



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"

Universidade Estadual Paulista (UNESP)

Faculdade de Ciências

Curso de Bacharelado em Ciência da Computação

Campus de Bauru – SP

---

# Reconhecimento de marcha humana utilizando poses 3D estimadas de múltiplas poses 2D

---

Pedro Luiz Cason Caldato

---

Disciplina: Projeto e Implementação de Sistemas 2023  
Professor Orientador: Aparecido Nilceu Marana

# Sumário

---

---

1. Introdução
  2. Objetivos
  3. Fundamentação teórica
  4. Método Proposto
  5. Resultados e discussão
  6. Conclusão
  7. Referências
- 
-



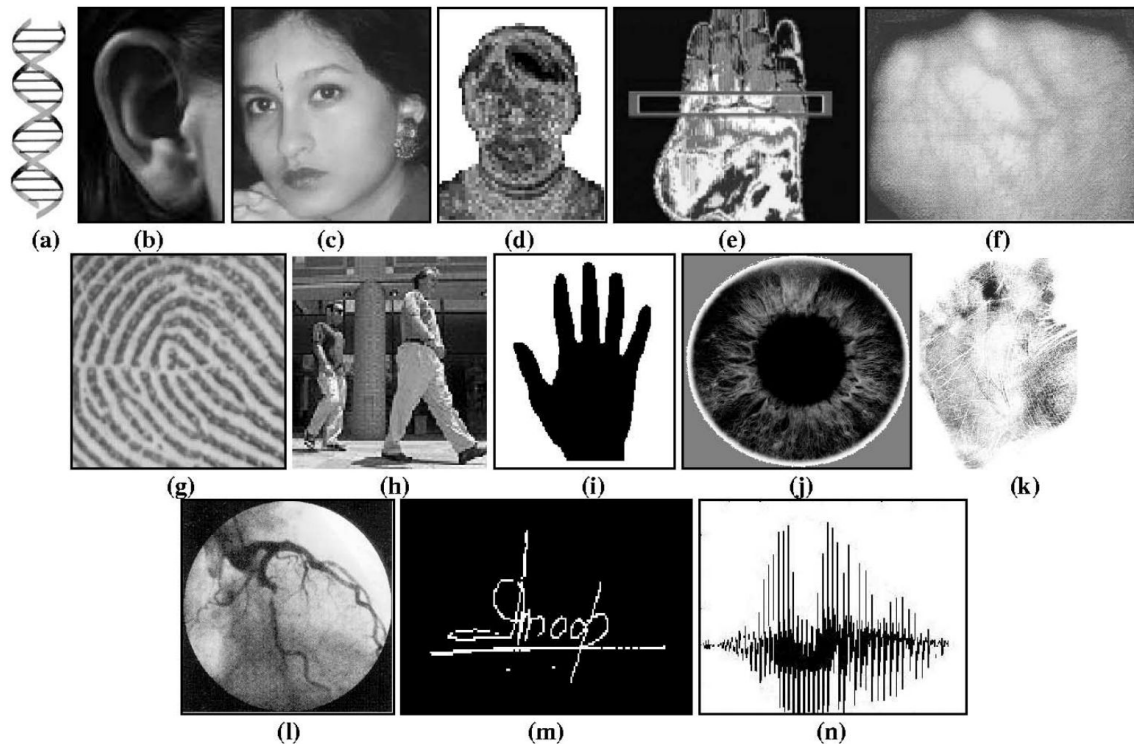
# Introdução



# Biométrie - O que é:

Análise de características físicas ou comportamentais para identificação do indivíduo, como:

- a) DNA
- b) Ouvido
- c) Face
- d) Termograma facial
- e) Termograma da mão
- f) Veias das mãos
- g) Impressão digital
- h) Caminhada**
  - i) Geometria da mão
  - j) Iris
  - k) Impressão da palma da mão
  - l) Retina
  - m) Assinatura
  - n) Voz

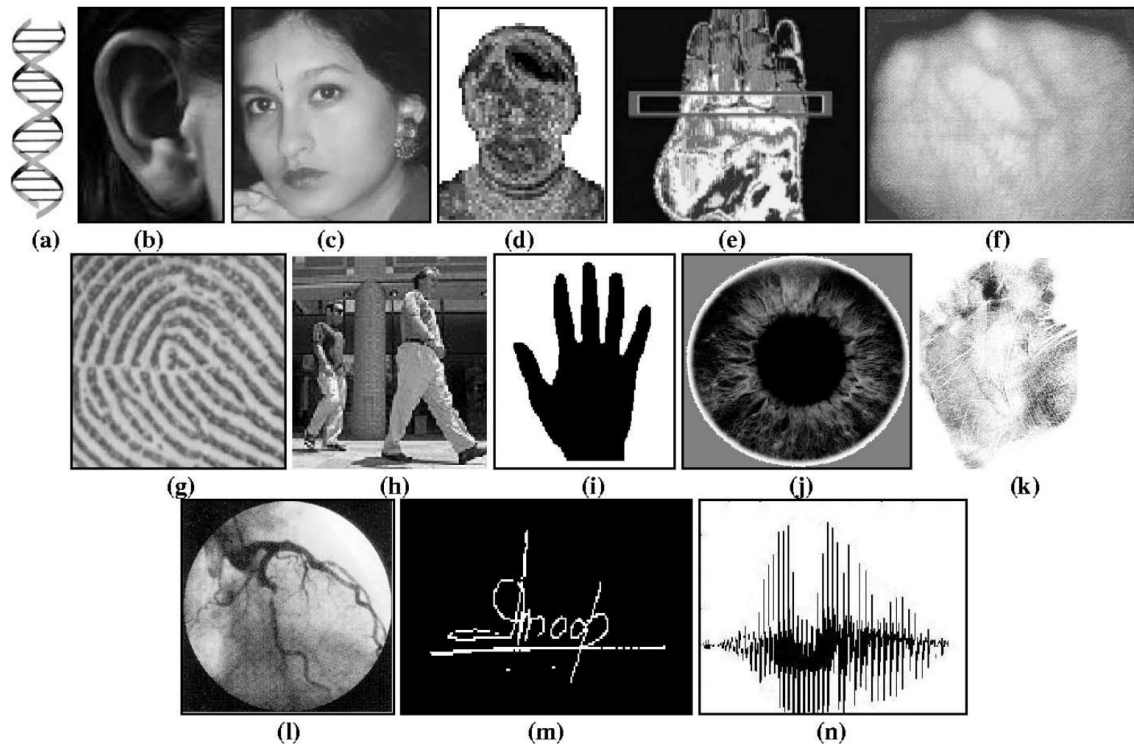


Fonte: Jain, Ross e Prabhakar (2004)

# Biométrie - O que é:

Utilizados diariamente para:

1. Autenticação de dispositivos móveis (celulares);
2. Confirmação de pagamentos bancários;
3. Liberação de acesso em ambientes;
4. Autenticação de documentos;
5. Controle de ambientes;



Fonte: Jain, Ross e Pabhakar (2004)

# Problemas

- Necessidade de **sensores específicos** para a aplicação, como no caso dos sistemas biométricos por digital ou íris;
- Requer, em sua grande maioria, o **contato direto e colaboração do indivíduo**;
- Os dados precisam ser de **boa qualidade** para fazer o reconhecimento biométrico, como para o reconhecimento facial, por íris, voz, ou digital;
- Algumas aplicações requerem a identificação da pessoa sem o contato direto e, às vezes, sem o consentimento:
  - Pode-se citar o monitoramento de ambientes como aeroportos e shoppings;

# Reconhecimento Biométrico por Marcha

- Identifica o indivíduo com base no padrão de caminhada do mesmo;
  - Utiliza características como ângulo máximo entre as pernas e braços ao caminhar -> movimento repetitivo;
  - Amplitude do movimento;
  - Angulação média e máxima entre membros (braços e antebraços, por exemplo);
- Possível de se utilizar em aplicações com vídeos de baixa resolução (JANGUA; MARANA, 2020);
- Pode ser feito o reconhecimento biométrico à distância e de forma anônima;
- Pode utilizar câmeras de vigilância já instalada nos locais;





# Objetivos





# Objetivo Geral

Desenvolver um método baseado em Rede Neural Artificial (RNA) para reconhecimento biométrico de pessoas através de reconhecimento de marcha humana utilizando poses 3D obtidas a partir de múltiplas poses 2D.

## Objetivos Específicos

- Pesquisar sobre reconhecimento biométrico de pessoas;
- Realizar levantamento bibliográfico de trabalho correlatos;
- Extração de poses 3D a partir de múltiplas poses 2D;
- Modelagem de RNA;
- Realização de testes com a RNA proposta;
- Analisar os resultados obtidos.

---

---

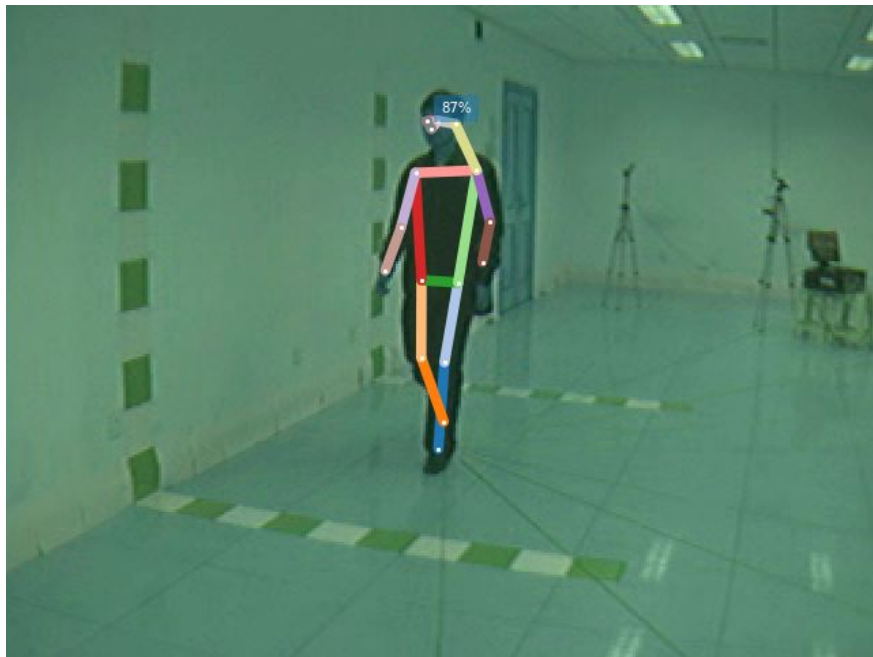
# Fundamentação Teórica

---

---

# Pose 2D/3D

Vértices no plano da imagem formado por pontos 2D (x, y);



Fonte: Elaborado pelo autor.

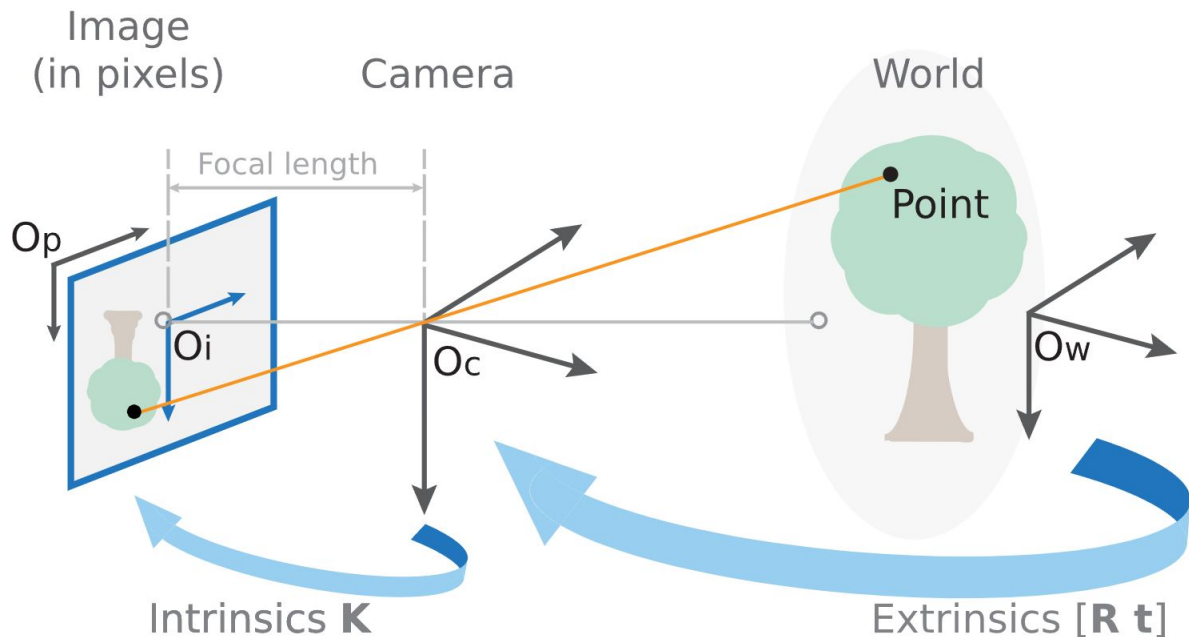
Vértices no frame do mundo formado por pontos 3D (x, y, z);



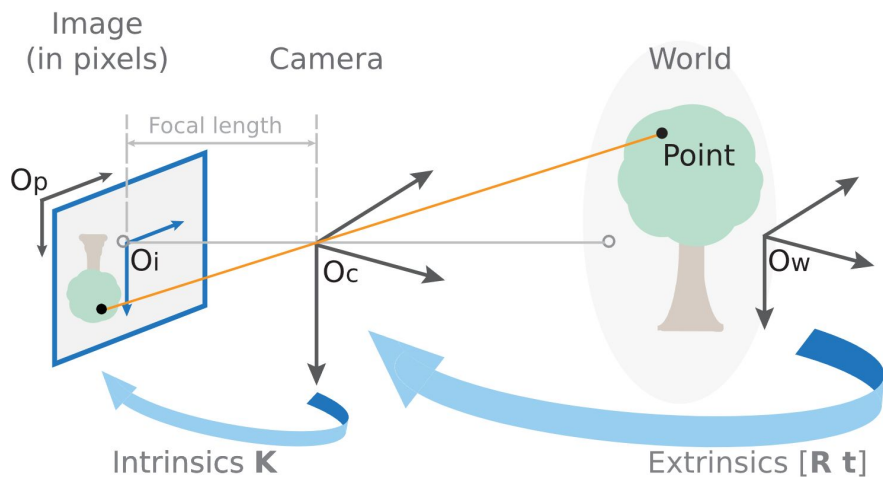
Fonte: Elaborado pelo autor.

# Parâmetros Extrínseco/Intrínseco de câmeras

- Descrevem a projeção de um ponto 3D do mundo no plano da imagem;
- Obtenção através do processo de calibração de câmeras;



# Parâmetros Extrínseco/Intrínseco de câmeras



$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

---

---

# Método Proposto

---

---

# Base de Dados

- CASIA Gait Dataset-A:
  - 20 indivíduos distintos;
  - 12 sequências de imagens por indivíduo;
  - 4 sequências para cada uma direção;



# Base de Dados

- CASIA Gait Dataset-B:
  - 124 indivíduos distintos;
  - 11 câmeras diferentes;
  - Câmeras com angulações de 18°;

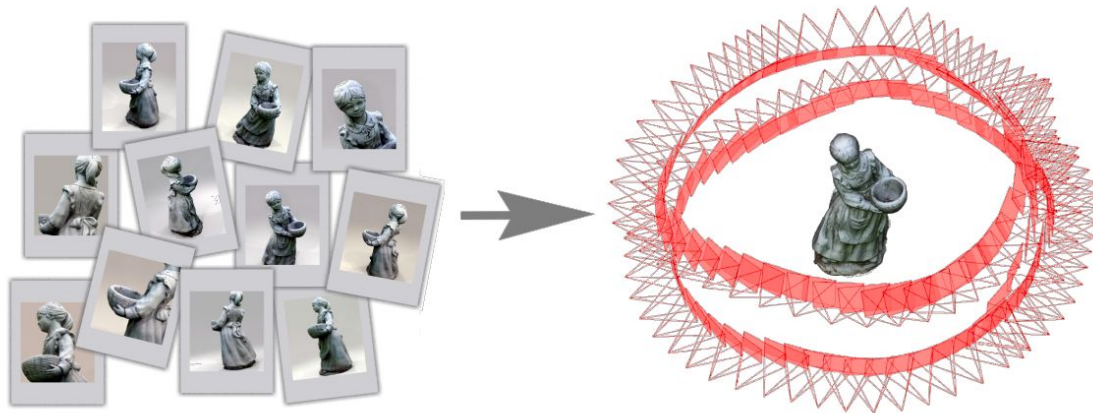


Fonte: Center for Biometrics and Security Research (2010).



# Calibração das Câmeras

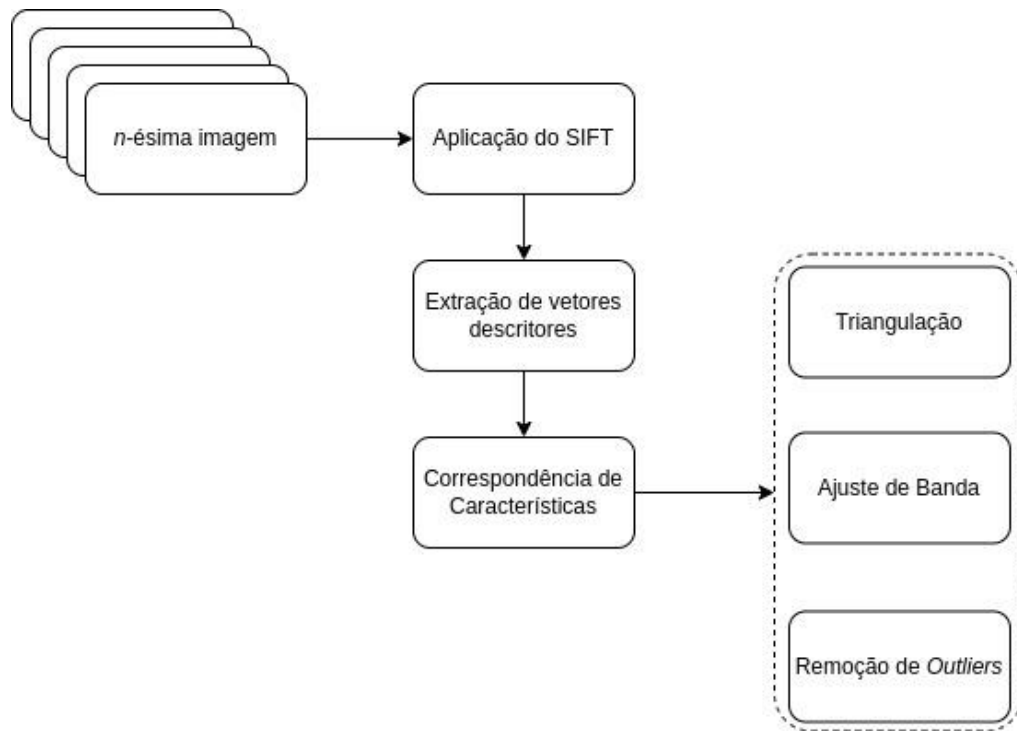
- Utilização do COLMAP -> Pipeline de SFM (Structure From Motion);
- Amplamente utilizado para reconstruções 3D (com textura) utilizando apenas imagens;
- Resolve sistemas de equações não-linear para reduzir a função de custo de reprojeção;



Fonte: Adaptado de Bianco, Ciocca e Marelli (2018)

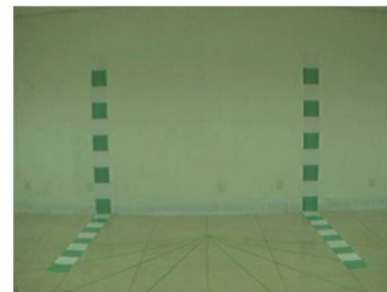
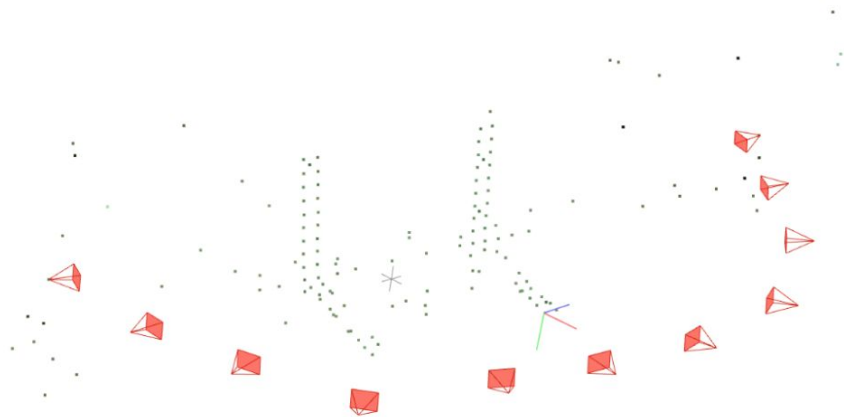
# Calibração das Câmeras

- Pipeline do COLMAP



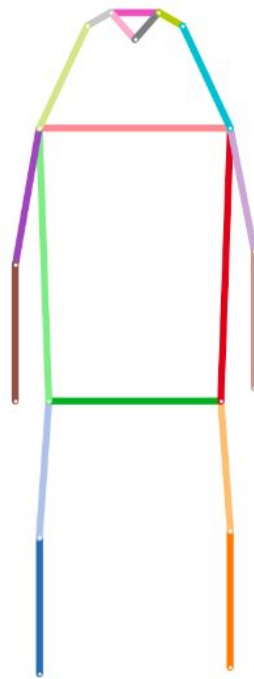
# Calibração das Câmeras

- Utilização do COLMAP para a estimação dos parâmetros das câmeras para o dataset CASIA-B;

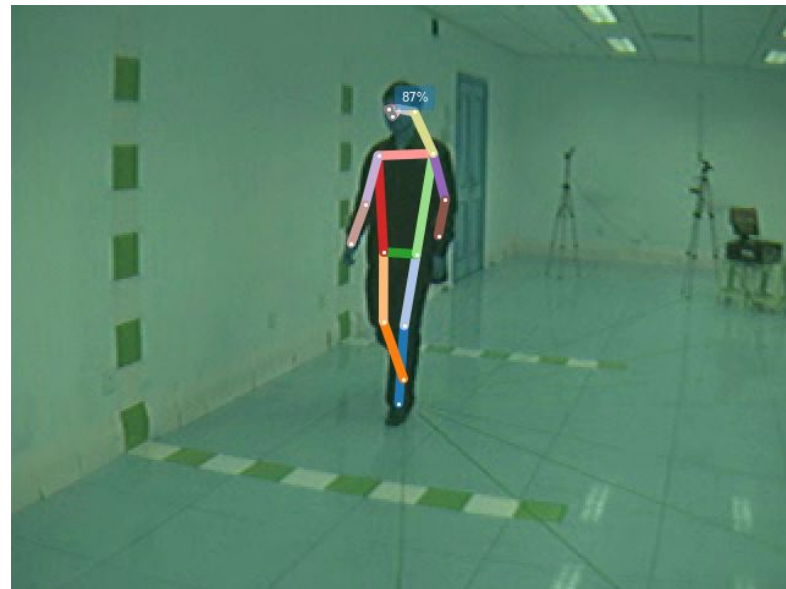


# Extração das poses 2D

- Utilização da Rede Neural Convolucional OpenPifPaf;
- Escolha pelo código aberto e facilidade do uso;
- Ao total, são 17 pontos-chave diferentes;



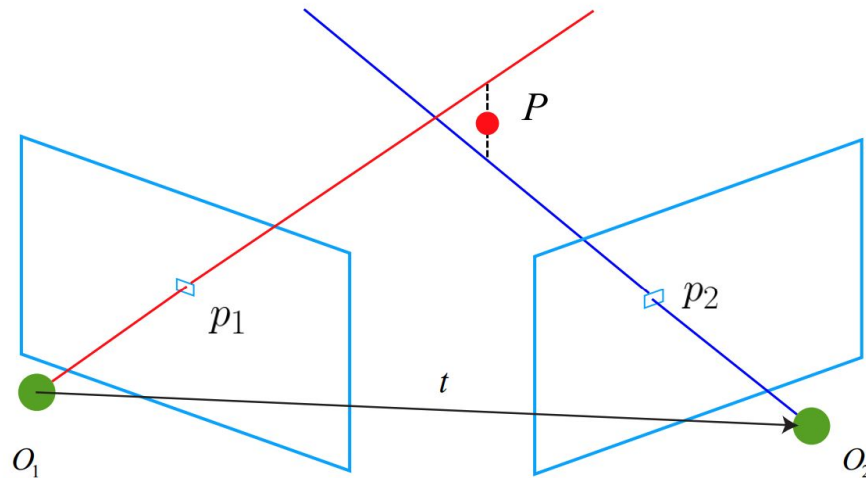
Fonte: Adaptado de Kreiss,  
Bertoni e Alahi (2021).



Fonte: Elaborado pelo autor.

# Estimação das poses 3D utilizando poses 2D

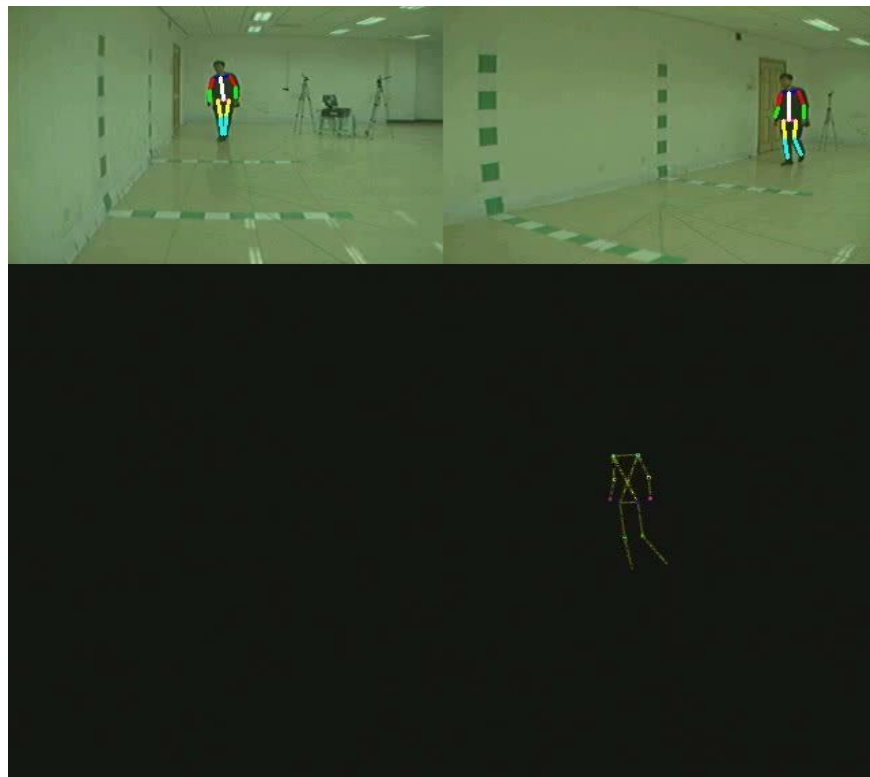
- Utilização da função `cv2.triangulatePoints` em Python:
  - Necessidade das coordenadas do mesmo ponto nas duas imagens diferentes;
  - Necessidade da matriz de projeção ( $K[R|t]$ );
  - Cálculo da *depth* através da disparidade  $\frac{z}{t} = \frac{f}{D}$
  - Estimação do ponto 3D;



Fonte: Adaptado de Kreiss, Bertoni e Alahi (2021).

# Estimação das poses 3D utilizando poses 2D

- Exemplo das poses 3D geradas no dataset CASIA-B;
- Foram utilizados apenas duas câmeras que proporciona ampla visão dos membros dos indivíduos;



Fonte: Elaborado pelo autor.

# RNA Proposta

- Objetivo de condensar a informação e representá-la em um espaço vetorial de menor dimensão:
  - Baseado no modelo NetVLAD, utilizado para abordagens de *Image Retrieval*;
  - Última camada é uma normalização, fazendo com que o vetor descritor tenha norma 1 e possibilitando a função de custo;
  - Função de custo utilizada é a *Triplet Loss*;

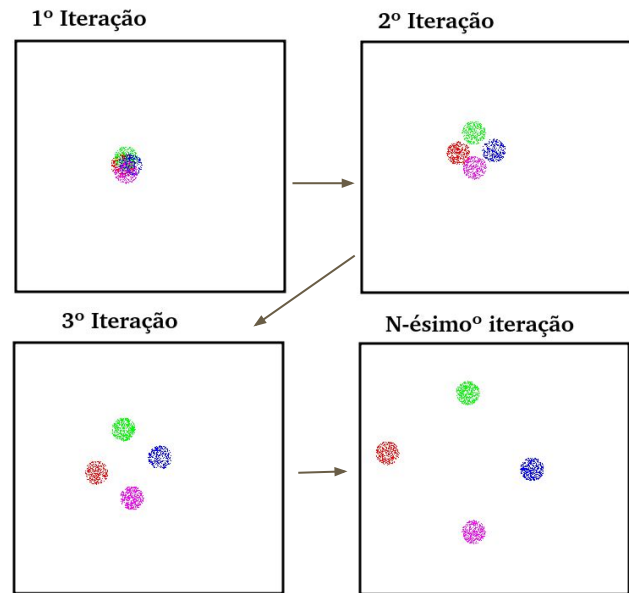
Camada	Canais de Entrada	Canais de Saída
<b>x</b>	$40 \times 17 \times 3$	-
CONV 2D	$40 \times 17 \times 3$	$64 \times 17 \times 3$
BATCH NORM	$64 \times 17 \times 3$	$64 \times 17 \times 3$
RELU	$64 \times 17 \times 3$	$64 \times 17 \times 3$
CONV 2D	$64 \times 17 \times 3$	$128 \times 17 \times 3$
BATCH NORM	$128 \times 17 \times 3$	$128 \times 17 \times 3$
RELU	$128 \times 17 \times 3$	$128 \times 17 \times 3$
CONV 2D	$128 \times 17 \times 3$	$256 \times 17 \times 3$
BATCH NORM	$256 \times 17 \times 3$	$256 \times 17 \times 3$
RELU	$256 \times 17 \times 3$	$256 \times 17 \times 3$
FLATTEN	$256 \times 17 \times 3$	$13056 \times 1$
LINEAR	$13056 \times 1$	$512 \times 1$
BATCH NORM	$512 \times 1$	$512 \times 1$
RELU	$512 \times 1$	$512 \times 1$
LINEAR	$512 \times 1$	$256 \times 1$
L2-NORM	$256 \times 1$	$256 \times 1$

Fonte: Elaborado pelo autor.

# RNC Proposta

- Para cada iteração, durante o treinamento da rede, a mesma “aprende” pesos que consigam gerar descritores diferentes para diferentes dados de entrada;
- Com o banco de dados criado, utiliza-se a distância cossenoidal entre o vetor descritor estimado com o banco para recuperar a pessoa que o dado de entrada/descritor pertence:

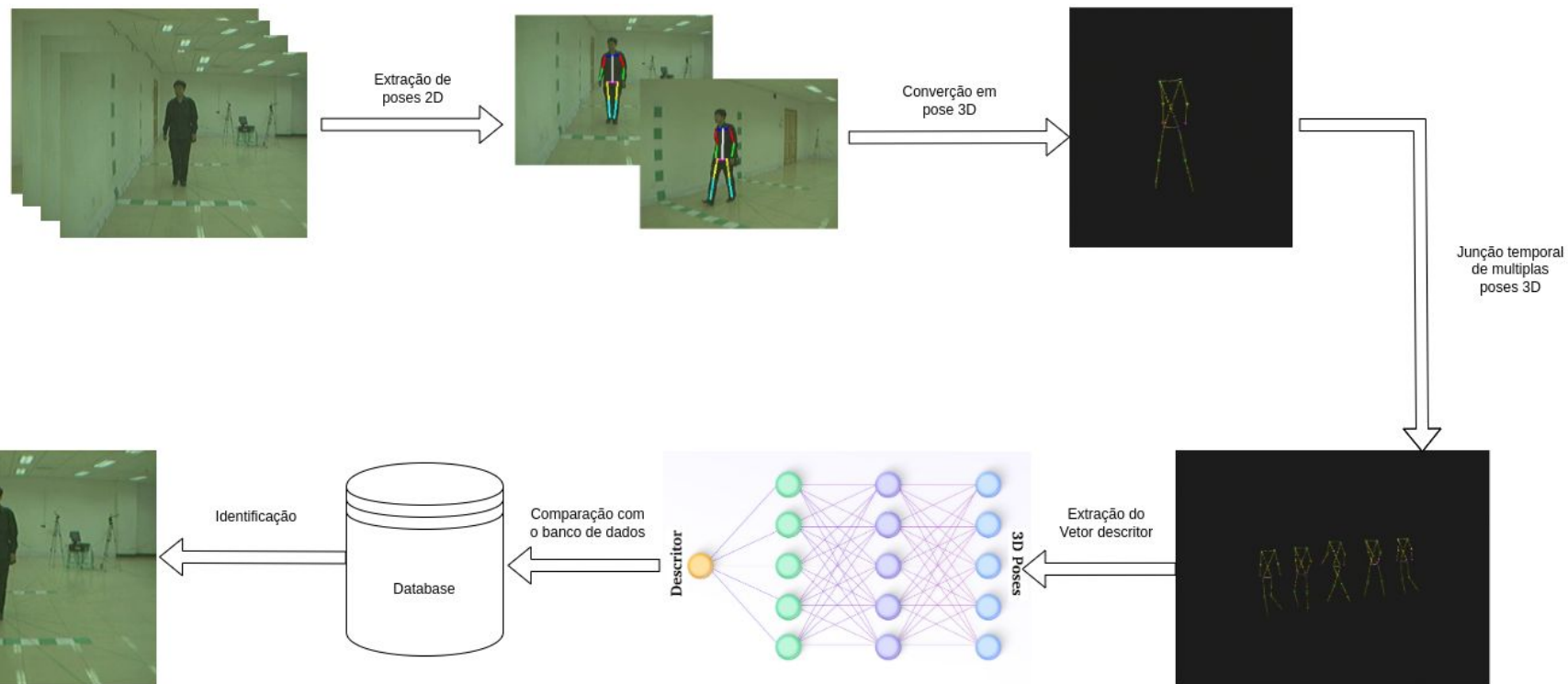
$$\cos \theta = \frac{x \odot y}{||x|| \cdot ||y||}.$$



Fonte: Elaborado pelo autor.



# Pipeline Proposta



Fonte: Elaborado pelo autor.

---

---

# Resultados e Discussão

---

---

# CASIA GAIT-A em Top-1

- Validação utilizando o dataset CASIA GAIT-A;
- Treinamento utilizando os vídeos sem acessórios;
- Utilização dos vídeos com acessórios para o cálculo da acurácia;
- Obtivemos uma boa acurácia quando treinado utilizando ambos datasets;

Método	Acurácia
<a href="#">Wang et al. (2003)</a>	88.75%
<a href="#">Liu et al. (2016)</a>	89.17%
<a href="#">Lima e Schwartz (2019)</a>	95.42%
<a href="#">Jangua e Marana (2020)</a> com distância Euclidiana	87.92%
<a href="#">Jangua e Marana (2020)</a> com distância Chi-quadrado	91.67%
Método Proposto Treinado no Conjunto GAIT-A	90.00%
Método Proposto Treinado no Conjunto GAIT-B	85.00%

Fonte: Elaborado pelo autor.

# CASIA GAIT-B em Top-1

- Validação utilizando o dataset CASIA GAIT-B;
- Treinamento utilizando os vídeos sem acessórios;
- Utilização dos vídeos com acessórios para o cálculo da acurácia;
- Obtivemos uma boa acurácia quando treinado utilizando o dataset CASIA GAIT-B;
- Baixa acurácia quando treinado usando o dataset CASIA GAIT-A:
  - Uma possível justificativa é a pequena quantidade de dados para treino;

Método	Acurácia
Yu et al. (2007)	83.50%
Chen et al. (2009)	91.10%
Lima e Schwartz (2019)	98.00%
Jangua e Marana (2020) (Distância Euclidiana)	91.26%
Jangua e Marana (2020) (Distância Chi-Quadrado)	94.22%
Método Proposto Treinado no Conjunto GAIT-A	56.45%
Método Proposto Treinado no Conjunto GAIT-B	86.10%

Fonte: Elaborado pelo autor.

# CASIA GAIT-B com casaco em Top-1

- Validação utilizando o dataset CASIA GAIT-B;
- Treinamento utilizando os vídeos com casaco;
- Utilização dos vídeos com/sem acessórios para o cálculo da acurácia;
- Obtivemos uma boa acurácia quando treinado utilizando o dataset CASIA GAIT-B:
  - Prova-se que nossa abordagem tem uma baixa dependência quanto aos acessórios;
- Baixa acurácia quando treinado usando o dataset CASIA GAIT-A;

Método	Acurácia
<a href="#">Lima e Schwartz (2019)</a>	95.16%
<a href="#">Jangua e Marana (2020)</a> (Distância Euclidiana)	86.29%
<a href="#">Jangua e Marana (2020)</a> (Distância Chi-Quadrado)	89.72%
Método Proposto Treinado no Conjunto GAIT-A	56.15%
Método Proposto Treinado no Conjunto GAIT-B	85.80%

Fonte: Elaborado pelo autor.

# Discussões

- Melhorar a estimação dos parâmetros (intrínsecos e extrínsecos) das câmeras devido seu impacto direto na qualidade das poses 3D:
  - Utilizar um processo de calibração através de *chess-board*;
- Refinar a precisão das poses 3D:
  - Analisar outros modelos de redes neurais para extração de poses 2D;
  - Refinar as poses 3D utilizando o erro de reprojeção das câmeras;
- Fazer um estudo do Top-N ao invés apenas do Top-1:
  - Em image retrieval, por exemplo, a prática de analisar os Top-N melhores resultados é comum.
- Adquirir um maior conjunto de dados para treinar a rede neural de extração do vetor descritor;
- Testar outros algoritmos de estimação de poses 3D utilizando mais de duas câmeras;

---

---

# Conclusão

---

---

# Conclusão

- Provou-se que é possível extrair poses 3D utilizando vídeos de diferentes câmeras e utilizar tal dado para fazer o reconhecimento biométrico;
  - Vantagem de lidar melhor com oclusões devido ao uso de diferentes câmeras;
  - Ideal para aglomerar informações de diferentes câmeras em uma única estimativa;
- Tal abordagem pode ser utilizada sem contato direto com o indivíduo que deseja-se reconhecer;
- Tal abordagem não tem uma dependência direta com vestimenta e acessórios que estejam sendo utilizados no momento;
- Obteve-se resultados competitivos, apesar de não se comparar aos métodos de Lima e Schwartz (2019);



---

---

# Bibliografia

---

---

BIANCO, S.; CIOCCA, G.; MARELLI, D. Evaluating the performance of structure from motion pipelines. Journal of Imaging, v. 4, n. 8, 2018. ISSN 2313-433X. Disponível em: <<https://www.mdpi.com/2313-433X/4/8/98>>.

Center for Biometrics and Security Research. Chinese Academy of Sciences (CASIA). 2010. Acesso em: 10 de Outubro, 2023. Disponível em: <<http://www.cbsr.ia.ac.cn/english/index.asp>>.

COLMAP. Tutorial - COLMAP. 2023. Acesso em: 3 de Outubro, 2023. Disponível em: <<https://colmap.github.io/tutorial.html>>.

GAO, X.; ZHANG, T. Introduction to Visual SLAM. 1st. ed. Singapura: Springer, 2021. ISBN 9811649383.

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.

JANGUA, D.; MARANA, A. A new method for gait recognition using 2d poses. In: Anais do XVI Workshop de Visão Computacional. Porto Alegre, RS, Brasil: SBC, 2020. p. 69–74. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/wvc/article/view/13483>>.

LIMA, V. de; SCHWARTZ, R. Gait recognition using pose estimation and signal processing. In: Iberoamerican on Pattern Ecognition. [S.l.]: CIARP, 2019. ISSN 0000-0000.

MathWorks. What Is Camera Calibration?. 2023. Acesso em: 10 de Novembro, 2023. Disponível em: <<https://www.mathworks.com/help/vision/ug/camera-calibration.html>>.



**Obrigado!**

