivíduo;
3. **Nível de decisão:** Os sistemas biométricos baseados nas características sendo combinadas calculam as respectivas pontuações de similaridade e tomam a decisão sobre a identidade do indivíduo. A decisão final sobre a identidade do indivíduo será dada, por exemplo, pela maioria dos votos, sendo que ponderamentos específicos podem ser aplicados para se chegar à decisão final.

A Figura 6 ilustra o esquema de um sistema biométrico bimodal com as diferentes possibilidades de fusão entre duas características biométricas, levando em consideração as abordagens em três níveis. As caixas azuis maiores representam os níveis do processo de reconhecimento de um sistema biométrico, sendo que o sentido para a direita corresponde ao modelo tradicional unimodal. Ao se direcionar para o círculo verde, a fusão é estabelecida com a outra característica, que segue o restante do processo de reconhecimento sendo ilustrado pelas caixas azuis menores com as informações combinadas.

Além desses três níveis de fusão, Panchal e Singh (2013) propõem um tipo de fusão de característica a nível do sensor de captura, isto é, características biométricas obtidas por meio de sensores diferentes ou em momentos diferentes, são combinadas para originar o dado de entrada para o sistema.

Além do método de fusão em diferentes níveis, é possível realizar a combinação de biometrias em um modelo cascata. A abordagem multimodal em diferentes níveis funciona

Figura 6 – Sistema biométrico bimodal mostrando os três níveis de fusão: características, pontuação e decisão.



Fonte: Adaptado de Ross e Jain (2003).

com dois ou mais sistemas biométricos em paralelo, que em determinado momento terão suas informações combinadas para se tomar uma decisão. O modelo de classificador em cascata, por sua vez, opera em um esquema de estágios, sendo um tipo de biometria definido como uma primeira etapa a ser cumprida, limitando o processo de busca de um segundo estágio ou simplesmente excluindo a necessidade de uma segunda verificação.

Um exemplo de biometria com classificador em cascata pode ser encontrado no trabalho de Zheng et al. (2012), onde informações de imagens cumulativas da pressão dos pés sobre o solo durante uma caminhada são utilizadas em um primeiro nível para limitar a base de dados de busca do segundo nível, com imagens sequenciais da forma de caminhar do indivíduo.

2.4.1 Fusão de Características *Soft Biometrics*

Não há muitos trabalhos na literatura utilizando *soft biometrics* como traço biométrico prioritário, sobretudo quando se trata do uso de fusão de características biométricas. Isso impulsiona ainda mais nosso trabalho, já que é uma área com um vasto campo ainda a ser

explorado. Ademais, os autores que estudam essa metodologia têm compartilhado ótimos resultados. GUO, NIXON e CARTER (2018) por exemplo, mostraram que características da categoria *soft biometrics*, tais como o dorso, vestimenta e forma de caminhar, atingiram resultados próximos à traços biométricos bem difundidos na literatura, como face, principalmente em imagens capturadas por câmeras posicionadas à uma distância significativa dos indivíduos. Neste trabalho, os autores mostraram também que a fusão das características da categoria *soft biometrics* com características da face melhoraram a acurácia geral em todas as distâncias analisadas: perto, médio e longe.

Do mesmo modo, SOBABE, DJARA e VIANOU (2019) apresentam um algoritmo que utiliza *soft biometrics* como metodologia biométrica secundária. Quando os níveis de *threshold* não são atingidos, são utilizadas características da categoria *soft biometrics* para auxiliar na identificação das pessoas.

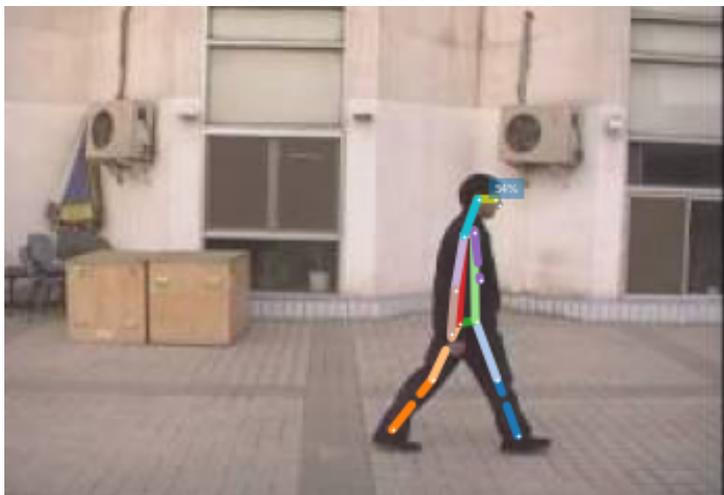
2.5 Estimativa de Pose Humana 2D

A estimativa de pose humana pode ser definida como a localização de articulações do corpo humano dada uma imagem (TOSHEV; SZEGEDY, 2014). Com a localização das articulações, é possível calcular medidas de diferentes partes do corpo e utilizá-las como características biométricas para a identificação da pessoa, como propõem-se nesta dissertação. A estimativa de pose serve também como uma ferramenta fundamental para várias aplicações práticas, como a interação humano-computador (MOESLUND; GRANUM, 2001), a re-identificação de pessoas (SU et al., 2017) e o reconhecimento de ações humanas (SILVA; MARANA, 2020).

Neste trabalho foram utilizados e avaliados dois diferentes métodos para estimar poses humanas 2D: *OpenPose* (CAO et al., 2018a) e *PifPaf* (KREISS; BERTONI; ALAHI, 2019b). As Figuras 7 e 8 mostram os resultados obtidos pelos métodos *PifPaf* e *OpenPose*, respectivamente, para estimar as poses 2D dos indivíduos em um quadro de um vídeo, e a Tabela 3 apresenta exemplos de informações usualmente retornadas pelos algoritmos de estimativa de pose. Nesta tabela, o *Score* faz referência à confiabilidade do algoritmo sobre afirmar que um determinado ponto representa, de fato, uma determinada parte do corpo, sendo 1,0 o seu valor máximo e 0,0 o seu valor mínimo, os valores de X e Y significam as coordenadas do ponto no espaço 2D. As dimensões desse espaço são definidas após a equalização de altura e largura de todas as imagens.

É possível perceber um detalhe de detecção de falso positivo mostrada na Figura 8, isso pode ser explicado, pois o *OpenPose* tem a característica de detectar pontos minuciosos e detalhistas nas imagens. O algoritmo é capaz de registrar até 156 pontos, porém o mesmo acaba se confundindo quando ocorrem oclusões ou quando os humanos não são bem destacados. No caso do *PifPaf*, o método apenas se preocupa em encontrar um esqueleto humano sem se importar muito com detalhes, por exemplo, e por este motivo o mesmo tem alcançado melhores

Figura 7 – Estimativas de poses 2D utilizando o método *PifPaf* (KREISS; BERTONI; ALAHI, 2019b).



Fonte: Elaborada pelo autor.

Figura 8 – Estimativas de poses 2D utilizando o método *OpenPose* (CAO et al., 2018a).



Fonte: Elaborada pelo autor.

resultados no nosso contexto.

2.5.1 PifPaf

Proposto por Kreiss, Bertoni e Alahi (2019b), o principal objetivo do método *PifPaf* é estimar poses humanas em imagens com várias pessoas. Para tanto, ele utiliza algumas redes neurais. Uma das redes estima precisamente as regiões de intensidade das partes do corpo humano, *Part Intensity Field* (PIF), ou seja as regiões de localização das juntas que compõem

Tabela 3 – Exemplo de informações geradas por algoritmos de estimativa de pose.

Identificador	Parte do Corpo	Eixo Y	Eixo X	Score
0	Nariz	0,55	0,12	0,81
1	Pescoço	0,49	0,23	0,74
2	Ombro Direito	0,38	0,23	0,67
3	Antebraço	0,31	0,35	0,71

Fonte: Elaborada pelo autor.

as partes do corpo humano. Já uma outra rede estima a associação entre tais partes, processo chamado de *Part Association Field* (PAF). O nome do método, *PifPaf*, origina da junção dessas duas etapas.

Os *Pifs* são definidos como uma estrutura composta de uma pontuação para a confiança, um vetor que aponta para a parte do corpo mais próxima de algum tipo particular e o tamanho da articulação.

Os *Pafs* são compostos de dois vetores para as duas partes que a associação se conecta e duas larguras. A localização desses pontos é feita com regressões e por utilizar métodos baseados em grade, não sofre discretização.

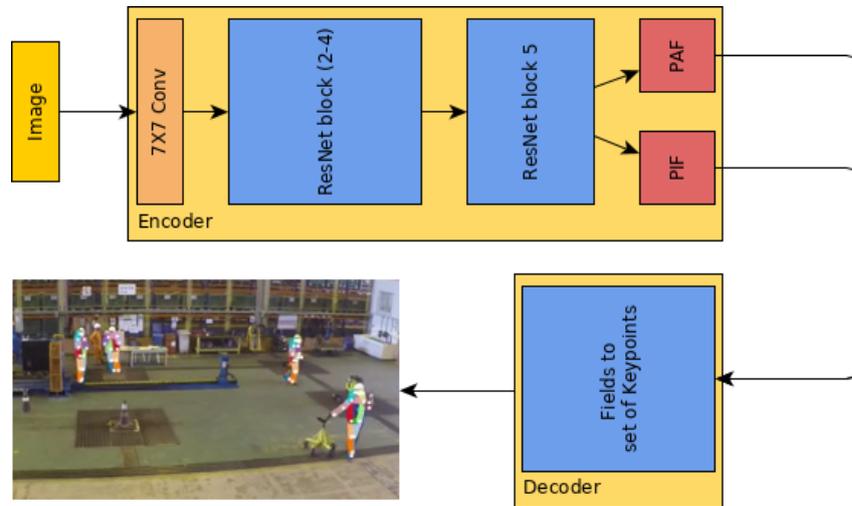
Para construir o *Paf* em um mapa de características é utilizado um algoritmo de duas etapas. Em primeiro lugar, a junção mais próxima de dois tipos é encontrada e, com isso, determina-se um dos componentes do vetor. Em seguida, o *ground truth* é utilizado para determinar o outro componente vetorial. Por fim, a junção de *Pifs* e *Pafs* constrói o esqueleto. A Figura 9 apresenta um diagrama do método *PifPaf*.

As Figuras 10 e 11 exemplificam o funcionamento do método *PifPaf*. A Figura 10(a) apresenta um mapa de confiança original gerado para os ombros esquerdos das pessoas presentes em uma determinada imagem. Para melhorar a localização do mapa de confiança, realiza-se uma fusão com a parte vetorial do PIF, apresentado na Figura 10(b), formando-se um mapa de confiança de alta resolução, conforme mostra a Figura 10(c).

A Figura 11 mostra o PAF que associa os ombros esquerdos aos respectivos quadris esquerdos das pessoas presentes na imagem. Cada localização do mapa de características é a origem de dois vetores que apontam para os ombros e quadris para associação. A Figura 11 (a) mostra a confiança das associações e a Figura 11 (b) mostra os componentes do vetor.

2.5.2 OpenPose

Proposto por Cao et al. (2018a), o método OpenPose baseia-se em um *pipeline* com vários estágios. Inicialmente, gera-se um conjunto de mapas de confiança 2D das partes do corpo. Após, gera-se um conjunto de campos vetoriais de afinidade parciais, que descrevem

Figura 9 – Diagrama do método *PifPaf*.

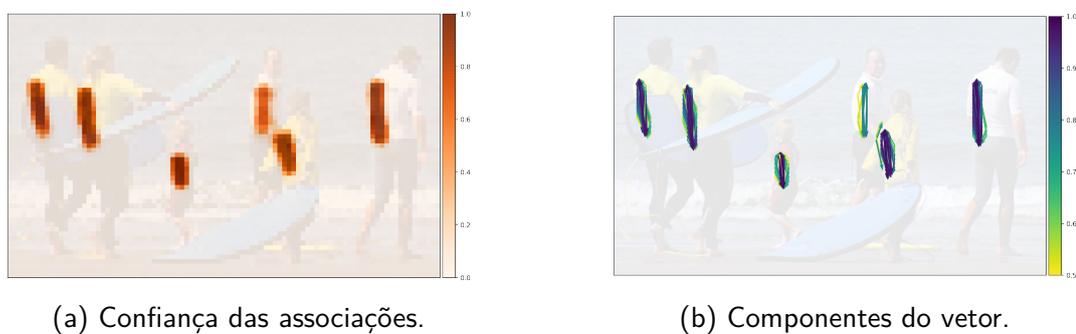
Fonte: Elaborada pelo autor.

Figura 10 – Componentes PIF para o ombro esquerdo.



Fonte: Kreiss, Bertoni e Alahi (2019b).

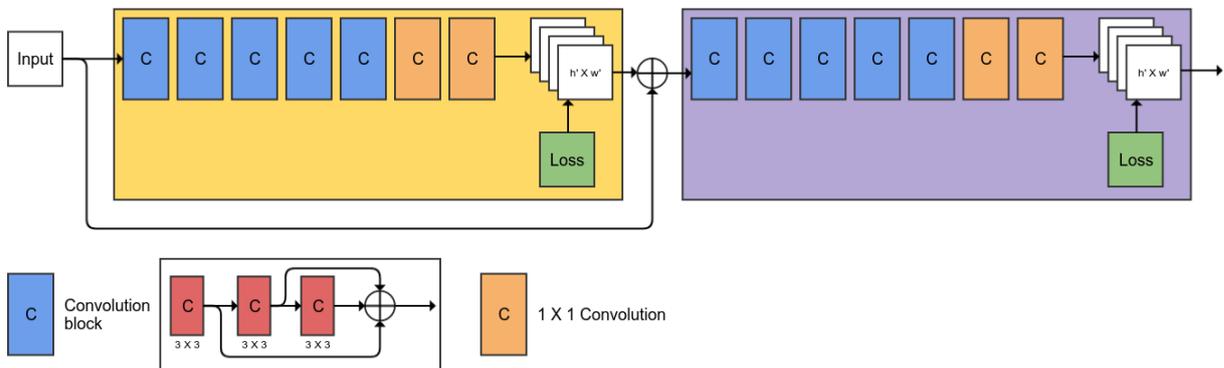
Figura 11 – PAF que associa ombros esquerdos com respectivos quadris esquerdos das pessoas presentes na imagem.



Fonte: Kreiss, Bertoni e Alahi (2019b).

as afinidades entre duas partes do corpo. A Figura 12 apresenta um diagrama do método *OpenPose*.

Figura 12 – Diagrama do pipeline do método *OpenPose*.



Fonte: Elaborada pelo autor.

Os mapas de confiança são representações em 2D da ideia de que, dada uma imagem que tem um corpo humano, uma parte dela pode ser localizada em qualquer pixel. Se uma imagem contiver apenas um indivíduo, então é esperado um único pico. Por outro lado, em um cenário com múltiplos indivíduos, deve haver um pico para cada parte do corpo de cada indivíduo.

Em suma, o método *OpenPose* recebe como entrada uma imagem colorida, conforme apresentado na Figura 13 (a), e, como saída, produz a localização das articulações do corpo em 2D e uma pontuação de confiança para cada pessoa encontrada na imagem, conforme apresentado na Figura 13 (e). Inicialmente, uma rede neural convolucional estima um mapa de confiança 2D para cada articulação do corpo (Fig. 13 (b)) e um campo vetorial 2D de campos de associação de partes (PAFs) para cada membro, que representa o grau de associação entre o corpo juntas (Fig. 13 (c)). No final, os mapas de confiança e os PAFs são analisados (Fig. 13 (d)) para a saída da pose 2D, que inclui coordenadas de articulações 2D e uma pontuação de confiança para todas as pessoas detectadas na imagem (CAO et al., 2018a).

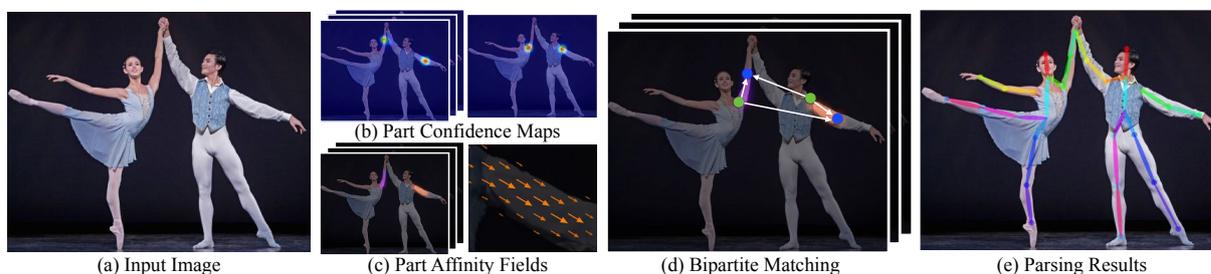
2.5.3 Tf-Pose-Estimation

O método *TF-Pose Estimation*¹ é uma implementação do método *OpenPose* usando a estrutura *Tensorflow* (ABADI et al., 2015). Ele possui algumas variações com mudanças na estrutura da rede, o que permite utilizá-la em tempo real na CPU ou em dispositivos embarcados.

Nesta dissertação, utilizou-se também esta implementação para verificar se há diferença no comportamento observável e se isso impacta de alguma forma o resultado obtido. No

¹ <https://github.com/ZheC/tf-pose-estimation>

Figura 13 – Pipeline geral do método OpenPose. (a) Imagem de entrada; (b) Mapas de confiança para detecção das articulações; (c) PAFs para associação parcial; (d) Conjunto de correspondência para associar as articulações que formam os membros; (e) As poses 2D de corpo inteiro construídas para todas as pessoas encontradas na imagem.



Fonte: Cao et al. (2018a).

Capítulo 6, os resultados obtidos dos experimentos com as duas implementações do método OpenPose, framework Caffe original e framework TensorFlow, são listados.

2.6 Métricas

Nesta seção apresentam-se as métricas comumente utilizadas para avaliação dos métodos de identificação da área de Biometria.

2.6.1 Curvas ROC e CMC

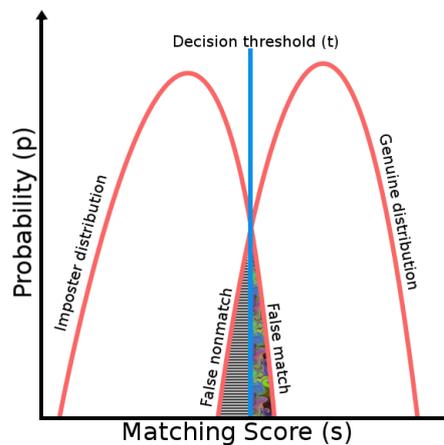
Devido ao ruído, às condições ambientais, às mudanças nas características pessoais e até mesmo na forma como o usuário interage com o sensor, duas amostras de uma característica biométrica da mesma pessoa nunca são iguais. Portanto, os sistemas biométricos calculam uma pontuação da comparação entre a característica de consulta (*probe*) e a característica do banco de dados (*gallery*) e essa pontuação determinará se a comparação é genuína ou impostora. Para tanto, um limiar t deve ser definido para regular a decisão do sistema. Se a pontuação for maior ou igual a t , considera-se que as características pertencem ao mesmo indivíduo.

Devido à sua natureza, os sistemas biométricos podem apresentar dois tipos de erros (PRABHAKAR; PANKANTI; JAIN, 2003):

- FMR (*False Match Rate*): as características biométricas de duas pessoas diferentes são consideradas como sendo proveniente da mesma pessoa;
- FNMR (*False Non-Match Rate*): as características biométricas de uma mesma pessoa são consideradas como sendo proveniente de duas pessoas distintas.

Para funcionar corretamente, um sistema biométrico deve fazer uma compensação entre essas duas taxas de erro. Se o projetista do sistema decidir diminuir a FMR (taxa de correspondência falsa), então a FNMR (taxa de não correspondência falsa) aumentará. Por outro lado, se o designer tentar facilitar o login do usuário, a FMR aumentará. A Figura 14 mostra a correlação entre essas duas taxas de erro.

Figura 14 – Taxas de erro do sistema biométrico: FMR (Taxa de correspondência falsa) e FNMR (Taxa de não correspondência falsa)



Fonte: (PRABHAKAR; PANKANTI; JAIN, 2003).

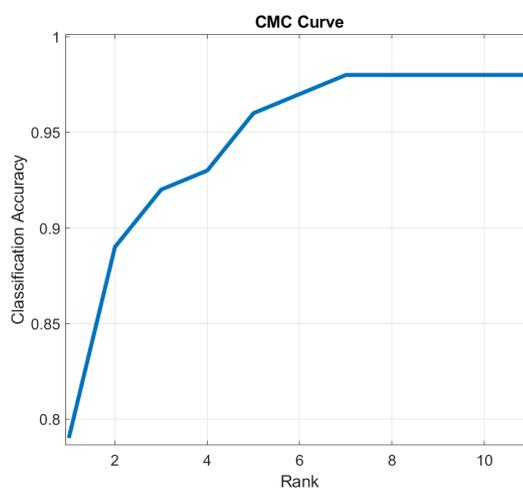
Outras taxas de erro são FAR (*False Accept Rate* - Taxa de Falsa Aceitação) e FRR (*False Reject Rate* - Taxa de Falsa Rejeição), elas são análogas a FMR e FNMR. FAR é a taxa em que um falso sujeito é aceito como genuíno e FRR é a taxa em que uma correspondência genuína é categorizada como impostora.

Dadas as duas taxas de erro, FAR e FRR, é possível calcular a EER (*Equal Error Rate* - Taxa Igual de Erro), que é uma medida de nível de segurança que identifica um limite quando as taxas FAR e FRR têm o mesmo valor.

Para avaliar os sistemas biométricos, também são utilizadas curvas de desempenho (MARTIN et al., 1997). A curva ROC (*Receiver Operating Characteristics*) é uma maneira de entender o desempenho de um sistema biométrico. Uma curva ROC é um gráfico bidimensional onde a taxa de verdadeiro positivo (TPR - *True Positive Rate*) é traçada no eixo Y, enquanto a taxa de falso positivo (TFR - *True False Rate*) é traçada no eixo X (FAWCETT, 2006).

Para aplicações forenses o FNMR é considerado mais crítico do que o FMR, pois, normalmente, este tipo de aplicação trata de identificação criminal em que é muito importante não deixar um suspeito passar sem ser identificado, mesmo que seja necessário fazer manualmente a identificação correta em um grupo de indivíduos selecionados. Para aplicações de alta segurança ocorre o contrário, uma pessoa genuína identificada de forma errada poderá não

Figura 15 – Um exemplo de curva CMC, em que no eixo das abscissas tem-se as posições do ranque e no eixo das ordenadas tem-se a acurácia obtida pelo classificador.



Fonte: (PRABHAKAR; PANKANTI; JAIN, 2003).

obter acesso ao seu próprio sistema. Para aplicações civis, um equilíbrio entre as duas taxas é o cenário desejado.

Outra forma de avaliar o desempenho de um sistema biométrico é utilizar a curva CMC (Cumulative Match Characteristic). Para construir uma curva CMC é necessário ordenar as pontuações obtidas quando se compara a característica biométrica de pesquisa (*probe*) com todas as características biométricas armazenadas no banco de dados (*gallery*). Desse modo, é possível calcular a probabilidade do indivíduo genuíno aparecer nas primeiras K posições (*rank-k*). A curva CMC é uma representação visual das K maiores classificações para um sistema biométrico, com $K \leq N$ e N sendo o número de identidades. A Curva CMC é uma métrica *baseada na classificação*, enquanto a Curva ROC é uma métrica *baseada em agregação* (DECANN; ROSS, 2013). A Figura 15 mostra um exemplo de uma Curva CMC.

Também é importante entender a relação entre a curva ROC e a curva CMC. Apesar de a Curva CMC não ter um limiar (*threshold*), a forma como a correspondência (casamento) é realizada pode se relacionar com o FAR e o FRR. Essa relação é dada pelo entendimento de que para cada busca de pontuação genuína, o valor da pontuação impostora pode ser considerada como um limiar virtual. Com isso em mente, é possível construir uma Curva CMC, dadas as distribuições de pontuações impostoras e genuínas em relação a um comparador 1 : 1 (BOLLE et al., 2005).

Neste trabalho, utilizam-se as curvas CMC, com destaque para o valor do Ranking-1, uma vez que para identificação baseada em medidas antropométricas e marcha os sistemas biométricos normalmente operam no modo de identificação (e não em modo de autenticação). Para o cálculo das distâncias (pontuações das comparações) utiliza-se a distância Euclidiana ou a distância *City-Block*.

2.6.2 Precisão, Revocação e Medida F1

Em reconhecimento de padrões, recuperação de informações e classificação (aprendizado de máquina ou classificadores), a precisão (também chamada de valor preditivo positivo) é a fração de instâncias relevantes entre as instâncias recuperadas, enquanto a revocação é a fração da quantidade total de instâncias relevantes que foram realmente recuperadas. A precisão e a revocação são, portanto, baseadas em uma compreensão e medida de relevância. As definições dessas duas métricas estão apresentadas nas equações 1 e 2, respectivamente (GOUTTE; GAUSSIER, 2005):

$$precisão = \frac{VP}{VP + FP} \quad (1)$$

$$revocação = \frac{VP}{VP + FN} \quad (2)$$

em que VP representa os verdadeiros positivos, FP representa os falsos positivos e FN representa os falsos negativos.

As medidas precisão e revocação podem ser usadas juntas na medida F1 (F1 Score), definida como a média harmônica simples, em que a precisão e a revocação têm o mesmo peso). A medida F1 está definida na equação 3 (GOUTTE; GAUSSIER, 2005):

$$F1 = 2 \cdot \frac{precisão \cdot revocação}{precisão + revocação} \quad (3)$$

2.7 Funções de Distância

As funções de distância são definidas dentro de um espaço métrico e podem ser formalmente descritas da seguinte forma: dado um conjunto O , e três pontos o_p, o_q e $o_r \in O$, uma função de distância métrica em O é uma função $d : O \times O \rightarrow R$ que possui as seguintes propriedades:

1. Simetria: $d(o_q, o_r) = d(o_r, o_q)$;
2. Não negatividade: $d(o_q, o_r) > 0$ se $o_q \neq o_r$ e $d(o_q, o_r) = 0$ se $o_q = o_r$;
3. Desigualdade triangular: $d(o_q, o_p) \leq d(o_q, o_r) + d(o_r, o_p)$.

Neste trabalho as funções de distância são utilizadas para medir a similaridade entre dois vetores de características biométricas. Dentre as funções de distância mais comumente utilizadas, encontram-se:

Distância de Minkowski : Considerando dois pontos o e o' , pertencentes a um espaço vetorial n -dimensional, ou seja, $o = \{o_1, o_2, \dots, o_n\}$ e $o' = \{o'_1, o'_2, \dots, o'_n\}$, a distância de Minkowski, L_p , entre o e o' é dada por:

$$L_p(o, o') = \sqrt[p]{\sum_{i=1}^n (o_i - o'_i)^p} \quad (4)$$

em que p é um parâmetro que modifica a distância de Minkowski resultando em outras medidas conhecidas. Por exemplo, se $p = 2$, tem-se a **Distância Euclidiana**, se $p = 1$, tem-se a **Distância City-Block**. Caso $p \geq 1$ tem-se uma distância métrica, e caso $p < 1$, a desigualdade triangular não é obedecida, o que resulta na inexistência de uma distância métrica.

Distância Qui-Quadrado (χ^2): Considerando dois pontos o e o' , pertencentes a um espaço vetorial n -dimensional, ou seja, $o = \{o_1, o_2, \dots, o_n\}$ e $o' = \{o'_1, o'_2, \dots, o'_n\}$, a distância χ^2 é definida como:

$$\chi^2(o, o') = \frac{1}{2} \sum_{i=1}^n \frac{(o_i - o'_i)^2}{o_i + o'_i} \quad (5)$$

Esta função de distância é uma métrica não linear. Ela realça as grandes discrepâncias existentes entre dois vetores de características e é amplamente utilizada para comparar histogramas (YANG et al., 2015).

2.8 Classificadores

2.8.1 K-Vizinhos Mais Próximos - KNN

O algoritmo de k-vizinhos mais próximos (KNN) é um algoritmo de classificação *lazy*, sendo assim não envolve a construção de um modelo prévio.

O algoritmo é definido por calcular uma medida de distância de um exemplo a ser classificado com todos os outros da base de treinamento que estão rotulados. O novo exemplo é rotulado com base na quantidade de votos dos k exemplos mais próximos sendo que cada voto tem a distância como peso (WETTSCHERECK; AHA; MOHRI, 1997).

2.8.2 Rede Neural Artificial - RNA

Uma rede neural artificial (RNA) é definida como uma coleção de unidades, normalmente conhecidas como neurônios, que são conectadas de forma a gerar diferentes estruturas. As propriedades de uma RNA são definidas por suas unidades e topologia (RUSSELL; NORVIG, 2016).

Devido a sua inspiração biológica no cérebro humano, a forma como o padrão de ativação é propagada pelas suas unidades é de suma importância. Essa propagação ocorre pelas

conexões entre neurônios, sendo assim, é necessária uma conexão entre um neurônio i para um neurônio j responsável pela propagação do padrão de ativação (RUSSELL; NORVIG, 2016). A saída de um neurônio é calculada utilizando uma função de ativação, que pode ser definida como um *hard-threshold*, uma função logística ou uma unidade linear retificada (ReLU) (NAIR; HINTON, 2010). Ao utilizar o *hard-threshold* como função de ativação o neurônio é conhecido como perceptron.

As ligações de entrada no neurônio j possuem peso $w_{i,j}$ e o valor da entrada é definida pela Equação 6:

$$in_j = \sum_{i=0}^n w_{i,j} a_i \quad (6)$$

sendo n o número de ligações de entrada, a_i o valor da ligação de entrada do neurônio para o j , in_j o valor de entrada para a função de ativação e $w_{i,j}$ o peso de cada ligação.

O processo de aprendizado de uma RNA é definido pela atualização dos pesos com base na minimização da função de *loss*. Uma função de *loss* é definida como a quantidade de utilidade que é perdida ao estimar $h(x) = \hat{y}$ quando a resposta é $f(x) = y$, sendo h uma hipótese (RUSSELL; NORVIG, 2016).

Como normalmente os problemas são não-lineares, é necessário a utilização do gradiente descendente. O gradiente descendente lida com o problema de otimização de uma função de *loss* conhecendo sua derivada de primeira ordem (WU; WARD; BOTTOU, 2018). O cálculo dos novos pesos é definido por:

$$w_{j+1} = w_j - \eta_j \nabla f(w_j) \quad (7)$$

sendo η a razão do aprendizado e f a função de *loss*.

2.8.2.1 MLP - *Multilayer Perceptron*

Uma *Multilayer Perceptron* (MLP) é uma rede neural artificial composta por mais de um perceptron. Eles são compostos por uma camada de entrada para receber o sinal, uma camada de saída que toma uma decisão ou previsão sobre a entrada, e entre esses dois, um número arbitrário de camadas ocultas que são o verdadeiro mecanismo computacional do MLP. MLPs com uma camada oculta são capazes de aproximar qualquer função contínua (ORRÙ et al., 2020).

3 Trabalhos Correlatos

Neste capítulo são apresentados trabalhos e métodos encontrados na literatura que compartilham dos mesmos objetivos e abordam os mesmos temas desta dissertação.

3.1 Métodos Baseados em Medidas Antropométricas

3.1.1 Método Baseado em Poses

Pietikäinen (2005) utiliza as medidas antropométricas capturadas por um Kinect RGB-D para a identificação de pessoas. Neste trabalho são agregadas todas as articulações que são encontradas pelo Kinect SDK para analisar os padrões de movimento em todos os frames. O autor foca em demonstrar que o método pode identificar pessoas realizando diversas ações (por exemplo, caminhar e correr) em diversos níveis de complexidade e cenário. É utilizada a fusão das medidas antropométricas e de marcha.

3.1.2 Métodos Não Baseados em Poses

Hasan e Babaguchi (2016) propõem uma assinatura antropométrica que descreve o tamanho e a forma do corpo humano. Tanto o tamanho e forma do corpo são estimados com marcos anatômicos detectados pelo Microsoft Kinect. A decisão da identidade dos indivíduos é feita por um classificador treinado com a assinatura antropométrica. Utilizando a base de dados BIWI RGBD-ID o trabalho atingiu 33.3% de taxa de reconhecimento em *rank-1* com imagens de pessoas paradas e 22.1% de taxa de reconhecimento em *rank-1* com pessoas se movimentando.

Velardo e Dugelay (2012) utilizam o peso, tamanho das pernas, tamanho dos braços, tamanho da cintura e altura para realizar um passo de reconhecimento inicial e o resultado desse passo é utilizado para a decisão final com um sistema de reconhecimento de faces. Para avaliar a performance dessa abordagem, uma base de dados é criada, gerando usuários virtuais utilizando-se as medidas antropométricas da base de dados NHANES e as faces dos indivíduos da base de dados FERET. No melhor dos casos, a inclusão das medidas antropométricas aumentou o desempenho do sistema em 20% na taxa de reconhecimento em *rank-1*.

3.2 Métodos Baseados em Marcha Humana

3.2.1 Métodos Baseados em Poses

Lima e Schwartz (2019) utilizaram estimativa de pose para obter as coordenadas das partes do corpo. Essas informações são processadas gerando sinais e histogramas de movimento, os quais são utilizados como características. Para realizar o reconhecimento dos padrões de marcha, os autores utilizam dois métodos. O primeiro método usa SDTW (*Subsequence Dynamic Time Warping*) para comparar os sinais de consulta com os do banco de dados, classificando com base no custo de distância de correspondência mínima, e o segundo método calcula a distância euclidiana dos histogramas de movimento para definir a pessoa do banco de dados que está mais próxima da pessoa sendo identificada. Finalmente, uma fusão de pontuação é usada no resultado desses dois métodos para determinar a identidade da pessoa. A abordagem proposta foi avaliada em todas as visões da base de dados CASIA A (ARAI; ANDRIE, 2001), alcançando uma precisão de reconhecimento acima de 92,5% em todas as visões.

Jangua e Marana (2020) propuseram um método para modelar a marcha humana analisando, a cada *frame* do vídeo, os ângulos e as distâncias das partes do esqueleto do indivíduo à posição do pescoço e construindo um sinal que representa como cada parte do corpo se comporta durante os ciclos de marcha. Para estimar as poses 2D, os autores utilizaram o método OpenPose (CAO et al., 2018b). O método proposto foi avaliado nas bases de dados CASIA Gait Dataset-A (ARAI; ANDRIE, 2001), composta por 20 sujeitos com 12 sequências de vídeo cada (4 sequências para cada posição da câmera: frontal, lateral e oblíqua), e CASIA Gait Dataset-B (YU, S.; TAN, D.; TAN, T., 2006), composta por 124 sujeitos caminhando em três condições diferentes (normal, vestindo um casaco e carregando uma bolsa), com 11 ângulos de visão cada. Os melhores resultados foram obtidos usando a função de distância qui-quadrado, que obteve 95,00% da taxa de reconhecimento de classificação-1 na CASIA Gait Dataset-A e 94,22% da taxa de reconhecimento de classificação-1 na CASIA Gait Dataset-B, que são comparáveis a resultados estado da arte.

FENG, Y., LI, Y. e LUO, J. (2016) desenvolveram o próprio algoritmo de estimativa de pose utilizando CNN para realizar a extração de mapas de calor. Geralmente sequências de marcha são representadas por GEI (*Gait Energy Image*), porém parte das informações temporais são perdidas durante a etapa de cálculo da média. Por este motivo neste trabalho é proposta a obtenção de mapas de calor a partir de *frames*, resultando de estimação de poses. Os autores utilizaram o CASIA Gait Dataset-B para testar o modelo obtendo o resultado de 81.9% de acurácia utilizando imagens em 90° (visão lateral).

3.2.2 Métodos Não Baseados em Poses

O método proposto por Chao et al. (2018) é baseado em uma rede chamada GaitSet, que utiliza conjuntos de estruturas independentes de marcha para aprender informações de identidade. No trabalho, nota-se que o método é robusto à permutação de *frames* e pode integrar *frames* de diferentes vídeos gravados em diferentes cenários. Ele alcançou uma precisão de classificação de *rank-1* de 95 % na base de dados CASIA Gait Dataset A (ARAI; ANDRIE, 2001) em seu melhor caso e 62,5 % no pior caso.

Ben et al. (2019a) construíram uma estrutura que utiliza representação tensorial aplicada ao reconhecimento de marcha em visão cruzada. Foram utilizados três critérios: *Coupled Multi-Linear Locality-Preserved* (CMLP), com a responsabilidade de preservar a estrutura de variedade tensorial, *Coupled Multi-Linear Marginal* (CMMF), com a responsabilidade de codificar compacidade intraclasse e separabilidade interclasse, e *Coupled Multi-Linear Discriminant Analysis* (CMDA), com a responsabilidade de minimizar a dispersão intraclasse e maximizar a dispersão interclasse. Este trabalho também foi validado na base de dados CASIA Gait Dataset A (ARAI; ANDRIE, 2001), obtendo uma precisão de classificação *rank-1* de 99 % no melhor caso e 62 % no pior caso.

O trabalho apresentado por BEN et al. (2019b) utiliza *Couple Patch Alignment* (CPA) para lidar com as mudanças na visão, isso é feito para combinar com diferentes ângulos. Cada captura é feita com uma amostra da marcha. Isso é seguido por uma função objetivo que equilibra as variações de visão cruzada, intra-classe e inter-classe. Os resultados obtidos na base de dados CASIA Gait Dataset A (ARAI; ANDRIE, 2001) variam de 48% a 100% da taxa de reconhecimento de classificação *rank-1*, dependendo do protocolo utilizado.

3.3 Métodos Baseados na Fusão de Características Antropométricas e de Marcha

Andersson, Dutra e Araújo (2014) investigaram o uso de pontos 3D das articulações do esqueleto de indivíduos, capturados por meio do sensor Microsoft Kinect. Esses pontos foram utilizados para extrair informações antropométricas e calcular ângulos descritos pelas articulações inferiores (quadril, joelhos e tornozelos) durante a caminhada, a fim de extrair parâmetros cinemáticos da marcha. O método proposto foi avaliado em uma base de dados própria, contendo dados de 166 indivíduos. Nos experimentos realizados com um pequeno número de indivíduos deste banco de dados, o uso apenas de parâmetros da marcha obteve altos índices de acertos. No entanto, ao se aumentar o número de indivíduos nos experimentos, foram observadas acurácias decrescentes. Os autores relatam que melhores precisões foram obtidas quando foram utilizadas medidas antropométricas para complementar as informações da marcha.

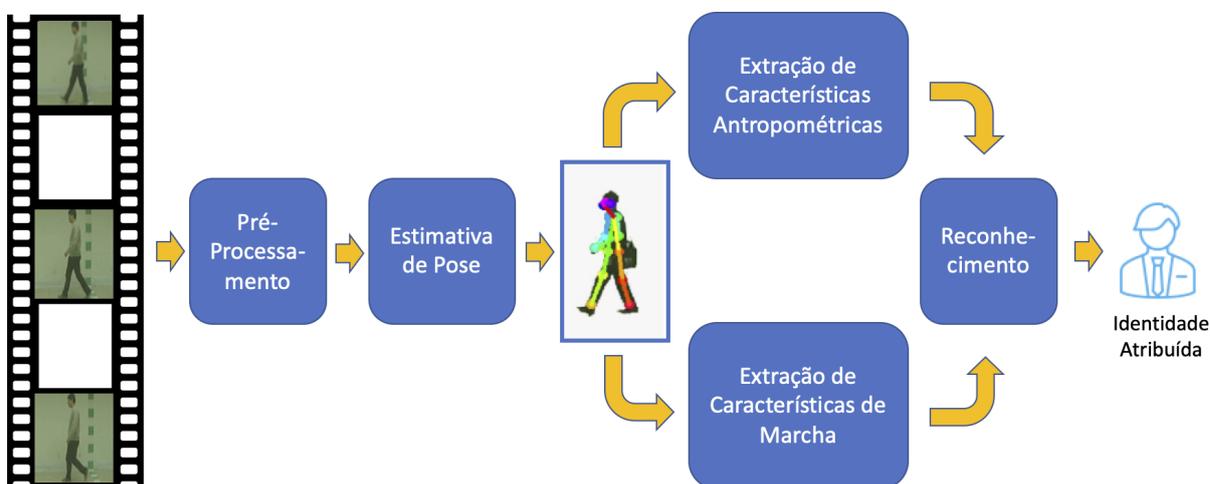
Liao et al. (2020) propuseram o PoseGait, um novo método de reconhecimento de marcha combinando um conjunto de características *hand-crafted* e uma CNN para extração de características temporais e espaciais de alto nível. Esse conjunto de características é definido por: pontos de junta estimados a partir de um algoritmo de estimativa de pose 3D, ângulos α e β entre a junta estimada para o ombro esquerdo e o cotovelo esquerdo, medidas dos membros a partir da pose estimada e o movimento das juntas. Os autores validam o método nas bases de dados CASIA B e E. Na CASIA B, o método consegue atingir, em média, 96.62% de taxa de reconhecimento quando o *probe* e a *gallery* são imagens na mesma angulação e o *probe* não possui nenhum tipo de variação, 71.25% com o *probe* tendo variações de objetos carregados e 54.17% com variações de roupa. Nos experimentos com a CASIA E, é possível confirmar que roupas afetam o desempenho da abordagem e que existe uma grande variação nas taxas de reconhecimento, indo de 91% nos casos sem nenhuma variação até menos de 10% nos casos com variação de roupa.

4 Método Proposto

Neste capítulo apresenta-se o método proposto nesta dissertação para a identificação de pessoas baseada em características antropométricas e de marcha, obtidas de poses 2D estimadas em imagens de vídeos. Nas aplicações alvo do método proposto, o objetivo é identificar uma pessoa caminhando em uma área sob monitoramento. Assim, a informação de entrada para o sistema biométrico será uma sequência de quadros de vídeo capturados por uma câmera de vigilância. Assume-se que o sistema biométrico terá em seu banco de dados (*gallery*) características biométricas obtidas de pelo menos um vídeo por pessoa, que serão comparadas com as características obtidas dos vídeos de consulta (*probe*).

A Figura 16 apresenta um diagrama do método proposto. Como pode-se observar, o método possui quatro etapas principais: pré-processamento, estimativa de pose, extração de características e reconhecimento do indivíduo, que estão descritas nas Seções 4.1, 4.2, 4.3 e 4.4, respectivamente.

Figura 16 – Diagrama de blocos do método proposto para identificação de pessoas baseada em medidas antropométricas (características *Soft Biometrics*) e características de marcha, obtidas de poses 2D.



Fonte: Elaborada pelo autor.

4.1 Pré-Processamento

Na primeira etapa do método realizam-se alguns pré-processamentos visando melhorar a qualidade das imagens a fim de se obter melhores estimativas de poses na etapa subsequente.

Inicialmente, a imagem de entrada é subtraída da imagem de fundo a fim de manter apenas as informações de interesse, no caso, a imagem da pessoa. Após, aplica-se o filtro de *Canny*, um filtro de convolução que suaviza o ruído e realça as bordas, combinando um filtro Gaussiano com um operador diferencial (GONZALEZ; WOODS, 2008). Na sequência, aplica-se um operador de fechamento da Morfologia Matemática visando também à eliminação de ruídos. O fechamento de uma imagem I pelo elemento estruturante S consiste na dilatação de I por S seguida da erosão do resultado por S (GONZALEZ; WOODS, 2008), conforme definido na equação 8:

$$I \bullet S = (I \oplus S) \ominus S \quad (8)$$

em que \oplus corresponde ao operador de dilatação e \ominus corresponde ao operador de erosão da Morfologia Matemática.

A Figura 17 mostra o resultado obtido após a etapa de pré-processamento de um *frame* do vídeo de entrada.

Figura 17 – Exemplo da aplicação do pré-processamento em um *frame* do vídeo de entrada.



Fonte: Elaborada pelo autor.

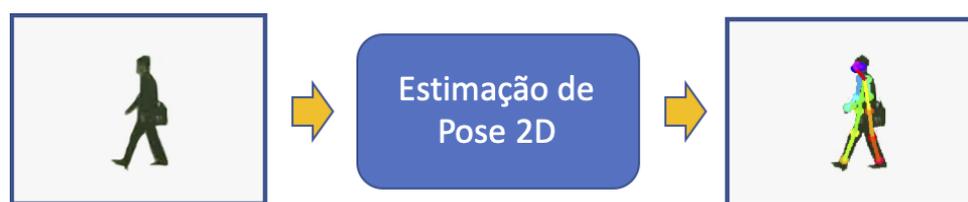
Nos experimentos realizados (vide Seção 6.2), observou-se que com a aplicação das operações de pré-processamento, o algoritmo de estimativa de pose apresenta melhores resultados do que quando não é aplicado nenhum pré-processamento.

4.2 Estimativas de Pose

Na segunda etapa do método, estimam-se as poses nas imagens pré-processadas utilizando-se um método de estimativa de pose 2D a fim de se obter os esqueletos das pessoas na cena. Neste trabalho, foram avaliados os métodos OpenPose (CAO et al., 2018b), PifPaf (KREISS; BERTONI; ALAHI, 2019a) e Tf-Pose-Estimation², descritos na Seção 2.5, porém, quaisquer outros métodos de estimação de poses 2D poderão ser utilizados. A Figura 18 mostra um exemplo de pose 2D estimada pelo método PifPaf, a partir de uma imagem obtida após o pré-processamento.

² <https://github.com/ZheC/tf-pose-estimation>.

Figura 18 – Exemplo de pose 2D estimada a partir de uma imagem pré-processada.



Fonte: Elaborada pelo autor.

4.3 Extração de Características

Na terceira etapa do método, a partir das poses 2D estimadas, extraem-se medidas antropométricas e as características de marcha que são utilizadas para a identificação biométrica das pessoas na etapa subsequente. As Subseções 4.3.1 e 4.3.2 descrevem, respectivamente, detalhes da extração das características antropométricas e de marcha utilizadas no método proposto.

4.3.1 Medidas Antropométricas

Neste trabalho, propõe-se utilizar oito medidas antropométricas para a identificação biométrica das pessoas, por meio de características *soft*. A primeira medida é o comprimento do úmero direito (UmD) do indivíduo, calculado medindo-se a distância entre as articulações 0 (ombro direito) e 2 (cotovelo direito) do esqueleto, utilizando-se o modelo apresentado na Figura 1. A segunda medida é o comprimento do úmero esquerdo (UmE), calculado medindo-se a distância entre as articulações 1 (ombro esquerdo) e 3 (cotovelo esquerdo). A terceira medida é o comprimento da Ulna direita (UD), calculado medindo-se a distância entre as articulações 2 (cotovelo direito) e 4 (punho direito). A quarta medida é o comprimento da Ulna esquerda (UE), calculado medindo-se a distância entre as articulações 3 (cotovelo esquerdo) e 5 (punho esquerdo). A quinta medida é o comprimento do fêmur direito (FD), calculado medindo-se a distância entre as articulações 6 (anca direita) e 8 (joelho direito). A sexta medida é o comprimento do fêmur esquerdo (FE), calculado medindo-se a distância entre as articulações 7 (anca esquerda) e 9 (joelho esquerdo). A sétima medida é o comprimento da tíbia direita (TD), calculado medindo-se a distância entre as articulações 8 (joelho direito) e 10 (tornozelo direito). Finalmente, a oitava medida é o comprimento da tíbia esquerda (TE), calculado medindo-se a distância entre as articulações 9 (joelho esquerdo) e 11 (tornozelo esquerdo). A Figura 19 apresenta as oito medidas antropométricas utilizadas neste trabalho, obtidas a partir das poses 2D de cada *frame* dos vídeos de entrada.

É importante destacar que todos os comprimentos são normalizados, pois os mesmos são alterados conforme o indivíduo se movimenta na cena, podendo ficar mais próximo ou mais distante da câmera. Esta normalização é realizada dividindo-se todas as medidas pela distância entre o pescoço e a anca.

Figura 19 – Medidas antropométricas obtidas das poses 2D dos indivíduos, utilizadas pelo método proposto.



Fonte: Elaborada pelo autor.

Como no método proposto a identificação do indivíduo é baseada nos vários *frames* do vídeo de entrada, e como pode haver alterações nas medidas de cada parte do esqueleto, nos diversos *frames* do vídeo, decorrentes de oclusões das articulações ou de imprecisões na localização exata das articulações, foram propostas três estratégias para a etapa de obtenção de um descritor para o vídeo, a partir dos seus vários *frames*: média das medidas antropométricas, concatenação das medidas antropométricas e histogramas das medidas antropométricas, descritas nas Subseções 4.3.1.1, 4.3.1.2 e 4.3.1.3, respectivamente.

Um aspecto importante do método proposto diz respeito às oclusões de partes do corpo que podem ocorrer durante a captura das imagens de vídeo. Quando ocorrem oclusões, os métodos de estimativa de pose não são capazes de localizar perfeitamente todas as juntas. Como os métodos OpenPose e PifPaf retornam uma medida de qualidade de cada junta detectada, no método proposto utiliza-se uma função que determina um limite de aceitação de uma determinada junta extraída. Desse modo, utiliza-se uma junta apenas se a medida de qualidade da detecção dessa junta for superior a 70% (este valor foi definido experimentalmente).

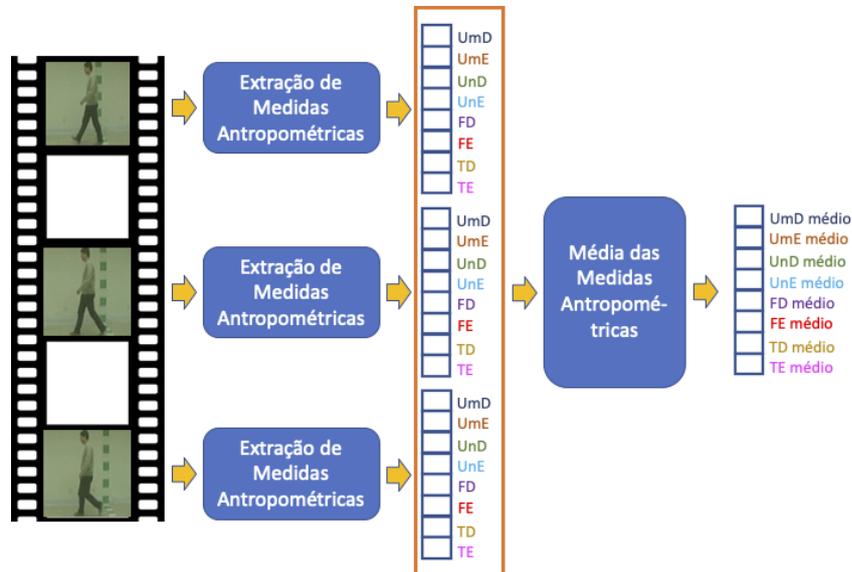
4.3.1.1 Média das Medidas Antropométricas

Nesta estratégia, o vetor de características antropométricas utilizado para representar todo o vídeo consiste no cálculo do comprimento médio de cada parte do esqueleto (úmero direito (UmD) e esquerdo (UmE), ulna direita (UnD) e esquerda (UnE), fêmur direito (FD) e esquerdo (FE), e tíbia direita (TD) e esquerda (TE)) em cada um dos n quadros do vídeo, conforme ilustra a Figura 20.

4.3.1.2 Concatenação das Medidas Antropométricas

Nesta estratégia, o vetor de características antropométricas utilizado para representar todo o vídeo consiste na concatenação dos vetores contendo os comprimentos de cada parte do esqueleto (úmero direito (UmD) e esquerdo (UmE), ulna direita (UnD) e esquerda (UnE),

Figura 20 – Vetor de características composto pela média dos comprimentos de cada parte do esqueleto humano: Úmero Direito (UmD), Úmero Esquerdo (UmE), Ulna Direita (UnD), Ulna Esquerda (UnE), Fêmur Direito (FD), Fêmur Esquerdo (FE), Tíbia Direita (TD) e Tíbia Esquerda (TE), em cada um dos n quadros do vídeo.



Fonte: Elaborada pelo autor.

fêmur direito (FD) e esquerdo (FE), e tíbia direita (TD) e esquerda (TE)) em cada um dos n quadros do vídeo, conforme ilustra a Figura 21.

4.3.1.3 Histogramas das Medidas Antropométricas

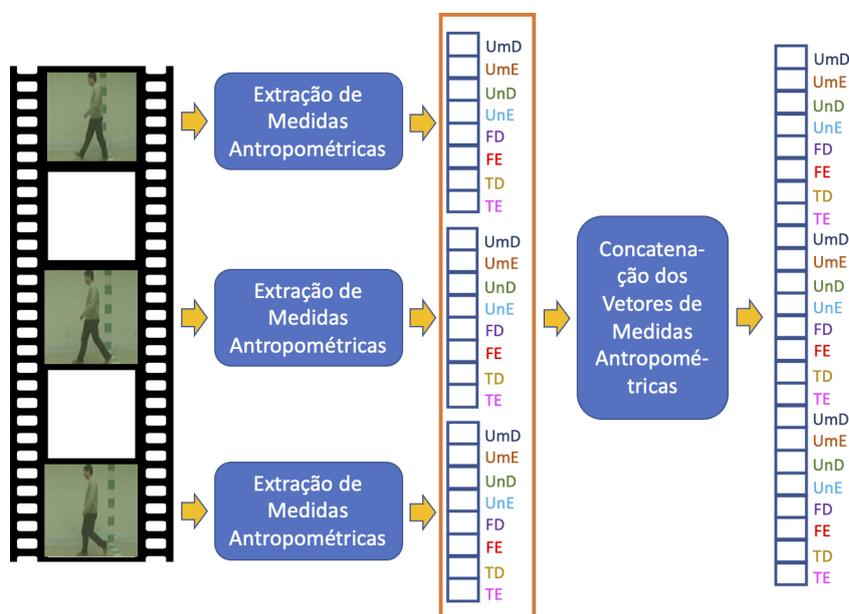
Nesta estratégia, o vetor de características antropométricas utilizado para representar todo o vídeo consiste no cálculo de histogramas dos comprimentos de cada parte do esqueleto: úmero direito (UmD) e esquerdo (UmE), ulna direita (UnD) e esquerda (UnE), fêmur direito (FD) e esquerdo (FE), e tíbia direita (TD) e esquerda (TE), e a subsequente concatenação destes histogramas, conforme ilustra a Figura 22.

4.3.2 Características de Marcha

Neste trabalho, propõe-se a extração de características da marcha humana, também obtidas a partir das poses 2D, para serem utilizadas em conjunto com as medidas antropométricas para a identificação das pessoas. Há na literatura alguns métodos propostos para esta finalidade, como os métodos de Lima e Schwartz (2019) e Jangua e Marana (2020), que são baseados nas distâncias entre as articulações ou nos ângulos entre as partes do esqueletos.

Inspirados no trabalho de FENG, Y., LI, Y. e LUO, J. (2016), que desenvolveram o próprio algoritmo de estimativa de pose utilizando CNN para realizar a extração de mapas de calor, utilizam-se, neste trabalho, os mapas de calor das articulações gerados pelo método

Figura 21 – Vetor de características composto pela concatenação dos vetores contendo os comprimentos de cada parte do esqueleto humano: Úmero Direito (UmD), Úmero Esquerdo (UmE), Ulna Direita (UnD), Ulna Esquerda (UnE), Fêmur Direito (FD), Fêmur Esquerdo (FE), Tíbia Direita (TD) e Tíbia Esquerda (TE), em cada um dos n quadros do vídeo.



Fonte: Elaborada pelo autor.

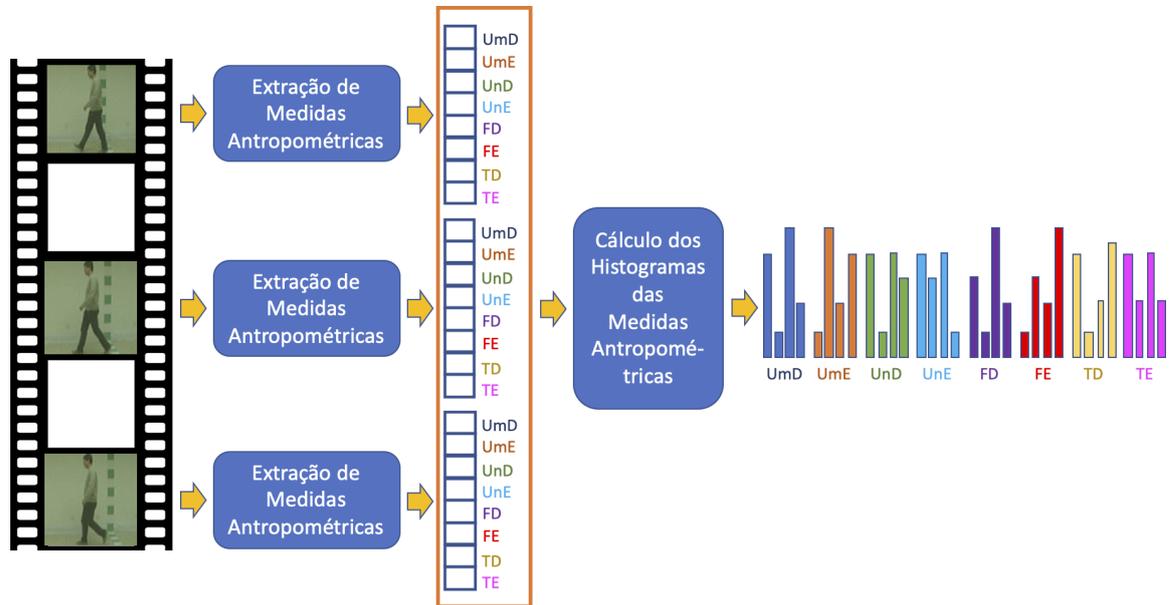
de estimativa de poses 2D em cada *frame* do vídeo, como característica biométrica para o reconhecimento de marcha³.

A Figura 23 apresenta um diagrama do método proposto para a extração das características da marcha a partir dos mapas de calor gerados na etapa de estimação de poses 2D. Pode-se observar que o método proposto, que possui uma etapa de estimativa de pose e outra de extração de características, utiliza o mapa de calor médio obtido a partir dos mapas de calor das articulações de cada *frame* do vídeo.

Na primeira etapa, aplica-se nos *frames* do vídeo de entrada um método de estimativa de pose 2D, como OpenPose (CAO et al., 2018b), PifPaf (KREISS; BERTONI; ALAHI, 2019a) ou TF-Pose-Estimation, a fim de se obter os mapas de calor das juntas em cada quadro do vídeo. Na segunda etapa, um mapa de calor médio das juntas é calculado a partir dos mapas de calor de cada *frame* do vídeo. Este mapa de calor médio constitui o descritor de marcha do indivíduo naquele vídeo.

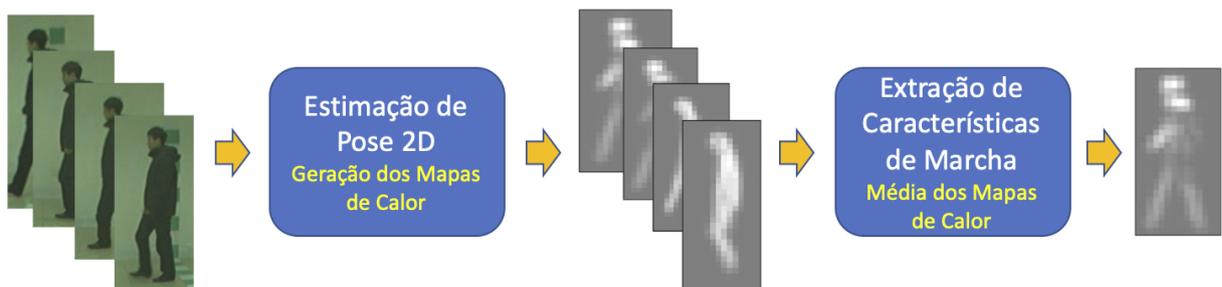
³ Detalhes sobre como os mapas de calor são gerados pelos métodos OpenPose e PifPaf podem ser obtidos em (CAO et al., 2018a) e Kreiss, Bertoni e Alahi (2019b), respectivamente.

Figura 22 – Vetor de características composto pela concatenação dos histogramas dos comprimentos das partes do esqueleto humano: Úmero Direito (UmD), Úmero Esquerdo (UmE), Ulna Direita (UnD), Ulna Esquerda (UnE), Fêmur Direito (FD), Fêmur Esquerdo (FE), Tíbia Direita (TD) e Tíbia Esquerda (TE).



Fonte: Elaborada pelo autor.

Figura 23 – Diagrama do método proposto para a extração das características de marcha humana baseado nos mapas de calor das articulações gerados na etapa de estimação de poses 2D. O mapa de calor médio é utilizado como o descritor de marcha do vídeo.



Fonte: Elaborado pelo autor.

4.4 Reconhecimento

Na quarta etapa do método proposto neste trabalho para a identificação biométrica de pessoas por meio de poses 2D, os métodos baseados nas medidas antropométricas e nos padrões de marcha são fundidos, visando, deste modo, aumentar as taxas de reconhecimento e tornar o processo de identificação mais robusto. As estratégias de fusão usadas neste trabalho são no nível de características ou no nível de pontuação, descritas na Seção 2.4.

Como já foi mencionado, nas aplicações alvo do método proposto neste trabalho, o objetivo é identificar uma pessoa caminhando em uma área sob vigilância. Assim, a informação de entrada para o sistema biométrico será uma sequência de quadros de vídeo capturados por uma câmera de vigilância. Assume-se que o sistema biométrico terá em seu banco de dados (*gallery*) características antropométricas e de marcha obtidas de pelo menos um vídeo por pessoa, e estas características são comparadas às características obtidas dos vídeo de consulta (*probe*) no momento da identificação.

Caso a identificação do indivíduo ocorra no modo de verificação, as distâncias entre os vetores de características antropométricas e de marcha obtidos do vídeo *probe* e os respectivos vetores de características obtidos do vídeo *gallery* associado à identidade reivindicada, deverão ser inferiores a um determinado valor de limiar (neste caso, tem-se uma comparação 1:1).

Caso a identificação do indivíduo ocorra no modo de reconhecimento, os vetores de características antropométricas e de marcha obtidos do vídeo *probe* são comparados a todos os vetores de características obtidos dos vídeos *gallery*, sendo a identidade do indivíduo associada à identidade referente ao vídeo *gallery*, cujos vetores de características apresentarem a menor distância com os vetores de características obtidas do vídeo *probe* (neste caso, tem-se uma comparação 1:n).

A forma como os vetores de características são comparados, bem como as funções de distâncias utilizadas nas comparações, variam de acordo com as particularidades dos vetores de características das medidas antropométricas e de marcha propostos neste trabalho. As próximas subseções detalham os processos de comparação propostos.

4.4.1 Medidas Antropométricas

Nesta subseção, descreve-se como é feito o reconhecimento dos indivíduos quando são utilizadas as medidas antropométricas, ou estatísticas destas medidas, nos vetores de características.

4.4.1.1 Média das Medidas Antropométricas

Neste trabalho, quando o vetor de características extraído do vídeo é composto pelos comprimentos médios de cada uma das partes do esqueleto obtido em cada um dos n frames do vídeo, utiliza-se o classificador 1-NN, definido na Seção 2.8.1, com a distância euclidiana ou *city-block*, definidas na Seção 2.7, para o reconhecimento dos indivíduos.

4.4.1.2 Concatenação das Medidas Antropométricas

Neste trabalho, quando o vetor de características do vídeo v é composto pela concatenação dos n vetores de características v_f , que contém os comprimentos das oito partes do esqueleto obtido em cada um dos n frames f do vídeo por um método de estimativa de poses

2D, utiliza-se o classificador 1-NN, definido na Seção 2.8.1, com a função de distância definida na equação 9, para o reconhecimento dos indivíduos.

$$D(v_p, v_g) = \sum_{i=1}^k d_i(v_{f_p}, v_{f_g}) \quad (9)$$

em que $d_i(v_{f_p}, v_{f_g})$ é a i -ésima menor distância euclidiana (ou *city-block*), obtida entre vetores de características v_{f_p} e v_{f_g} , obtidos de *frames* dos vídeos *probe*, v_p , e *gallery*, v_g , respectivamente, e k é uma constante cujo valor é obtido empiricamente entre 1 e n .

4.4.1.3 Histogramas das Medidas Antropométricas

Neste trabalho, quando o vetor de características extraído do vídeo é composto pela concatenação dos histogramas dos comprimentos de cada uma das oito partes do esqueleto obtido em cada um dos n *frames* do vídeo, utiliza-se o classificador 1-NN, definido na Seção 2.8.1, com a distância qui-quadrado (χ^2) ou a distância euclidiana, definidas na Seção 2.7, para o reconhecimento dos indivíduos.

4.4.2 Características de Marcha

Neste trabalho, quando o vetor de características extraído do vídeo é constituído pelo mapa de calor médio obtido calculando-se a média dos mapas de calor das articulações estimadas em cada um dos n *frames* do vídeo, utiliza-se um classificador baseado em redes neurais artificiais, mais especificamente uma *Multilayer Perceptron* (MLP), conforme descrito na Seção 2.8.2.1, para o reconhecimento dos indivíduos.

5 Bases de Dados

Neste trabalho, foram utilizadas três bases de dados para avaliação do método proposto para identificação de pessoas baseada em características antropométricas e de marcha extraídas de poses 2D, sendo duas públicas, CASIA Gait Dataset-A (TAN et al., 2006) e CASIA Gait Dataset-B, e uma base de dados privada, contendo vídeos fornecidos pela Petrobrás, empresa financiadora deste trabalho.

5.1 CASIA Gait Dataset-A

O conjunto de dados CASIA Gait Dataset-A (antigo NLPR Gait Database) (TAN et al., 2006) foi criado em 2001, contendo vídeos de 20 pessoas caminhando. Para cada pessoa tem-se 12 sequências de imagens (vídeos), sendo 4 sequências para cada uma das três direções: 0 graus (paralela), 45 graus (diagonal) e 90 graus (perpendicular) ao plano da imagem. Os vídeos desta base de dados possuem entre 37 a 127 *frames*. O tamanho total do CASIA Gait Dataset-A é de cerca de 2,2 GB e o banco de dados inclui 19139 imagens (*frames*). A Figura 24 apresenta três imagens do CASIA Gait Dataset-A, sendo as imagens à esquerda, centro e à direita obtidas de vídeos capturados na direção paralela, perpendicular e diagonal à câmera, respectivamente.

Figura 24 – Imagens do banco de dados CASIA Gait Dataset-A de pessoas caminhando nas direções paralela (imagem da esquerda), perpendicular (imagem central) e diagonal (imagem da direita) com relação ao plano da câmera.



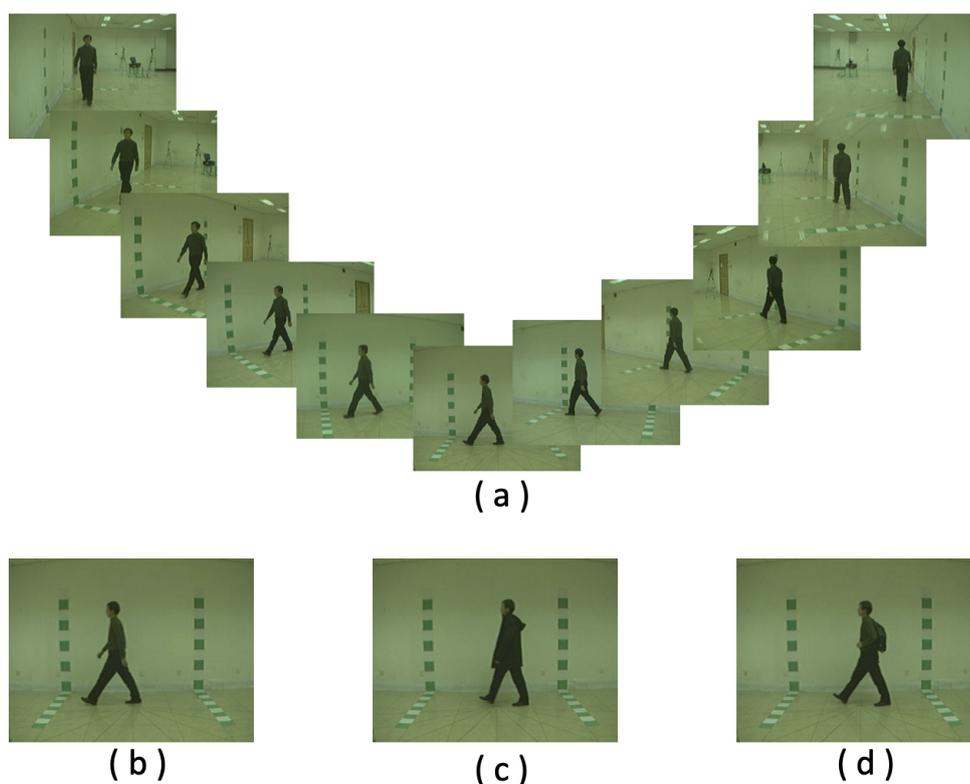
Fonte: (TAN et al., 2006) .

5.2 CASIA Gait Dataset-B

O conjunto de dados CASIA Gait Dataset-B (WANG et al., 2005; YU, S.; TAN, D.; TAN, T., 2006) foi criado em 2005, contendo vídeos capturados simultaneamente por 11 câmeras de 124 indivíduos caminhando em um ambiente. Trata-se, portanto, de um conjunto

de dados com múltiplas visões do mesmo dado. Para aumentar a variabilidade intraclasse desta base de dados, foram obtidos vídeos dos mesmos indivíduos caminhando, porém carregando objetos (bolsa ou mochila) e vestindo casaco. A Figura 25 apresenta imagens do CASIA Gait Dataset-B. Na Figura 25(a) pode-se observar as onze imagens capturadas pelas onze câmeras de um mesmo indivíduo durante uma caminhada. Na Figura 25(b)-(d) pode-se observar imagens de indivíduos sem objetos e sem casaco, com casaco e sem objeto e com objeto e sem casaco, respectivamente.

Figura 25 – Exemplos de imagens do banco de dados CASIA Gait Dataset-B; (a) Múltiplas visões da marcha, capturadas por 11 câmeras; (b) Marcha normal do indivíduo; (c) Marcha com o indivíduo vestindo casaco; (d) Marcha com o indivíduo carregando uma mochila.



Fonte: (YU, S.; TAN, D.; TAN, T., 2006).

Para os experimentos realizados neste trabalho foram utilizados 1240 vídeos da vista perpendicular (ângulo de visão igual a 90 graus) do CASIA Gait Dataset-B, com as seguintes informações:

- Os vídeos possuem em média 66 quadros (*frames*);
- De cada quadro foram detectados 14 pontos de interesse (articulações dos esqueletos estimados);

- Para cada indivíduo foram utilizados os vídeos com as três variações (pessoa caminhando sem objetos e sem casaco, pessoa carregando bolsa e pessoa vestindo casaco);
- Para a variação da pessoa utilizando casaco são disponibilizados 2 vídeos por pessoa;
- Para a variação da pessoa utilizando mochila são disponibilizados 2 vídeos por pessoa;
- Para a variação da pessoa sem objetos são disponibilizados 6 vídeos por pessoa;

5.3 Base de Dados Privada

A base de dados privada utilizada neste trabalho foi fornecida pela Petrobrás e consiste em quatro vídeos, com cerca de 20 minutos cada, nos quais aparecem até cinco trabalhadores caminhando em uma das instalações da empresa. Nesses vídeos, os trabalhadores aparecem utilizando equipamentos de proteção individual (EPI), compostos por: macacão, capacete, luvas e óculos de segurança. A Figura 26 apresenta um quadro extraído de um dos vídeos dessa base de dados.

Como pode-se observar na Figura 26, esta base de dados é bastante desafiadora, pois os vídeos foram capturados em um ambiente não controlado, à distância, com várias pessoas presentes simultaneamente na cena, trajando uniformes e utilizando EPI (equipamentos de proteção individual) que dificultam a identificação das pessoas mesmo para observadores humanos.

Esta base de dados foi rotulada por um observador humano, com o auxílio do Deep SORT (*Simple Online and Realtime Tracking*), um método de rastreamento desenvolvido por WOJKE, BEWLEY e PAULUS (2017), para facilitar a identificação dos trabalhadores nas cenas dos vídeos. A Seção 5.3.1 apresenta detalhes do Deep SORT.

Dadas as características dos vídeos, em alguns trechos o algoritmo de rastreamento gerou falsos positivos ou falsos negativos, detectando algo que não seja um humano ou não detectando um humano presente na cena. Porém, como já foi mencionado, o algoritmo de rastreamento foi utilizado para auxiliar a tarefa de anotação, que foi realizada por um observador humano.

A Figura 27 mostra duas imagens extraídas de um dos vídeos da base de dados privada em que é possível observar os *bounding boxes* ao redor dos trabalhadores detectados pelo Deep SORT, descrito na seção 5.3.1, e os respectivos esqueletos no interior, estimados pelo algoritmo de estimação de poses.

Cada um dos vídeos foram separados em três trechos: treinamento (75% dos *frames* iniciais dos vídeos), teste (20% dos *frames* seguintes dos vídeos) e validação (5% dos *frames* finais dos vídeos). Foram utilizados apenas trechos dos vídeos em que existem trabalhadores na cena. Esses trechos correspondem a 76% de todos os vídeos.

Figura 26 – Quadro de um dos vídeos da base de dados privada, disponibilizada pela Petrobras, na qual é possível observar funcionários em sua rotina de trabalho. As características do ambiente, as oclusões, a distância da câmera e a utilização de equipamentos de proteção pelos funcionários dificultam sobremaneira o processo de identificação.



Fonte: Elaborada pelo autor.

Figura 27 – Exemplos da aplicação do algoritmo Deep SORT no processo de rastreamento e anotação de indivíduos nos vídeos da base de dados privada.



Fonte: Elaborada pelo autor.

Os vídeos que compõem esta base de dados têm as seguintes características:

- Vídeo 1: Este vídeo tem duração de 23 minutos e contém imagens de três trabalhadores. 81% do vídeo foi aproveitado para uso no *dataset*;
- Vídeo 2: Este vídeo tem duração de 24 minutos e contém imagens de quatro trabalhadores. 76% do vídeo foi aproveitado para uso no *dataset*;

- Vídeo 3: Este vídeo tem duração de 21 minutos e contém imagens de quatro trabalhadores. 85% do vídeo foi aproveitado para uso no *dataset*;
- Vídeo 4: Este vídeo tem duração de 16 minutos e contém imagens de cinco trabalhadores. 51% do vídeo foi aproveitado para uso no *dataset*.

5.3.1 Deep SORT

O Deep SORT (*Simple Online and Realtime Tracking*) é um método de código aberto proposto por WOJKE, BEWLEY e PAULUS (2017) para rastreamento 2D de múltiplos objetos em imagens (*frames*) de vídeo. Trata-se de uma abordagem estendida do SORT (BEWLEY et al., 2016), um método anterior que tinha o mesmo intuito.

O Deep SORT é considerado um detector de pessoas estado da arte que permite o rastreamento e a reidentificação de objetos na cena após longos períodos de oclusão, integrando, para tanto, informações de aparência dos objetos. Este método tem obtido resultados experimentais que competem com outros métodos como o próprio SORT em acurácia e precisão do rastreamento, se destacando em poucas trocas de identidade dos objetos durante o processo. O Deep SORT adota uma metodologia de rastreamento de hipótese única, utilizando filtragem recursiva de Kalman e associação *frame a frame* de dados, seguindo as componentes chave do SORT que são: i) detecção, realizada por redes neurais convolucionais, responsáveis por extrair características e propor regiões para classificar os objetos presentes nela; ii) modelo de estimação, representação do objeto identificado e o modelo de movimento para propagação da identidade para o próximo *frame*, atribuindo uma caixa delimitadora, modelada através da representação vertical e horizontal do pixel, área e proporção em tela do alvo, usada para atualização do estado através de filtragem de Kalman; iii) identidades de rastreamento, são criados rastreadores para qualquer detecção com sobreposição menor que uma distância entre cada detecção e caixas delimitadoras previstas, e encerrados caso não haja detecção após uma certa quantidade de *frames*.

O Deep SORT trabalha com algumas modificações dos princípios mencionados. Quando surge uma hipótese de um novo rastreamento não associado a detecções anteriores, este é classificado como uma tentativa de novo rastreamento, sendo analisado durante os primeiros três *frames* após sua concepção. Durante esse tempo, o algoritmo espera associar de maneira assertiva as métricas da modelagem, tentando associá-las a detecção do indivíduo e, caso não seja possível, a tentativa do rastreamento é descartada. A correspondência do rastreamento é feita em cascata, priorizando objetos vistos frequentemente e há implementação de descritores de aparência nas CNNs, visando reduzir trocas no rastreamento de entidades, bem como o número de falsos positivos.

6 Resultados Experimentais

Neste capítulo são apresentados os resultados experimentais obtidos com os métodos propostos neste trabalho para identificação de pessoas a partir de medidas antropométricas e características de marcha obtidas de poses 2D.

6.1 Avaliação dos Métodos de Estimação de Pose

Inicialmente, foi realizado um experimento sobre o subconjunto de vídeos da base de dados CASIA Gait Dataset-B capturados a 90 graus do plano de captura das imagens (ou seja, com a vista lateral dos indivíduos) utilizando-se os descritores antropométricos (estratégia de concatenação) e os descritores de marcha (média dos mapas de calor das articulações), propostos neste trabalho, com o objetivo de avaliar os desempenhos destes descritores obtidos a partir das poses 2D estimadas por três métodos propostos na literatura: OpenPose, PifPaf e Tf-Pose-Estimation (esses métodos são descritos na Seção 2.5).

A RNA-MLP utilizada neste experimento para a identificação dos indivíduos por meio dos descritores de marcha era composta por uma camada de entrada, três camadas ocultas e uma camada de saída, sendo utilizada a estratégia *feedforward* para o treinamento. Esta RNA foi implementada usando o Tensorflow (ABADI et al., 2015) e o Keras.

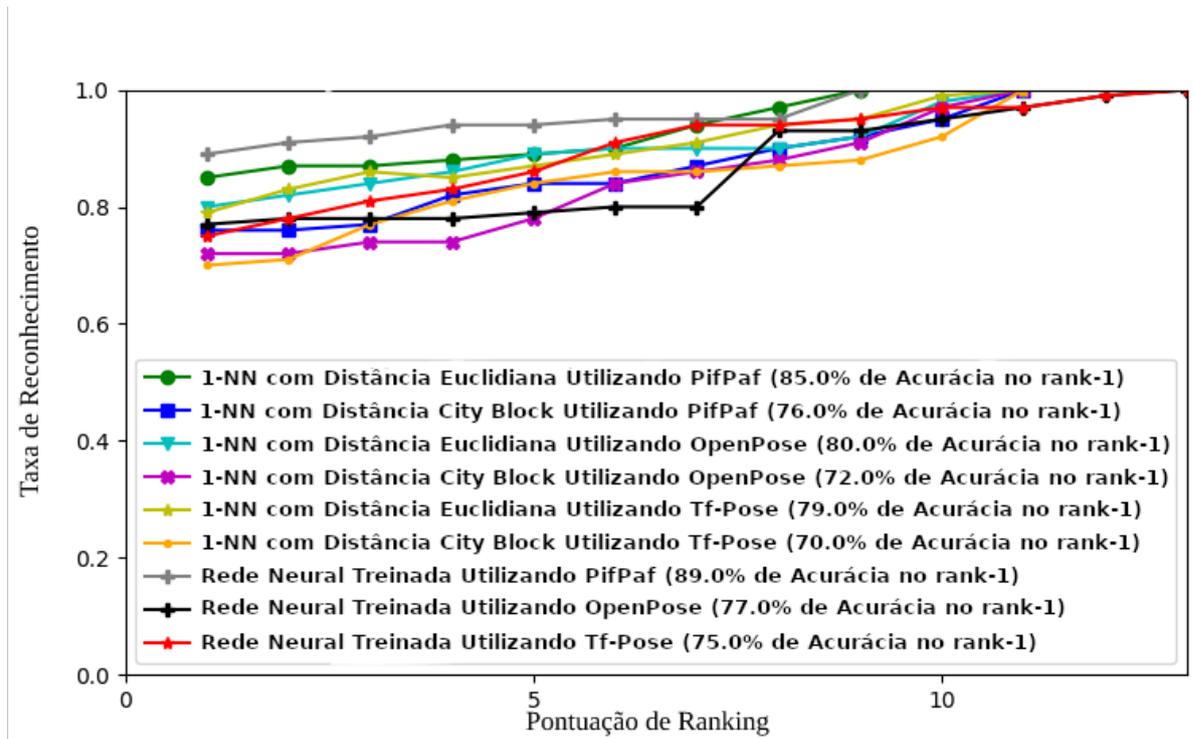
Para a identificação dos indivíduos por meio dos descritores antropométricos, foi utilizado o classificador 1-NN com as funções de distância euclidiana e *city block*.

A Figura 28 apresenta as curvas CMC (Cumulative Match Characteristics) obtidas neste experimento. Como já foi mencionado, a curva CMC é uma medida comumente usada para avaliar o desempenho de sistemas biométricos que operam no modo de identificação (comparações 1: n).

Como pode-se observar na Figura 28, o melhor resultado em relação à taxa de acurácia rank-1 (89%) foi obtido com o método de estimativa de pose PifPaf, utilizando-se as características de marcha (obtidas dos mapas de calor das juntas) com o classificador RNA. O segundo melhor resultado (85%) também foi obtido pelo método PifPaf, mas com as características antropométricas (estratégia de concatenação) e com o classificador 1-NN, utilizando-se a distância euclidiana. O pior resultado (70%) foi obtido com o método TF-Pose-Estimation, com as características antropométricas (estratégia de concatenação) e o classificador 1-NN, utilizando-se a distância *city block*.

Em relação ao desempenho geral medido pelas áreas sob as curvas (AUC, do inglês Area Under the Curve), conforme mostra a Tabela 4, o melhor resultado (0,92) também foi obtido com o método de estimativa de pose PifPaf, utilizando-se as características de marcha

Figura 28 – Curvas CMC obtidas utilizando-se os descritores antropométricos (estratégia da concatenação) e os descritores de marcha obtidos de poses 2D estimadas por três métodos de estimativa de pose: PifPaf, OpenPose e Tf-Pose-Estimation, no subconjunto de dados CASIA Gait Dataset-B com vista lateral - 90°, utilizando-se um classificador RNA para os descritores de marcha e o classificador 1-NN, com as distâncias euclidiana e *city-block*, para os descritores antropométricos.



Fonte: Elaborada pelo autor.

obtidas dos mapas de calor das juntas com o classificador RNA. E o segundo melhor resultado (0,86) também foi obtido pelo método PifPaf, mas com as características antropométricas (estratégia de concatenação) e com o classificador 1-NN, utilizando-se a distância euclidiana. Por fim, o pior resultado geral (0,69) também foi obtido com o método TF-Pose-Estimation, com as características antropométricas (estratégia de concatenação) e o classificador 1-NN, utilizando-se a distância *city block*.

Um segundo experimento foi realizado de forma análoga ao primeiro, porém, utilizando-se a estratégia da média das medidas antropométricas. A Figura 29 mostra os resultados obtidos com este experimento. Pode-se observar nesta figura que os resultados confirmam o melhor desempenho do método PifPaf em relação ao OpenPose. Quanto aos resultados obtidos com os descritores antropométricos, observa-se que a estratégia da concatenação apresenta resultados um pouco superiores do que a estratégia da média. Porém, deve-se observar que a estratégia da média requer um tempo menor de processamento do que a estratégia da concatenação, uma vez que esta última requer $n \times m$ comparações, sendo n e m as quantidades de *frames*

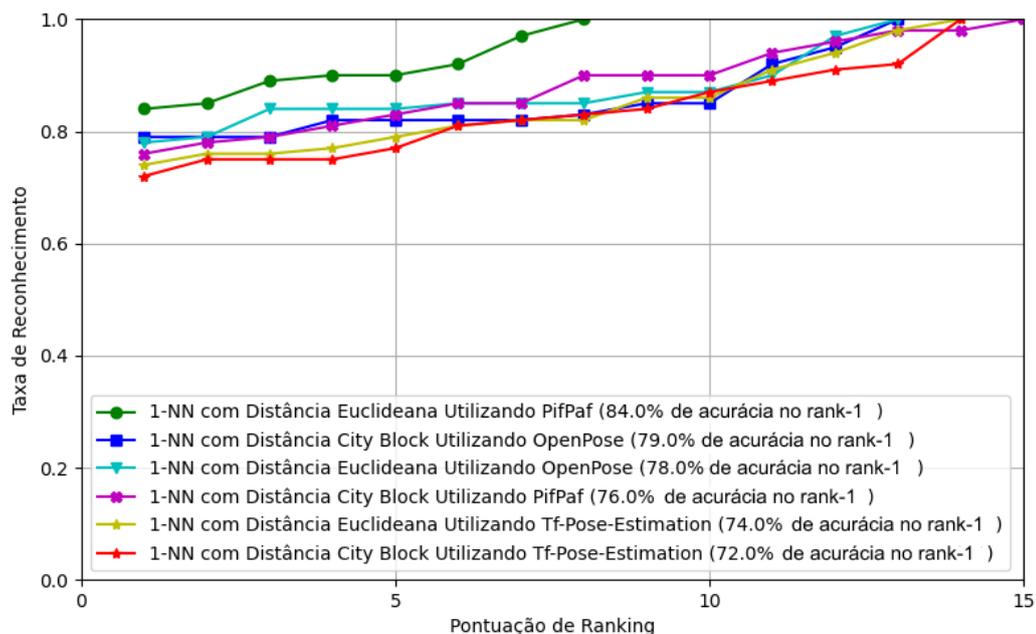
Tabela 4 – Valores da AUC referentes às curvas CMC apresentadas na Figura 28.

Método	AUC
1-NN com Distância Euclidiana Utilizando PifPaf	0,86
1-NN com Distância City Block Utilizando PifPaf	0,76
1-NN com Distância Euclidiana Utilizando OpenPose	0,79
1-NN com Distância City Block Utilizando OpenPose	0,72
1-NN com Distância Euclidiana Utilizando Tf-Pose-Estimation	0,78
1-NN com Distância City Block Utilizando Tf-Pose-Estimation	0,69
Rede Neural Treinada Utilizando PifPaf	0,92
Rede Neural Treinada Utilizando OpenPose	0,79
Rede Neural Treinada Utilizando Tf-Pose-Estimation	0,82

Fonte: Elaborada pelo autor.

dos dois vídeos sendo comparados, ao passo que a estratégia da média requer apenas uma comparação entre os vetores de características médios de cada vídeo.

Figura 29 – Curvas CMC obtidas utilizando-se os descritores antropométricos (estratégia da média) e os descritores de marcha obtidos de poses 2D estimadas por três métodos de estimativa de pose: PifPaf, OpenPose e TF-Pose-Estimation, no subconjunto de dados CASIA Gait Dataset-B com vista lateral - 90°, utilizando-se um classificador RNA para os descritores de marcha e o classificador 1-NN, com as distâncias euclidiana e *city-block*, para os descritores antropométricos.

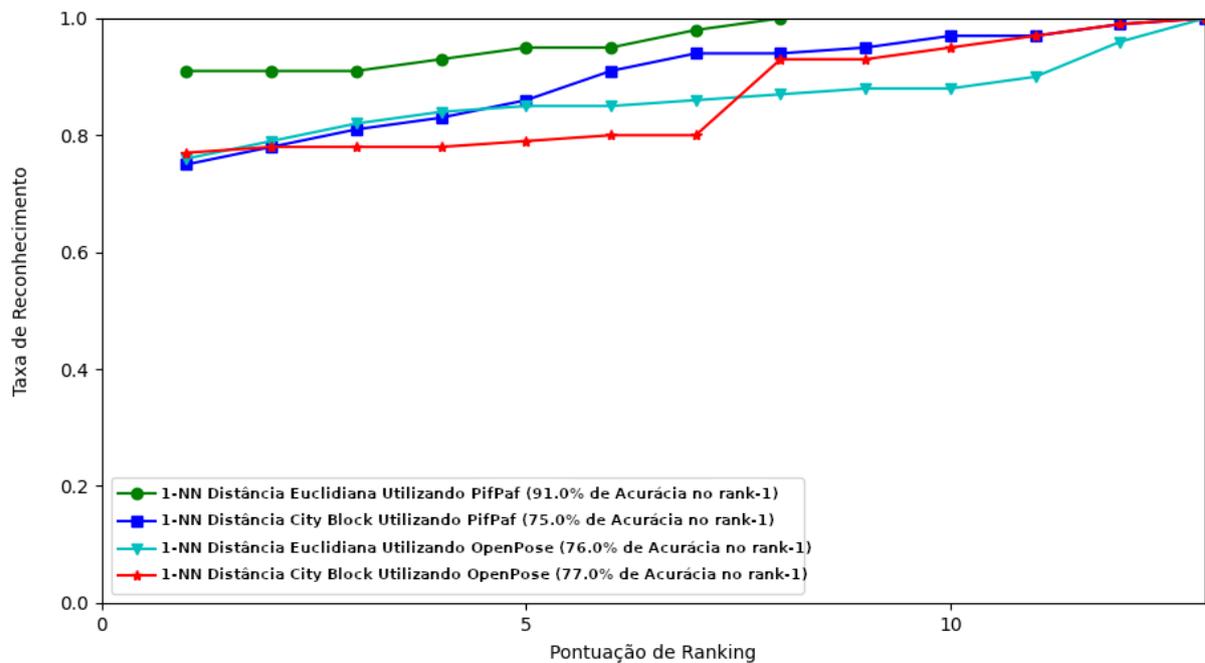


Fonte: Elaborada pelo autor.

Um terceiro experimento foi realizado com o objetivo de avaliar os desempenhos dos

métodos de estimação de poses 2D, aplicados na tarefa de reconhecimento de indivíduos. Neste experimento foram avaliados os métodos OpenPose e PifPaf, aplicados na base de dados CASIA Gait Dataset-A, utilizando-se os descritores antropométricos (estratégia de concatenação) obtidos de todos os vídeos (das três visões: lateral, frontal e diagonal), com o classificador 1-NN com as distâncias euclidiana e *city block*. A Figura 30 mostra as curvas CMC obtidas neste experimento. Observa-se nesta Figura que o método PifPaf obteve, novamente, resultados superiores aos obtidos pelo método OpenPose.

Figura 30 – Curvas CMC obtidas na base dados CASIA Gait Dataset-A, utilizando-se os métodos PifPaf e OpenPose, com os descritores antropométricos (estratégia de concatenação) e o classificador 1-NN, com as funções de distância euclidiana e *city block*.



Fonte: Elaborada pelo autor.

De modo geral, foi observado nestes três experimentos que o algoritmo PifPaf mostrou-se mais eficaz do que o OpenPose e o Tf-Pose-Estimation para estimar as poses nas imagens das bases de dados utilizadas neste trabalho. As Figuras 31(a) e 31(b) apresentam, respectivamente, as poses 2D estimadas pelos métodos PifPaf e OpenPose em um *frame* de um vídeo da base de dados privada descrita na Seção 5.3. Pode-se observar na Figura 31(a) que, dos cinco indivíduos presentes na cena, o PifPaf detectou quatro deles, enquanto que na Figura 31(b) pode-se observar que o método OpenPose detectou apenas dois indivíduos, sendo que os membros inferiores de um deles foram estimados incorretamente.

Figura 31 – Métodos de estimação de pose PifPaf e OpenPose aplicados em um mesmo *frame* de um vídeo da base de dados privada. Pode-se observar que dos cinco indivíduos na cena, o PifPaf (a) detectou quatro deles, enquanto que o OpenPose (b) detectou apenas dois, sendo que os membros inferiores de um dos indivíduos foram estimados erroneamente.



(a) PifPaf



(b) OpenPose

Fonte: Elaborada pelo autor.

Figura 32 – Exemplo de uma imagem em que o método PifPaf detectou e estimou corretamente a pose 2D de um trabalhador (a), enquanto que o método OpenPose falhou (b).



Fonte: Elaborada pelo autor.

A Figura 32 apresenta um caso aparentemente simples em que o PifPaf obteve êxito (Figura 32(a)), enquanto que o OpenPose não conseguiu detectar e estimar a pose do indivíduo (Figura 32(b)).

6.2 Avaliação dos Preprocessamentos

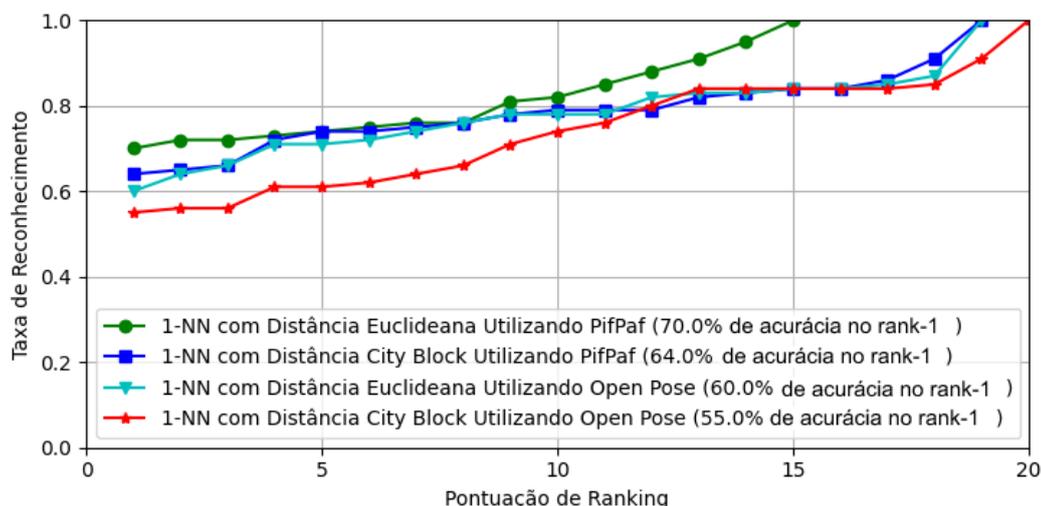
Foi realizado também um experimento com o objetivo de avaliar a influência dos preprocessamentos descritos na Seção 4.1 sobre os métodos de estimativa de poses 2D sendo utilizados neste trabalho.

A Figura 33 mostra as curvas CMC obtidas neste experimento sobre a base de dados CASIA Gait Dataset-B sem aplicar os preprocessamentos, utilizando-se o descritor baseado nas medidas antropométricas (estratégia da concatenação). Comparando-se estas curvas com aquelas mostradas na Figura 28, obtidas com a aplicação dos preprocessamentos, pode-se observar que os preprocessamentos propiciam uma melhora significativa nas taxas de reconhecimento dos indivíduos. Por exemplo, sem o preprocessamento, o método baseado nas medidas antropométricas extraídas das poses 2D estimadas pelo PifPaf, com o classificador 1-NN e a distância euclidiana, obtém uma acurácia *rank* 1 de 70,0%, enquanto que com o preprocessamento esta acurácia aumenta para 85,0%. Neste mesmo método, se for utilizado o OpenPose, a acurácia *rank* 1 aumenta de 60,0% para 80,0% quando se utiliza o preprocessamento.

6.3 Avaliação do Impacto do Número de Indivíduos

Com o objetivo de avaliar o desempenho do descritor de marcha baseado nos mapas de calor das articulações das poses estimadas pelos métodos OpenPose, PifPaf e Tf-Pose-Estimation em um conjunto reduzido de pessoas, foi realizado um experimento sobre um

Figura 33 – Curvas CMC obtidas utilizando-se os descritores antropométricos obtidos de poses 2D estimadas pelos métodos PifPaf e OpenPose, no subconjunto de dados CASIA Gait Dataset-B com vista lateral - 90°, utilizando-se o classificador 1-NN, com as distâncias euclidiana e city-block, sem aplicar os preprocessamentos.



Fonte: Elaborada pelo autor.

subconjunto da base de dados CASIA Gait Dataset-B composto pelos vídeos das visões laterais (90°) das 10 primeiras pessoas.

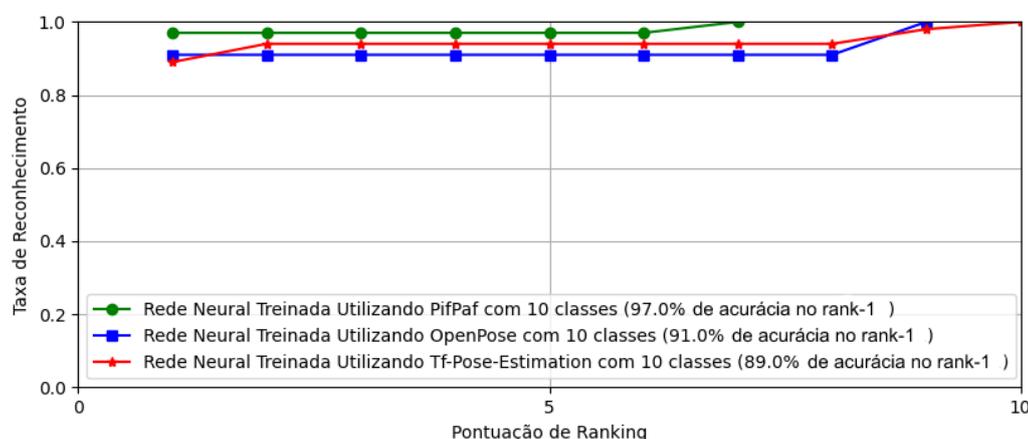
Este experimento foi desenhado para simular um cenário onde seria necessário identificar à distância um grupo pequeno de pessoas conhecidas (isto é, um problema de identificação de conjunto fechado) em um ambiente privado. Esse cenário poderia ser, por exemplo, o chão de fábrica de uma indústria.

A Figura 34 mostra as curvas CMC obtidas neste experimento. Pode-se observar que a melhor taxa de acurácia de rank-1 foi de 97% com o método PifPaf. Ou seja, para um grupo pequeno de pessoas, a identificação baseada no descritor de marcha proposto neste trabalho pode alcançar taxas próximas a 100%, utilizando-se a vista lateral da marcha das pessoas.

As medidas de precisão, revocação e F1-score obtidas para cada um dos dez indivíduos avaliados neste experimento utilizando-se o método PifPaf para estimação das poses 2D e o descritor de marcha com o classificador RNA são mostradas na Tabela 5. Como pode-se observar, as taxas médias de precisão e revocação atingiram 96%.

Para efeito de comparação, as medidas de precisão, revocação e F1-score também foram obtidas considerando-se todos os indivíduos da base de dados CASIA Gait Dataset B. A Tabela 6 mostra os resultados obtidos. Neste caso, a medida F1-Score foi de 89%. Como era de se esperar, quando se aumenta o número de indivíduos a serem identificados a tendência

Figura 34 – Curvas CMC obtidas usando o descritor de marcha baseado nos mapas de calor das articulações de poses estimadas pelos métodos PifPaf, OpenPose e TF-Pose-Estimation, para um conjunto de 10 indivíduos da base de dados CASIA Gait Dataset B.



Fonte: Elaborada pelo autor.

Tabela 5 – Medidas de precisão, revocação e F1-score utilizando-se os descritores de marcha extraídos dos mapas de calor das juntas das poses estimadas pelo método PifPaf em um conjunto de 10 indivíduos da base de dados CASIA Gait Dataset-B.

Indivíduo	Precisão	Revocação	F1-Score
0	0.97	0.98	0.98
1	0.98	0.98	0.98
2	0.97	0.96	0.96
3	0.95	0.95	0.95
4	0.94	0.97	0.95
5	0.96	0.93	0.95
6	0.97	0.97	0.97
7	0.99	0.92	0.95
8	0.95	0.95	0.95
9	0.90	0.96	0.93
accuracy			0.96
macro avg	0.96	0.96	0.96
weighted avg	0.96	0.96	0.96

Fonte: Elaborada pelo autor.

é diminuir a taxa de acurácia, e no caso dos descritores de marcha esta diminuição pode ser ainda mais acentuada, uma vez que este tipo de característica biométrica tem uma unicidade baixa, como mostra a Tabela 3, a partir dos estudos de Jain, Ross e Prabhakar (2004).

Tabela 6 – Medidas de Precisão, Revocação e F1-Score obtidas na base de dados CASIA Gait Dataset-B utilizando-se os descritores de marcha extraídos dos mapas de calor das juntas das poses estimadas pelo método PifPaf para todos os indivíduos da base.

	Precisão	Revocação	F1-Score
acurácia			0.89
macro avg	0.89	0.89	0.89
weighted avg	0.89	0.89	0.89

Fonte: Elaborada pelo autor.

6.4 Avaliação da Fusão no Nível de Pontuação

Conforme descrito na Seção 4.4, no método proposto neste trabalho para a identificação biométrica de pessoas por meio de poses 2D, objetivando melhorar as taxas de reconhecimento dos indivíduos os métodos baseados nas medidas antropométricas e nos padrões de marcha devem ser fundidos.

Para avaliar a eficácia da fusão dos métodos no nível de pontuação foi realizado um experimento, utilizando-se a base de dados CASIA Gait Dataset B e as poses obtidas pelo método PifPaf, no qual foi calculada a soma ponderada das pontuações, s_f , conforme definido na equação 10.

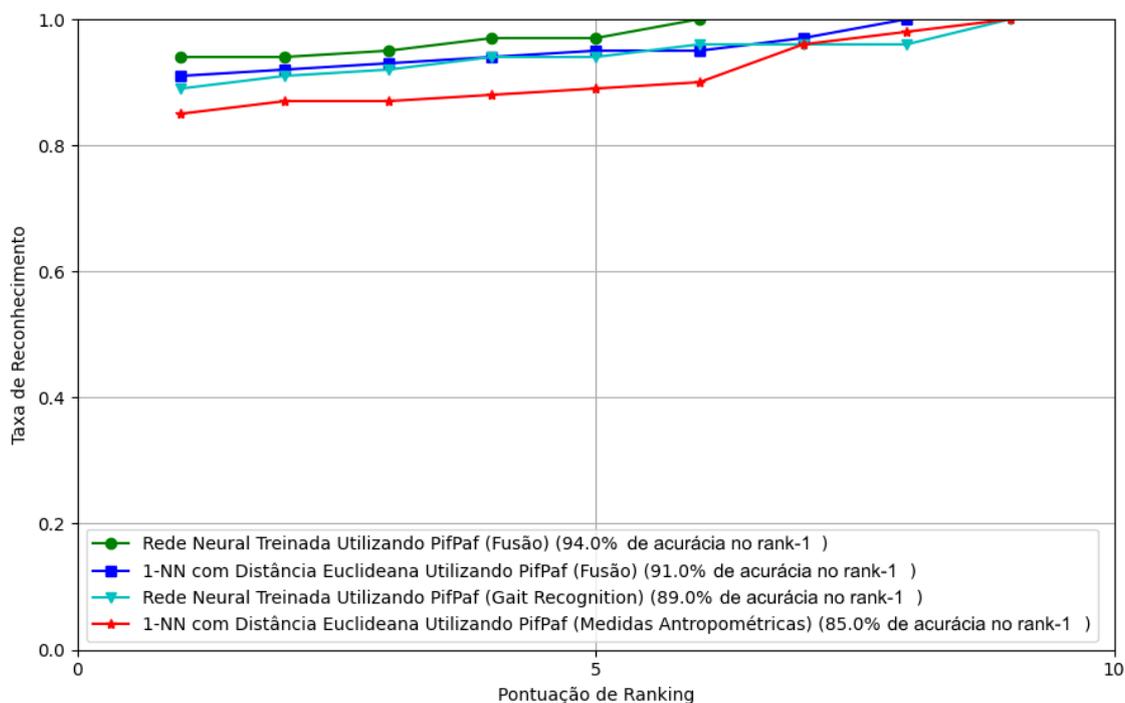
$$s_f = \omega_a \times s_a + \omega_g \times s_g \quad (10)$$

em que s_a é a pontuação obtida pelo método baseado nos descritores antropométricos (estratégia da concatenação), s_g é a pontuação obtida pelo método baseado nos descritores de marcha (média dos mapas de calor das articulações), ω_a é o peso da pontuação dos descritores antropométricos e ω_g é o peso da pontuação dos descritores de marcha, com $\omega_a + \omega_g = 1$. No experimento realizado, foram atribuídos valores de 0 a 1 para ω_a , com um incremento de 0,1 e $\omega_g = 1 - \omega_a$, sendo que o melhor resultado foi obtido com $\omega_a = 0,7$ e $\omega_g = 0,3$. A Figura 35 mostra as curvas CMC obtidas neste experimento utilizando-se os métodos baseados nos descritores antropométricos (estratégia da concatenação) e nos descritores de marcha individualmente e também aplicando-se a fusão de ambos, com pesos $\omega_a = 0,7$ e $\omega_g = 0,3$. Pode-se observar que a fusão propiciou um ganho nas taxas de reconhecimento para todos os valores de *rank*. Para *rank* igual a 1, por exemplo, a melhor taxa individual era de 89,0% e com a fusão esta taxa aumentou para 94,0%.

6.5 Avaliação da Fusão no Nível de Características

Como descrito na Seção 2.4, outra estratégia possível para a fusão biométrica é a fusão no nível de características, na qual os vetores de características obtidos das diferentes modalidades podem ser, por exemplo, concatenados. Este experimento visou avaliar esta

Figura 35 – Curvas CMC obtidas com os métodos baseados nos descritores antropométricos (estratégia da concatenação) e de marcha (média dos mapas de calor) individualmente e também aplicando-se a fusão de ambos, com pesos de 70% para os descritores antropométricos e 30% para os descritores de marcha.



Fonte: Elaborada pelo autor.

estratégia de fusão. Para tanto, foi utilizado como descritor antropométrico o histograma das medidas antropométricas e, como descritor de marcha, foi utilizado o método de extração de características proposto por Jangua e Marana (2020), pois este também gera histogramas como descritores, o que permitiu efetuar a fusão por meio da concatenação de vetores de características.

Um segundo objetivo deste experimento foi comparar os resultados obtidos pelo método proposto nesta dissertação com outros trabalhos encontrados na literatura, como o trabalho de Jangua e Marana (2020), que também é baseado em poses 2D. Para tanto, foi utilizado o método OpenPose para a estimação de poses, pois este foi o método utilizado por Jangua e Marana (2020), e utilizadas as bases de dados CASIA Gait Dataset A, com as visões lateral, frontal e oblíqua (diagonal), e a base de dados CASIA Gait Dataset B, com a visão lateral.

Neste experimento, o vetor de características das medidas antropométricas foi composto pela concatenação de oito histogramas (um para cada parte do esqueleto) contendo 16 *bins*

cada⁴, totalizando 128 características. Para o cálculo das distâncias entre os vetores de características após a fusão (concatenação) com os descritores de marcha do método de Jangua e Marana (2020), foram avaliadas as funções de distância euclidiana e qui-quadrado, e utilizados os seguintes pesos obtidos empiricamente: 20% para os histogramas dos ângulos (descritor de marcha), 10% para os histogramas das distâncias (descritor de marcha) e 70% para os histogramas das medidas antropométricas.

As Tabelas 7 e 8 apresentam os resultados das acurácias para *rank* 1 nas bases de dados CASIA Gait Dataset A e CASIA Gait Dataset B, respectivamente. Na base de dados CASIA Gait Dataset-A, o método proposto nesta dissertação, porém utilizando o método de reconhecimento de marcha proposto por Jangua e Marana (2020), alcançou 100% de acurácia no *rank* 1 na visão frontal e obteve também resultados superiores aos reportados por trabalhos correlatos nas visões lateral (95% de acurácia em *rank* 1) e oblíqua (97,5% de acurácia em *rank* 1). Na base de dados CASIA Gait Dataset-B, o método proposto nesta dissertação, porém utilizando o método de reconhecimento de marcha proposto por Jangua e Marana (2020), alcançou 97,85% de acurácia no *rank* 1 na visão lateral, muito próximo ao melhor resultado dos trabalhos correlatos, obtido por Lima e Schwartz (2019), que foi de 98%.

Tabela 7 – Acurácia Rank-1 - CASIA Gait Dataset-A.

Método	Lateral	Oblíqua	Frontal
(WANG, L. et al., 2003)	88,75%	87,50%	90,00%
(LIU D. et al., 2016)	85,00%	87,50%	95,00%
(LIMA; SCHWARTZ, 2019)	92,50%	96,25%	97,50%
(JANGUA; MARANA, 2020) (Euclidiana) - Marcha	80,00%	87,50%	96,25%
(JANGUA; MARANA, 2020) (Qui-quadrado) - Marcha	87,50%	92,50%	95,00%
Histograma das Medidas Antropométricas (Euclid.)	78,75%	85,00%	90,00%
Histograma das Medidas Antropométricas (Qui-quad)	85,00%	91,25%	88,75%
(JANGUA; MARANA, 2020) e Histogramas (Euclid.)	95,00%	96,25%	100,00%
(JANGUA; MARANA, 2020) e Histogramas (Qui-quad.)	95,00%	97,50%	100,00%

6.6 Avaliação em uma Aplicação Real

Finalmente, foi realizado um experimento utilizando-se a base de dados privada, apresentada da Seção 5.3, composta por vídeos capturados nas instalações de uma empresa durante o expediente rotineiro. Para garantir a segurança dos funcionários no local de trabalho, a empresa tem interesse em saber: 1) quem são os funcionários; 2) se eles estão trabalhando em áreas seguras; 3) quais ações eles estão executando; 4) se eles estão utilizando equipamentos de proteção individual (EPI); e 5) quando eles estão executando as ações. Nesta dissertação, focou-se na primeira tarefa. As Figuras 26 e 31 apresentam exemplos de *frames* de vídeos desta

⁴ Mesmo número de *bins* dos histogramas do método de Jangua e Marana (2020).

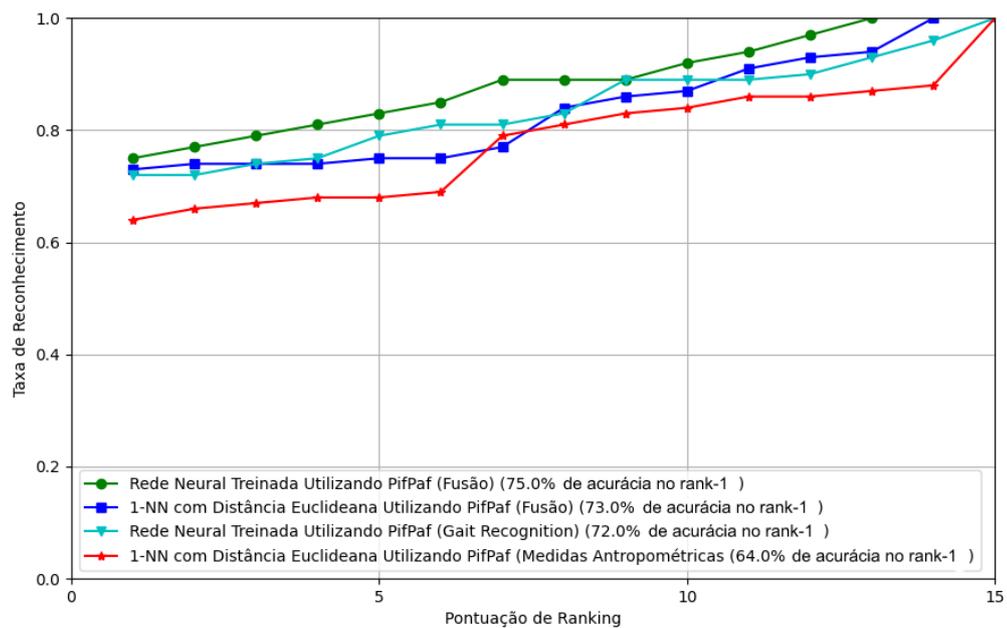
Tabela 8 – Acurácia Rank-1 - CASIA Gait Dataset-B.

Método	Lateral
(YU, S. et al., 2007)	83,50%
(CHEN C. et al., 2009)	91,10%
(LIMA; SCHWARTZ, 2019)	98,00%
(JANGUA; MARANA, 2020) (Euclidiana) - Marcha	91,26%
(JANGUA; MARANA, 2020) (Qui-quadrado) - Marcha	94,22%
Histograma das Medidas Antropométricas (Euclidiana)	85,89%
Histograma das Medidas Antropométricas (Qui-quadrado)	86,96%
(JANGUA; MARANA, 2020) e Histog. Medidas Antrop. (Euclidiana)	97,85%
(JANGUA; MARANA, 2020) e Histog. Medidas Antrop. (Qui-quadrado)	97,18%

base de dados. Observa-se que esta tarefa é bastante desafiadora, uma vez que os indivíduos estão distantes da câmera, movem-se em direções não controladas e podem ter partes do corpo obstruídas. Para identificar os funcionários neste cenário de forma contínua e natural durante a execução das suas tarefas, à distância, sem que fossem necessárias mudanças na rotina e nas instalações, optou-se por utilizar características biométricas de marcha e *soft*. Como os funcionários usam uniformes e EPIs muito similares, e que cobrem grande parte do corpo, inclusive a cabeça, as medidas antropométricas, mais especificamente os comprimentos de partes dos membros superiores (braços) e inferiores (pernas), foram as características *soft* escolhidas.

A Figura 36 apresenta as curvas CMC (Cumulative Match Characteristics) obtidas neste experimento. Foram utilizados os descritores antropométricos e os descritores de marcha obtidos de poses 2D estimadas pelo método PifPaf. Para os descritores de marcha utilizou-se um classificador RNA e para os descritores antropométricos utilizou-se um classificador 1-NN, com a distância euclidiana. Observa-se que os resultados foram inferiores aos resultados apresentados na Seção 6.3, Figura 34, o que era de se esperar, tendo em vista os desafios já mencionados da base de dados obtida no cenário real e não controlado. A Figura 36 mostra que no cenário real, a fusão das características antropométricas e de marcha também tendem a melhorar as taxas de reconhecimento obtidas individualmente com estas características. O método proposto neste trabalho atingiu uma acurácia de 75% em *rank* 1, o que pode ser considerado um bom resultado para esta base de dados.

Figura 36 – Curvas CMC obtidas utilizando-se os descritores antropométricos e os descritores de marcha obtidos de poses 2D estimadas pelo método PifPaf, na base de dados privada, utilizando-se um classificador RNA para os descritores de marcha e o classificador 1-NN, com a distância euclidiana para os descritores antropométricos.



Fonte: Elaborada pelo autor.

7 Conclusão e Trabalhos Futuros

Neste trabalho, foi proposto um método para a identificação de pessoas baseado na fusão de características antropométricas e de marcha, classificadas como características biométricas *soft* e comportamental, respectivamente, obtidas a partir de poses 2D estimadas em *frames* de vídeos.

Para as características antropométricas foram propostas três estratégias: média das medidas obtidas dos esqueletos (poses 2D) estimados em cada *frame*, concatenação das medidas obtidas dos esqueletos (poses 2D) estimados em cada *frame* e histograma das medidas obtidas dos esqueletos (poses 2D) estimados em cada *frame*. A média das medidas, apesar de ser mais simples, intuitiva e rápida para ser calculada, apresentou resultados inferiores às outras duas estratégias, as quais se justificam pelo fato das medidas antropométricas de um mesmo membro do esqueleto terem variações uma vez que são obtidas a partir dos comprimentos das projeções dos membros na imagem e também por causa de obstruções que podem ocorrer.

Para as características de marcha foi proposto um método baseado na média dos mapas de calor das articulações dos esqueletos (poses 2D) estimados em cada *frame*. Além deste método de extração de características de marcha, também foi avaliado neste trabalho os resultados da fusão utilizando-se um outro método baseado em poses 2D, proposto por Jangua e Marana (2020), para reconhecimento de marcha humana.

Os métodos propostos foram avaliados nas bases de dados CASIA Gait Dataset A e CASIA Gait Dataset B (vista lateral), que são amplamente utilizadas em trabalhos relacionados ao reconhecimento de marcha humana, e também em uma base de dados privada bastante desafiadora, descritas no Capítulo 5. Como os métodos propostos dependem crucialmente da qualidade das poses 2D, foram avaliados neste trabalho os desempenhos dos métodos propostos utilizando-se os seguintes métodos de estimação de poses 2D: OpenPose, PifPaf e Tf-Pose-Estimation, descritos na Seção 2.5.

Os resultados experimentais obtidos neste trabalho permitem concluir que:

1. O reconhecimento biométrico de indivíduos baseado em características *soft* e de marcha, embora tenham baixa unicidade, pode ser bastante efetivo quando se deseja realizar a identificação à distância, sem a colaboração das pessoas sendo identificadas e em cenários pouco controlados do cotidiano, corroborando, assim, outros trabalhos correlatos encontrados na literatura;
2. O reconhecimento biométrico baseado em marcha tende a apresentar resultados melhores do que os obtidos por meio das medidas antropométricas;

3. A fusão das características antropométricas e de marcha proposta neste trabalho tendem a apresentar resultados melhores do que os obtidos individualmente, tanto para a fusão no nível de pontuação quanto no nível de características;
4. O método PifPaf apresentou melhores resultados do que os métodos OpenPose e Tf-Pose-Estimation no contexto deste trabalho;
5. A aplicação de métodos de pré-processamento podem influenciar positivamente na qualidade das poses 2D estimadas no contexto deste trabalho;
6. A identificação de pessoas por meio das características antropométricas e de marcha é mais efetiva para pequenos grupos de indivíduos, devido à baixa unicidade destas características;
7. O uso de métodos estado da arte para a extração de características de marcha, como o de Lima e Schwartz (2019), associado ao método de extração de medidas antropométricas proposto neste trabalho pode gerar resultados ainda melhores.

7.1 Trabalhos Futuros

Dentre os vários trabalhos futuros que este trabalho enseja, destacam-se:

1. A investigação sobre métodos de estimação de poses 3D para possibilitar a extração das características antropométricas e de marcha de forma mais robusta;
2. A investigação mais aprofundada sobre métodos de rastreamento, para permitir a extração totalmente automática de descritores biométricos de vários indivíduos presentes simultaneamente nos vídeos.

7.2 Publicações Realizadas

Este trabalho ensejou, até o momento, a publicação dos seguintes artigos científicos:

- Tavares, H. L., Neto, J. B. C., Papa, J. P., Colombo, D., & Marana, A. N. (2019, September). Tracking and re-identification of people using soft-biometrics. In 2019 XV Workshop de Visao Computacional (WVC) (pp. 78-83). IEEE.
- Tavares, H. L., Neto, J. B. C., Papa, J. P., Colombo, D., & Marana, A. N. (2020, October). People Identification Based on Soft Biometrics Features Obtained from 2D Poses. In Brazilian Conference on Intelligent Systems (pp. 318-332). Springer, Cham.

Referências

- ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANÉ, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCHE, V.; VASUDEVAN, V.; VIÉGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <<http://tensorflow.org/>>.
- ANDERSSON, V.; DUTRA, R.; ARAÚJO, R. Anthropometric and human gait identification using skeleton data from kinect sensor. In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. [S.l.: s.n.], 2014. p. 60–61.
- ARAI, K.; ANDRIE, R. Gait recognition method based on wavelet transformation and its evaluation with chinese academy of sciences (casia) gait database as a human gait recognition dataset. In: *2001 Ninth International Conference on Information Technology - New Generations*. [S.l.: s.n.], 2001. p. 656–661.
- ARANTES, M. Método de reconhecimento da marcha humana por meio da fusão das características do movimento global. Escola de Engenharia de São Carlos, 2010.
- ARANTES, M. Método de reconhecimento da marcha humana por meio da fusão das características do movimento global. In: . [S.l.: s.n.], 2010.
- BEN, X.; GONG, C.; ZHANG, P.; JIA, X.; WU, Q.; MENG, W. Coupled patch alignment for matching cross-view gaits. *IEEE Transactions on Image Processing*, v. 28, n. 6, p. 3142–3157, June 2019b. ISSN 1057-7149.
- BEN, X.; ZHANG, P.; LAI, Z.; YAN, R.; ZHAI, X.; MENG, W. A general tensor representation framework for cross-view gait recognition. *Pattern Recognition*, v. 90, p. 87 – 98, 2019a. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320319300251>>.
- BENABDELKADER, C.; DAVIS, L. S.; CUTLER, R. Motion-based recognition of people in eigengait space. *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, p. 267–272, 2002.
- BEWLEY, A.; GE, Z.; OTT, L.; RAMOS, F.; UPCROFT, B. Simple online and realtime tracking. In: *2016 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2016. p. 3464–3468.
- BOLLE, R. M.; CONNELL, J. H.; PANKANTI, S.; RATHA, N. K.; SENIOR, A. W. The relation between the roc curve and the cmc. In: IEEE. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*. [S.l.], 2005. p. 15–20.
- BOULGOURIS, N. V.; CHI, Z. X. Gait recognition using radon transform and linear discriminant analysis. *IEEE Transactions on Image Processing*, v. 16, n. 3, p. 731–740, March 2007.

CAO, Z.; HIDALGO, G.; SIMON, T.; WEI, S.-E.; SHEIKH, Y. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In: *arXiv preprint arXiv:1812.08008*. [S.l.: s.n.], 2018.

CAO, Z.; HIDALGO, G.; SIMON, T.; WEI, S.; SHEIKH, Y. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008, 2018. Disponível em: <<http://arxiv.org/abs/1812.08008>>.

CHAO, H.; HE, Y.; ZHANG, J.; FENG, J. Gaitset: Regarding gait as a set for cross-view gait recognition. *CoRR*, abs/1811.06186, 2018. Disponível em: <<http://arxiv.org/abs/1811.06186>>.

CHEN C.; LIANG J.; ZHAO H.; HU H.; TIAN J. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 2009.

DANTCHEVA, A.; VELARDO, C.; D'ANGELO, A.; DUGELAY, J.-L. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, Springer, v. 51, n. 2, p. 739–777, 2011.

DECANN, B.; ROSS, A. Relating roc and cmc curves via the biometric menagerie. In: IEEE. *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. [S.l.], 2013. p. 1–8.

DITTRICH, W. H. Action categories and the perception of biological motion. *Perception*, v. 22, n. 1, p. 15–22, 1993. PMID: 8474831. Disponível em: <<https://doi.org/10.1068/p220015>>.

FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006.

FENG, Y.; LI, Y.; LUO, J. Learning effective gait features using lstm. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2016. p. 325–330.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN 9780131687288 013168728X 9780135052679 013505267X. Disponível em: <<http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>>.

GOUTTE, C.; GAUSSIÉ, E. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In: SPRINGER. *European conference on information retrieval*. [S.l.], 2005. p. 345–359.

GUO, B. H.; NIXON, M. S.; CARTER, J. N. Fusion analysis of soft biometrics for recognition at a distance. In: *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*. [S.l.: s.n.], 2018. p. 1–8.

HASAN, M.; BABAGUCHI, N. Long-term people reidentification using anthropometric signature. In: IEEE. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. [S.l.], 2016. p. 1–6.

HAVASI, L.; SZLAVIK, Z.; SZIRANYI, T. Detection of gait characteristics for scene registration in video surveillance system. *IEEE Transactions on Image Processing*, v. 16, n. 2, p. 503–510, Feb 2007.

- JAIN, A. K. Biometric recognition: Overview and recent advances. In: RUEDA, L.; MERY, D.; KITTLER, J. (Ed.). *Progress in Pattern Recognition, Image Analysis and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 13–19. ISBN 978-3-540-76725-1.
- JAIN, A. K.; DASS, S. C.; NANDAKUMAR, K. Can soft biometric traits assist user recognition? *age*, v. 20, p. 39, 2004.
- JAIN, A. K.; KUMAR, A. Biometric recognition: an overview. In: *Second generation biometrics: The ethical, legal and social context*. [S.l.]: Springer, 2012. p. 49–79.
- JAIN, A. K.; NANDAKUMAR, K.; ROSS, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, Elsevier, v. 79, p. 80–105, 2016.
- JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, IEEE, v. 14, n. 1, p. 4–20, 2004.
- JAIN, A. K.; ROSS, A. A.; NANDAKUMAR, K. *Introduction to biometrics*. [S.l.]: Springer Science & Business Media, 2011.
- JANGUA, D.; MARANA, A. A new method for gait recognition using 2d poses. In: SBC. *Anais do XVI Workshop de Visão Computacional*. [S.l.], 2020. p. 69–74.
- JOHANSSON, G. Visual motion perception. *Scientific American*, v. 232, n. 6, p. 76–88, 1975.
- KREISS, S.; BERTONI, L.; ALAHI, A. Pifpaf: Composite fields for human pose estimation. *CoRR*, abs/1903.06593, 2019. Disponível em: <<http://arxiv.org/abs/1903.06593>>.
- KREISS, S.; BERTONI, L.; ALAHI, A. Pifpaf: Composite fields for human pose estimation. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2019.
- LAM, T. H.; CHEUNG, K. H.; LIU, J. N. Gait flow image: A silhouette-based gait representation for human identification. *Pattern recognition*, Elsevier, v. 44, n. 4, p. 973–987, 2011.
- LIAO, R.; YU, S.; AN, W.; HUANG, Y. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, Elsevier, v. 98, p. 107069, 2020.
- LIMA, V. C. de; SCHWARTZ, W. R. Gait recognition using pose estimation and signal processing. In: SPRINGER. *Iberoamerican congress on pattern recognition*. [S.l.], 2019. p. 719–728.
- LIU D.; YE M.; LI X.; ZHANG F.; LIN L. Memory-based gait recognition. *BMVC*, p. 1–12, 2016.
- MAGUIRE, M. The birth of biometric security. *Anthropology Today*, Wiley Online Library, v. 25, n. 2, p. 9–14, 2009.
- MARCONDES, J. S. Sistema biométrico: O que é, como funciona? *Gestão de Segurança Privada*, 08 2020.
- MARTIN, A.; DODDINGTON, G.; KAMM, T.; ORDOWSKI, M.; PRZYBOCKI, M. *The DET curve in assessment of detection task performance*. [S.l.], 1997.
- MOESLUND, T. B.; GRANUM, E. A survey of computer vision-based human motion capture. *Comput. Vis. Image Underst.*, Elsevier Science Inc., New York, NY, USA, v. 81, n. 3, p. 231–268, mar. 2001. ISSN 1077-3142. Disponível em: <<http://dx.doi.org/10.1006/cviu.2000.0897>>.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. USA: Omnipress, 2010. (ICML'10), p. 807–814. ISBN 978-1-60558-907-7. Disponível em: <<http://dl.acm.org/citation.cfm?id=3104322.3104425>>.

NIXON, M. et al. Model-based gait recognition. Springer, 2009.

ORRÙ, P. F.; ZOCCHEDDU, A.; SASSU, L.; MATTIA, C.; COZZA, R.; ARENA, S. Machine learning approach using mlp and svm algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. *Sustainability*, v. 12, n. 11, 2020. ISSN 2071-1050. Disponível em: <<https://www.mdpi.com/2071-1050/12/11/4776>>.

PANCHAL, T.; SINGH, A. Multimodal biometric system. *International Journal of Advanced Research in Computer Science and Software Engineering*, v. 3, n. 5, 2013.

PERRY, J. Análise da marcha. *Manoele*, n. 7, p. 125, 2005.

PIETIKÄINEN, M. Image analysis with local binary patterns. *Image Analysis*, Springer, p. 6–9, 2005.

PRABHAKAR, S.; PANKANTI, S.; JAIN, A. K. Biometric recognition: Security and privacy concerns. *IEEE security & privacy*, IEEE, v. 1, n. 2, p. 33–42, 2003.

ROSS, A.; JAIN, A. Information fusion in biometrics. *Pattern recognition letters*, Elsevier, v. 24, n. 13, p. 2115–2125, 2003.

RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. [S.l.]: Pearson, 2016.

SARKAR, S.; PHILLIPS, P. J.; LIU, Z.; VEGA, I. R.; GROTH, P.; BOWYER, K. W. The humanoid gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, p. 162–177, 2005.

SILVA, M. V. da; MARANA, A. N. Human action recognition in videos based on spatiotemporal features and bag-of-poses. *Applied Soft Computing*, Elsevier, v. 95, p. 106513, 2020.

SOBABE, A.; DJARA, T.; VIANOU, A. A framework for combination of sequential architecture and soft biometrics in multibiometric scores fusion. In: *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)*. [S.l.: s.n.], 2019. p. 1–4.

SU, C.; LI, J.; ZHANG, S.; XING, J.; GAO, W.; TIAN, Q. *Pose-driven Deep Convolutional Model for Person Re-identification*. 2017.

TAN, D.; HUANG, K.; YU, S.; TAN, T. Efficient night gait recognition based on template matching. In: IEEE. *18th International Conference on Pattern Recognition (ICPR'06)*. [S.l.], 2006. v. 3, p. 1000–1003.

TIAN, Y.; WEI, L.; LU, S.; HUANG, T. Free-view gait recognition. *PLOS ONE*, Public Library of Science, v. 14, n. 4, p. 1–24, 04 2019.

TOME, P.; FIERREZ, J.; VERA-RODRIGUEZ, R.; NIXON, M. S. Soft biometrics and their application in person recognition at a distance. *IEEE Transactions on information forensics and security*, IEEE, v. 9, n. 3, p. 464–475, 2014.

TONG, S. bing; FU, Y. zhuo; LING, H. fei. Cross-view gait recognition based on a restrictive triplet network. *Pattern Recognition Letters*, v. 125, p. 212 – 219, 2019. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865518307475>>.

TOSHEV, A.; SZEGEDY, C. Deeppose: Human pose estimation via deep neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun 2014. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2014.214>>.

VELARDO, C.; DUGELAY, J.-L. Improving identification by pruning: A case study on face recognition and body soft biometric. In: IEEE. *2012 13th International Workshop on Image Analysis for Multimedia Interactive Services*. [S.l.], 2012. p. 1–4.

WANG, L.; TAN, T.; HU, W.; NING, H. Automatic gait recognition based on statistical shape analysis. *IEEE transactions on pattern analysis and machine intelligence*, v. 12, n. 9, p. 1120–1131, 2003.

WANG, Y.; DU, B.; SHEN, Y.; WU, K.; ZHAO, G.; SUN, J.; WEN, H. Ev-gait: Event-based robust gait recognition using dynamic vision sensors. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2019.

WANG, Y.; YU, S.; WANG, Y.; TAN, T. Gait recognition based on fusion of multi-view gait sequences. In: ZHANG, D.; JAIN, A. K. (Ed.). *Advances in Biometrics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 605–611. ISBN 978-3-540-31621-3.

WETTSCHERECK, D.; AHA, D. W.; MOHRI, T. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, Springer, v. 11, n. 1, p. 273–314, 1997.

WINTER, D. Biomechanics and motor control of human gait: normal, elderly and pathological - 2nd edition. *Transport Research Laboratory*, Transport Research Laboratory, v. 2, 1991.

WOJKE, N.; BEWLEY, A.; PAULUS, D. Simple online and realtime tracking with a deep association metric. In: *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017. p. 3645–3649.

WU, X.; WARD, R.; BOTTOU, L. WNGrad: Learn the Learning Rate in Gradient Descent. *ArXiv e-prints*, mar. 2018.

YANG, W.; XU, L.; CHEN, X.; ZHENG, F.; LIU, Y. Chi-squared distance metric learning for histogram data. *Mathematical Problems in Engineering*, Hindawi, v. 2015, 2015.

YU, S.; TAN, D.; HUANG, K.; TAN, T. Reducing the effect of noise on human contour in gait recognition. *Internat. Conf. on Biometrics*, 2007.

YU, S.; TAN, D.; TAN, T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: *18th International Conference on Pattern Recognition (ICPR'06)*. [S.l.: s.n.], 2006. v. 4, p. 441–444. ISSN 1051-4651.

ZHENG, S.; HUANG, K.; TAN, T.; TAO, D. A cascade fusion scheme for gait and cumulative foot pressure image recognition. *Pattern Recognition*, Elsevier, v. 45, n. 10, p. 3603–3610, 2012.