



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

THIAGO BUARQUE DE GUSMÃO LAFAYETTE

**VALIDAÇÃO DE ESTIMATIVA ANGULAR BASEADA EM DADOS DE RASTREIO
CORPORAL DE CÂMERAS RGB-D E RGB PARA AVALIAÇÃO BIOMECÂNICA**

Recife
2022

THIAGO BUARQUE DE GUSMÃO LAFAYETTE

**VALIDAÇÃO DE ESTIMATIVA ANGULAR BASEADA EM DADOS DE RASTREIO
CORPORAL DE CÂMERAS RGB-D E RGB PARA AVALIAÇÃO BIOMECÂNICA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de mestre em Engenharia Biomédica.

Área de concentração: Computação Biomédica.

Orientadora: Profa. Dra. Alana Elza Fontes da Gama.

Coorientador: Prof. Dr. Marcelo Xavier Natário Teixeira.

Recife

2022

Catálogo na fonte:
Bibliotecária Sandra Maria Neri Santiago, CRB-4 / 1267

L161v Lafayette, Thiago Buarque de Gusmão.
Validação de estimativa angular baseada em dados de rastreamento corporal de câmeras RGB-D e RGB para avaliação biomecânica / Thiago Buarque de Gusmão Lafayette. – 2022.
92 f.: il., figs., tabs., abrev. e siglas.

Orientadora: Profa. Dra. Alana Elza Fontes da Gama.
Coorientador: Prof. Dr. Marcelo Xavier Natário Teixeira.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Biomédica. Recife, 2022.
Inclui referências.

1. Engenharia biomédica. 2. Sensor RGB-D. 3. Avaliação biomecânica. 4. Kinect V2. 5. Mediapipe. I. Gama, Alana Elza Fontes da (Orientadora). II. Teixeira, Marcelo Xavier Natário (Coorientador). III. Título.

UFPE

610.28 CDD (22. ed.) BCTG/2022-426

THIAGO BUARQUE DE GUSMÃO LAFAYETTE

**VALIDAÇÃO DE ESTIMATIVA ANGULAR BASEADA EM DADOS DE RASTREIO
CORPORAL DE CÂMERAS RGB-D E RGB PARA AVALIAÇÃO BIOMECÂNICA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de mestre em Engenharia Biomédica. Área de concentração: Bioengenharia.

Aprovada em: 23/09/2022.

BANCA EXAMINADORA

Profa. Dra. Fabiana Frata Furlan Peres (Examinador Externo)
Universidade Estadual do Oeste do Paraná

Prof. Dr. Wellington Pinheiro dos Santos (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Sidney Marlon Lopes de Lima (Examinador Externo)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Gostaria de agradecer a todos que me ajudaram no percurso da realização desse trabalho e conclusão da titulação de mestre tanto na esfera acadêmica, quanto na esfera pessoal. Gostaria de iniciar meus agradecimentos a minha orientadora Prof.^a Dr.^a Alana Elza, quem me ajudou e me apresentou a pesquisa e a apaixonante área da engenharia de reabilitação. Em seguida, ao meu coorientador Prof. Dr. João Marcelo Teixeira por ter sido a ponte para o contato com a Prof.^a Alana e uma grande fonte de conhecimento teórico e técnico. Não poderia faltar o Laboratório Voxar Labs, que me acolheu como pesquisador mesmo sendo de uma área não correlata com programação, por isso agradeço a Prof.^a Dr.^a Veronica Teichrieb, chefe do laboratório e de certa forma a madrinha acadêmica de todos, além disso, não poderia faltar meus agradecimentos para Maria de Lourdes Timoteo Euzebio, a pessoa que conseguiu mover montanhas burocráticas, como também um apoio positivo para as dificuldades de equipamento, escala, agenda... entre outros. A Prof.^a Dr.^a Cinthia Vasconcelos do LACAF por abrir o laboratório para mim e me permitir utilizar o Qualisys. Aos alunos da graduação Victor Hugo Kunst (ciência da computação) e Pedro Melo (fisioterapia).

No âmbito pessoal não poderia deixar de agradecer a minha família, primeiramente a minha mãe Helena Gama, que me apoiou em todo o percurso mesmo sem entender nada do que eu faço ou aceitar o fato que eu passava o dia todo no computador sem ser para jogar, mas ela sempre fazia o que podia para me dar o suporte que só uma mãe pode dar. A minha irmã Kyvia Buarque e ao meu pai Arimatea Lafayette. Um agradecimento especial a minha companheira para a vida toda, Maria Carolina Viana, que sem ela, eu teria sem sombras de dúvidas surtado no mestrado, ela que sempre deu amparo aos meus conflitos de ansiedade e frustração com a pesquisa e a pessoa que mais me deu força e coragem para entrar em uma “aventura” nova durante o mestrado, por me ajudar a reconhecer o que estava adormecido em mim e que o mestrado colocou para fora, o desejo de ser médico. Por isso e outros fatores eu sou o homem que sou por sua causa e não há palavras de agradecimentos que possam expressar o quão especial e única ela é para mim.

Agradeço a todos que cruzaram meu caminho nessa jornada acadêmica que é um mestrado. E por fim, a Deus, por tornar tudo possível.

RESUMO

Análise de movimento é uma área com diversas aplicações para saúde, esporte e entretenimento. No campo da saúde, a análise de movimento é um parâmetro básico e essencial para o acompanhamento e evolução de pacientes em reabilitação, entretanto, o alto custo dos equipamentos de ponta inviabiliza a aplicação dessa técnica na rotina das clínicas. Nesse viés, equipamentos RGB-D e RGB, os quais apresentam ferramentas de rastreamento de articulações, estão sendo testados com soluções portáteis e de baixo custo com o intuito de viabilizar a análise de movimento computacional. O recente lançamento do Google MediaPipe, uma técnica de rastreamento de inferência de articulações com câmeras RGB convencionais, pode ser considerado um marco devido à capacidade de estimar coordenadas de profundidade em imagens planares. Diante disso, este trabalho visa avaliar a medição de variação angular mensurada a partir de dados de sensores RGB-D e RGB em frente ao padrão ouro Qualisys Tracking Manager. Um total de 60 gravações foram realizadas para cada movimento de membros superiores e inferiores a partir de 6 voluntários realizando movimentos biomecânicos puros em duas configurações de posição diferentes em relação aos sensores resultando em 600 movimentos angulares das articulações. A utilização do Google MediaPipe obteve resultados próximos em comparação ao sensor Kinect V2 nos aspectos inerentes ao erro absoluto, RMS e correlação ao padrão ouro, apresentando valores de dispersão e métricas de erro menores, ou seja, mais positivo, apresentando apenas a correlação mais baixa que o Kinect V2. No comparativo com equipamentos comumente utilizados em avaliações físicas, o MediaPipe teve um erro dentro da escala de erro de goniômetros de braços curtos e longos. Por se tratar de uma técnica que usa apenas uma câmera RGB convencional, os resultados apontam uma possível alternativa mais acessível para avaliação biomecânica. Nessa visão foi desenvolvida uma aplicação que utilizava o MediaPipe para simular a captura de dados do Kinect V2 em aplicações já desenvolvidas.

Palavras-chave: sensor RGB-D; avaliação biomecânica; kinect V2; mediapipe.

ABSTRACT

Motion analysis is an area with several applications for health, sports, and entertainment. In the health field, motion analysis is a basic and essential parameter for the monitoring and evolution of patients in rehabilitation, however, the high cost of state-of-the-art equipment makes it unfeasible to apply this technique in the clinics' routine. In this vein, RGB-D and RGB equipment, which have joint tracking tools, are being tested with portable and low-cost solutions to enable computational motion analysis. The recent release of Google MediaPipe, a joint inference tracking technique that uses conventional RGB cameras, can be considered a milestone due to its ability to estimate depth coordinates in planar images. In light of this, this work aims to evaluate the measurement of angular variation from RGB-D and RGB sensor data against the Qualisys Tracking Manager gold standard. A total of 60 recordings were performed for each upper and lower limb movement from 6 volunteers performing pure biomechanical movements in two different position configurations concerning the sensors resulting in 600 joint angular movements. Google's MediaPipe usage obtained close results compared to Kinect V2 sensor in the inherent aspects of absolute error, RMS, and correlation to the gold standard, presenting lower dispersion values and error metrics, that is, more positive, presenting only the lower correlation compared to Kinect V2. In the comparison with equipment commonly used in physical evaluations, MediaPipe had an error within the error range of short and long arm goniometers. Because it is a technique that uses only a conventional RGB camera, the results point to a possible more affordable alternative for biomechanical assessment. In this vision, an application was developed that used MediaPipe to simulate and replace the data capture of Kinect V2 in applications already developed.

Keywords: RGB-D; biomechanical analysis; kinect V2; mediapipe.

LISTA DE ILUSTRAÇÕES

Figura 1 –	- Sistema de captura de movimentos óptico passivo.	21
Figura 2 –	Pipeline do método de Mocap usado no Kinect.	23
Figura 3 –	Vista frontal do Kinect V2 ilustrando a posição da câmera e dos sensores.	25
Figura 4 –	Modelo do esqueleto utilizado pelo Kinect V2.	26
Figura 5 –	O processamento de imagens com YOLO é simples e direto.	28
Figura 6 –	A detecção de modelos YOLO como um problema de regressão. Divide a imagem numa grelha $S \times S$ e para cada célula da grelha prediz caixas de delimitação B , confiança para essas caixas, e Probabilidades de classe C . Estas previsões são codificadas como uma $S \times S \times (B \times 5 + C)$ tensor.	29
Figura 7 –	PipeLine Global do OpenPose	30
Figura 8 –	- Fluxograma do funcionamento do MediaPipe.	32
Figura 9 –	O fluxograma de estimativa da pose humana da MediaPipe	33
Figura 10 –	Homem vitruviano alinhado com dois pontos-chave virtuais previstos pelo modelo BlazePose.	34
Figura 11 –	Pontos de referência da MediaPipe.	34
Figura 12 –	Fluxograma geral das etapas e processos realizados na pesquisa.	37
Figura 13 –	- Vista do equipamentos RGB-D. a) Vista traseira dos equipamentos alinhados com a pista de gravação. b) Visão frontal dos equipamentos e tripes de elevação.	39
Figura 14 –	- Representação esquemática da disposição dos equipamentos (sensores RGB-D) e 6 câmeras da Qualisys.	40
Figura 15 –	- Esquemático de posição de marcadores Coda Pelvis.	41
Figura 16 –	- Esquemático de posição de marcadores Thigh 3.	42
Figura 17 –	- Esquemático de posição de marcadores Shank 4.	43
Figura 18 –	Esquemático de posição de marcadores One Segment Feet.	43
Figura 19	- Esquemático de posição de marcadores Thorax/Adbomen Model 2.	44
Figura 20 –	- Esquemático de posição de marcadores Upper Arm and	45

	Lower Arm Segment Model 1.	
Figura 21 –	- Movimentos e posicionamento dos voluntários durante os exercícios. a) Movimento de MI de frente aos sensores. b) Movimento de MS de frente aos sensores. c) Movimento de MI inclinado aos sensores. d) Movimento de MS inclinado aos sensores.	47
Figura 22 –	- Representação do processo de semelhança angular. a) Projeção da frente da câmera. b) Representação sem considerar pontos principais. c) Vista do Eixo X. d) Vista do Eixo Y.	49
Figura 23 –	- Representações das definições da ISB. a) Planos anatômicos. Referências utilizadas para sistema de coordenada: b) ombro e cervical; c) pélvis e lombar; d) cotovelo e joelho; e) punho e tornozelo.	50
Figura 24 –	- Código da criação dos vetores. var y definida como o vetor de origem na espinha do ombro e final na espinha média. Var x produto vetorial do vetor y e o vetor de origem no ombro esquerdo e final no ombro direito (vetor normal ao plano horizontal). Var z produto vetorial entre x e y.	52
Figura 25 –	Gráficos do processo de tratamento dos dados.	55
Figura 26 –	- Dados não sincronizados e interpolados.	56
Figura 27 –	- Dados sincronizados e interpolados.	56
Figura 28 –	- Sistema cartesiano Kinect V2 utilizado.	60
Figura 29 –	- Uso de memória compartilhada na parte em Python.	61
Figura 30 –	- Uso de memória compartilhada no código em C++.	63
Figura 31 –	- Box Plox do Erro Absoluto dos dados gerais. Significância $p < 0,001$ para todos os sensores segundo o teste de rank-sum de Friedman.	68
Figura - 32 –	Gráfico de RMS do Erro Absoluto comparando todos em sensores quanto a dispersão do erro.	69
Figura 33 –	- Box Plot do Erro Absoluto de cada sensor em relação ao Qualisys, para o movimento de membros inferiores.	70
Figura 34 –	- RMS do Erro Absoluto para cada sensor em frente a todos os	71

	dados de Membros inferiores em graus.	
Figura 35 –	- Box Plots agrupados do Erro Absoluto de cada movimento de MMII.	72
Figura 36 –	- Gráficos de RMS do Erro Absoluto agrupado para todos os sensores para coleta de MMII.	74
Figura 37 –	- Box Plot do Erro Absoluto de cada sensor em relação ao Qualisys, para o movimento de membros superiores.	76
Figura 38 –	- Gráfico de RMS do Erro Absoluto para cada sensor em frente a todos os dados de Membros superiores.	77
Figura 39 –	- Box Plots agrupados do Erro Absoluto de cada movimento de MMSS.	78
Figura 40 –	- Gráficos de RMS do Erro Absoluto agrupado para todos os sensores para coleta de MMSS.	80

LISTA DE TABELAS

Tabela 1 –	Especificações das características dos Kinects.	24
Tabela 2 –	- Características dos voluntários.	38
Tabela 3 –	- Especificação dos Equipamentos utilizados	39
Tabela 4 –	- Correlação entre as articulações do Kinect e do MediaPipe.	59
Tabela 5 –	- Classes do Virtual Kinect implementadas com base nas interfaces originais definidas pelo Kinect SDK.	64
Tabela 6 –	- Tabela dos movimentos gravados, explicitando quais não foram analisados, identificados com "o" os positivos e "x" os negativos.	66
Tabela 7 –	- Correlação de Pearson para os dados de variação angular. Para todos os sensores $p < 0,001$ Azul ($>0,75$ e $<0,9$) e Amarelo ($>0,5$ e $<0,75$).	69
Tabela 8 –	- Correlação de Pearson dos sensores para todos os dados para dados de MMII.	71
Tabela 9 –	- Tabela de Medianas do Erro Absoluto para membros inferiores em graus ($^{\circ}$).	73
Tabela 10 –	- Valores de RMS do Erro Absoluto e média para membros inferiores em graus ($^{\circ}$).	74
Tabela 11 –	- Correlação de Pearson dos sensores para cada movimento de membros inferiores. Verde ($>0,9$), Azul ($>0,75$ e $<0,9$), Amarelo ($>0,5$ e $<0,75$) e Vermelho ($<0,5$).	75
Tabela 12 –	- Correlação de Pearson dos sensores para todos os dados de MMSS. Azul ($>0,75$ e $<0,9$).	77
Tabela 13 –	- Tabela dos valores de mediana do Erro Absoluto para cada movimento avaliado.	79
Tabela 14 –	- Valores de RMS do Erro Absoluto e média para membros superiores.	80

LISTA DE ABREVIATURAS E SIGLAS

3D	Espaço tridimensional
E _A	Erro Absoluto
FPS	Frames per second
iOS	IPhone Operating System
IoT	Internet of things
IR	Infravermelho
LACAF	Laboratório de Cinesiologia e Avaliação Funcional
ML	Machine Learning
MMII	Membros Inferiores
MMSS	Membros Superiores
Mocap	Captura de Movimento
PLM	Pose Landmark Model
RGB	Red-Green-Blue
RGB-D	Red-Green-Blue-Depth
ROI	Região de Interesse
SDK	Software Development Kit
ToF	Time-of-flight
UFPE	Universidade Federal de Pernambuco
YOLO	You Only Look Once

SUMÁRIO

1	INTRODUÇÃO	14
1.1	JUSTIFICATIVA	16
1.2	OBJETIVOS GERAIS	16
1.3	OBJETIVOS ESPECÍFICOS	17
2	REVISÃO DA LITERATURA	18
2.1	APLICAÇÕES EM SAÚDE	18
2.2	MOTION CAPTURE	19
2.2.1	Mocaps Ópticos	20
2.2.1.1	Marcadores Passivos	20
2.2.1.2	Marcadores Ativos	21
2.2.1.3	Sem Marcadores	22
2.3	MICROSOFT KINECT	23
2.3.1	Kinect V2: Especificações	25
2.3.2	Kinect V2 Software Development Kit (SDK)	27
2.4	SOLUÇÕES DE ESTIMATIVA/RECONHECIMENTO DE POSE	27
2.4.1	Yolo	27
2.4.2	OpenPose	29
2.4.3	Google MediaPipe	31
3	METODOLOGIA	36
3.1	SUJEITOS	37
3.2	MATERIAIS E SETUP	38
3.3	PROTOCOLO EXPERIMENTAL	46
3.4	PROCESSAMENTO E TRATAMENTO DOS DADOS	47
3.4.1	Sobreposição da imagem RGB e profundidade	48
3.4.2	Cálculos angulares	50
3.4.3	Tratamento dos Dados	53
3.4.4	Variáveis de validação e estatística	57
3.5	VIRTUAL KINECT	58
4	RESULTADOS	65
4.1	PERDA DE DADOS	65
4.2	DADOS GERAIS	67

4.3	MEMBROS INFERIORES	69
4.3.1	Análise Geral	70
4.3.2	Análise por movimento	71
4.4	MEMBROS SUPERIORES	75
4.4.1	Análise Geral	75
4.4.2	Análise por movimento	77
5	DISCUSSÃO	82
5.1	LIMITAÇÕES E DIFICULDADES	85
5.2	TRABALHOS FUTUROS	85
	REFERÊNCIAS	87

1 INTRODUÇÃO

A avaliação biomecânica é uma ferramenta bastante útil no auxílio de tratamentos indicados, tanto para pacientes em geral, que apresentam algum grau de comprometimento motor, quanto para sua utilização por atletas que buscam o aperfeiçoamento de técnicas (BAKER, 2006). A evolução tecnológica da captura de movimento (Mocap) possibilitou a mudança da forma de avaliar o movimento do corpo humano, partindo-se da percepção visual do profissional para a interpretação computacional do movimento (CAPPOZZO e colab., 2005). Essa interpretação é possível devido à reprodução do mundo físico para o meio digital realizada pelos Mocaps através de um software de computador que traduz a posição do corpo no mundo real para uma representação 3D e, assim, tornando possível o reconhecimento de movimentos. Diversos tipos e modelos de Mocap são encontrados no mercado, os quais podem ser diferenciados entre si pelo princípio físico utilizado: mecânicos, magnéticos, eletrônicos e ópticos. Dentre esses modelos destacam-se os sistemas Mocap ópticos de marcadores reflexivos, sendo considerados padrão ouro para avaliação biomecânica, devido à alta precisão métrica espacial (GIL e MOURA, 2015).

Os sistemas Mocap possibilitam a captura de dados quantitativos objetivos dos pacientes ao decorrer do movimento analisado, como amplitude articular, comparação de execução ao padrão biomecânico, postura e compensações corporais. Para esses fins alguns critérios são fundamentais, como a estimativa especial das articulações e variáveis cinemáticas (espaço, velocidade e aceleração), pois o equipamento precisa apresentar acurácia e precisão para representar adequada e suficientemente a anatomia do paciente e a capacidade de aquisição de dados condizente com a velocidade e frequência da realização do movimento estudado (CAPPOZZO e colab., 2005). Ademais, destacam-se ainda os custos para aquisição do equipamento, bem como o espaço de instalação e manutenção, uma vez que são critérios de viabilidade de compra do equipamento. A relevância da avaliação desses critérios se apresenta na medida que equipamentos Mocap de marcadores reflexivos apresentam altos índices de acurácia e precisão, entretanto, seu elevado custo de compra torna sua aplicação restrita para terapia de reabilitação (YANG e colab., 2014).

Nesse viés, alternativas de baixo custo representam uma solução viável para aplicar métodos Mocap em clínicas e laboratórios com menor poder aquisitivo. Dentre as alternativas passíveis de utilização, citam-se as Câmeras RGB-D (*Red-Green-Blue-Depth*), dispositivos que possuem dois tipos de câmeras que gravam respectivamente imagem de cor (RGB) e imagem de profundidade (*Depth - D*) (D'ANTONIO e colab., 2021). Igualmente, podem ser utilizados dispositivos como *Microsoft Kinect*, *Intel RealSense*, *Orbbec Astra* e *OpenCV OAK-D*, os quais constituem exemplos de câmeras RGB-D que apresentam uma melhor relação custo-benefício, uma vez que possibilitaram a aplicação de captura de movimento de modo mais acessível do ponto de vista econômico-financeiro, além de serem portáteis e não apresentarem a necessidade de utilização de marcadores físicos externos, tornando sua utilização mais prática e cômoda para o paciente. Contudo, apesar das vantagens econômicas inerentes, é um fato inconteste que essas ferramentas apresentam um *trade off* em relação a acurácia e precisão e, assim, sua utilização para alguns tipos de análise se apresenta limitada.

Ademais, o desenvolvimento de processamento de imagem para rastreamento juntamente com a aprendizagem de máquina possibilitou a aquisição de dados de profundidade apenas com imagem de cor RGB (Red-Green-Blue) (VISWAKUMAR e colab., 2019). Algoritmos, como OpenPose, MediaPipe, AlphaPose, Detectron2 etc., permitem a identificação do corpo humano e estimativa da posição de articulações em um espaço tridimensional (3D) (CAO e colab., 2021). Dessa maneira, a aplicação de técnicas como essas pode viabilizar a utilização de Mocap a um custo ainda inferior aos modelos RGB-D, visto que essa técnica utiliza apenas imagens RGB convencionais providas de qualquer tipo de câmera, como câmeras digitais e webcams, e não dispositivos compostos por dois tipos diferentes de sensores.

A utilização de Mocap em ambiente clínico demanda acurácia, visto que sutis mudanças na amplitude de movimento representam melhora ou piora do quadro de saúde do paciente, sendo necessária a observância quanto à adequação do equipamento para o fim proposto. De modo geral, nota-se que equipamentos de captura RGB-D e RGB se apresentam como uma nova perspectiva para avaliação biomecânica, devido as suas características vantajosas quanto ao custo e portabilidade. Diante disso, faz-se necessária uma maior compreensão da possibilidade de aplicação e intercambialidade dos dispositivos e dados gerados.

Além disso, o contexto de pandemia tornou evidente a questão das necessidades de avanços em telemedicina e acompanhamento remoto, como também a dificuldade em pesquisas e compartilhamento de equipamentos. Com base nessa necessidade, este trabalho tem por objetivo a realização de uma análise e avaliação comparativas entre diferentes dispositivos RGB-D, comparando-os com um algoritmo para RGB, tendo como parâmetro de confiança um Mocap de marcadores reflexíveis, além de um método de substituição em uma aplicação de reabilitação do equipamento RGB-D por um RGB.

1.1 JUSTIFICATIVA

A análise de movimento é uma ferramenta de grande poder dentro da área de saúde, especificamente na área de reabilitação. Equipamentos RGB-D são ferramentas de captura de movimento com estudos de sua aplicação em diversas modalidades na área de reabilitação, ao passo que equipamentos RGB representam a possível vanguarda tecnológica sobre o tema. A diminuição de custos pode facilitar a difusão e utilização da análise computacional do movimento, fato este que pode representar um impacto substancial na prestação de serviço de saúde, em especial, na qualidade dos tratamentos propostos para os pacientes. Nesse ponto, destaca-se a transição de pesquisas de aplicação de aparelhos RGB-D para RGB, uma vez que a mudança de algoritmos pela alteração da fonte dos dados resultaria em reinício de pesquisas e atrasos de entregas, dessa forma técnicas que simulem métodos RGB-D podem ser uma ferramenta útil para a continuidade de pesquisas e meios de transição de técnicas de pesquisa. Considerando esta realidade, o adequado conhecimento da capacidade de rastreamento de estimativa de articulações é fundamental para a utilização de equipamentos RGB-D com sistemas Mocap, bem como a própria possibilidade de realização de substituições dos equipamentos, com intuito de ofertar um melhor serviço no âmbito da saúde.

1.2 OBJETIVOS GERAIS

O presente trabalho visa determinar o nível de estimativa angular de modelos baseados em esqueletos gerados a partir de imagens de câmeras RGB-D, tendo por parâmetro um equipamento Mocap padrão ouro, além de utilizar as imagens de cor

em um algoritmo de estimativa de articulações para compará-lo a outros equipamentos. Igualmente, tem-se por objetivo o desenvolvimento de um algoritmo que permita a utilização de método RGB em um software programado para um dispositivo RGB-D, com vistas a avaliar a possibilidade da substituição simples.

1.3 OBJETIVOS ESPECÍFICOS

Os objetivos específicos desse trabalho são:

- Estudar e selecionar modelos de equipamentos RGB-D para o trabalho;
- Estudar e selecionar modelos de equipamentos RGB para o trabalho;
- Rastrear posicionamento 3D de articulações a partir de imagens RGB;
- Comparar dados dos equipamentos com o padrão ouro;
- Desenvolver um algoritmo que permita a substituição de um dispositivo RGB-D por um RGB em uma aplicação já implementada;
- Simular a equipamento RGB-D com câmera RGB em aplicação;
- Aplicar processo de revisão e aprimoramento ao longo do desenvolvimento.

2 REVISÃO DA LITERATURA

O trabalho apresenta um conteúdo interdisciplinar entre conhecimentos de engenharia e análise de movimento, abordando a análise comparativa entre diversos equipamentos de captura ópticos em frente a um equipamento padrão ouro. Para isso, o entendimento das tecnologias atuais e seu funcionamento nortearam a escolha do método e sua aplicabilidade, considerando os conceitos sobre os fenômenos físicos abordados para a utilização da ferramenta. Desse modo essa seção conterá: aplicações de análise de movimento, as classificações de sistemas de captura de movimento e técnicas de reconhecimento de pose e estimativa de articulações.

2.1 APLICAÇÕES EM SAÚDE

A necessidade de quantificar o movimento faz o Mocap ser uma ferramenta de grande auxílio em tratamento e acompanhamento de doenças (ABDULHAY e colab., 2018). A análise biomecânica não necessita obrigatoriamente da utilização de métodos computacionais, a evolução tecnológica possibilitou análises mais refinadas e robustas, entretanto cada pesquisa apresenta suas especificidades. A fotometria é a avaliação biomecânica a partir de fotos, podendo ser considerada o início do Mocap, atualmente ainda aplicadas em situação clínicas para mãos (O'BRIEN e colab., 2021) até avaliação postural (BOSSO e GOLIAS, 2012).

Há diversas formas e técnicas de avaliação e com objetivos diversos, desde casos clínicos a aperfeiçoamento e estudo de esporte. Em esporte o entendimento de biomecânica pode auxiliar na evolução e compreensão do esporte, por exemplo na avaliação do movimento do tronco em golfe (BEAK e colab., 2013), como também no lançamento em beisebol (BERNARDINA e colab., 2019).

Ademais, no aspecto de saúde diversas aplicações de Mocap podem ser vistas ao emprego, desde tratamento e identificação do grau de doenças até cuidados hospitalar e residencial. Doenças com impacto na marcha como Parkinson (ABDULHAY e colab., 2018; OTTE e colab., 2020; SHETTY, [S.d.]), AVC (DOLATABADI e colab., 2017; VISWAKUMAR e colab., 2019) e até o próprio envelhecimento são largamente estudadas por sistemas Mocap (ALBERT e colab., 2020). As combinações de sistema Mocap com diferentes princípios de

funcionamento visa melhorar a captura na coleta de dados redundantes. Ademais, muitas vezes ocorre a utilização de sistemas ópticos com marcadores reflexivos (padrão ouro) para a validação de outros métodos de captura (CIPPITELLI e colab., 2015; D'ANTONIO e colab., 2021; SONG e colab., 2016; VILAS-BOAS, Maria Do Carmo e CHOUPINA e colab., 2019). Após os métodos ópticos, é notória a atualização de métodos de equipamentos de princípio inercial (BOLINK e colab., 2016; HAFER e colab., 2020; MARTÍN e colab., 2021).

2.2 MOTION CAPTURE

Os sistemas Mocap são as tecnologias que capturam e transferem para um computador o movimento do mundo real para um espaço virtual, havendo atualmente uma variação de maneiras que podem ser utilizadas para a captura. Devido a essa diversidade, uma classificação baseada no principal recurso físico é utilizada para diferenciar os tipos, sendo classificados como mecânicos, magnéticos ou ópticos. Nesse ponto, ressalta-se que não é possível afirmar que tipo é melhor em relação a outro, uma vez que cada tipo apresenta vantagens e desvantagens, a depender do fim para o qual se deseja sua aplicação (GOMIDE e colab., 2009). Quando o aspecto rastreado é a qualidade do corpo do movimento e graus de liberdade, por exemplo, os sistemas ópticos podem ser considerados o estado da arte (MENACHE, 2011). Assim, essa tecnologia tem sido usada em esportes para analisar parâmetros biomecânicos, como comprimento da passada no beisebol (HANSEN e colab., 2016), rotação do tronco no golfe (BEAK e colab., 2013), padrões de impacto de pé em corredores e melhoria de execuções técnicas (MANUEL e colab., 2017).

Contudo, essa tecnologia também apresenta um alto potencial para aplicações clínicas como, por exemplo, o reconhecimento de marcha de pacientes com Parkinson (DOLATABADI e colab., 2017; LUN e ZHAO, 2015). Os sistemas ópticos Mocap podem ser baseados em marcadores ativos ou passivos, ou até mesmo sem marcadores. Esta revisão incidirá sobre as situações que envolvem marcadores passivos e aquelas sem marcadores, uma vez que estes compreendem a tecnologia que fundamenta o presente estudo.

2.2.1 Mocaps Ópticos

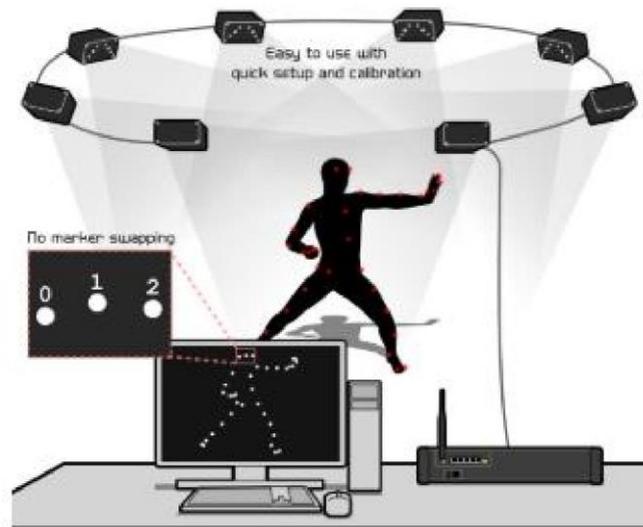
Mocaps ópticos funcionam a partir do tratamento de imagens, ou seja, tratamento de luz capturada por câmeras. O conceito de luz, como onda eletromagnética, é utilizado além dos padrões de luz branca visível e, assim, a faixa de operação desses sistemas abrange luz branca visível e luz infravermelha. Esses sistemas podem ser reconhecidos e classificados por apresentar um equipamento o qual é ligado ao sujeito para refletir ou emitir a luz, a fim de fornecer informações do sujeito a um outro que recebe a informação (KURIHARA e colab., 2002).

2.2.1.1 Marcadores Passivos

Dada a definição sobre os elementos que compõem o sistema, os sistemas passivos são aqueles formados pelo conjunto de marcadores que refletem a luz proveniente do ambiente ou de um equipamento para uma câmera que coleta os dados. Este método é o mais comumente utilizado para captura de corpo inteiro, pois propicia uma maior qualidade de captura e sensibilidade dos movimentos. Existem recursos do modelo infravermelho (IR) que exigem configuração específica, espaço e marcadores retro reflexivos e apresentam alto nível de precisão, conforme representado na Figura 1 (GOMIDE e colab., 2009).

A qualidade da captura é proporcional à quantidade de câmeras e sua distribuição, visto que o próprio movimento pode obstruir os marcadores criando perda de dados e, sendo assim, o aumento de câmeras reduz significativamente esse efeito. Devido a esses fatores, os modelos ópticos IR são uma das tecnologias com mais alto custo, tornando essa tecnologia inviável economicamente para ser utilizada na maioria dos ambientes clínicos (DOLATABADI e colab., 2017).

Figura 1 - Sistema de captura de movimentos óptico passivo.



Fonte: Motta (2016).

Sistemas que utilizam luz visível são chamados de sistemas de câmeras RGB, base de cores para interpretação computacional, as quais consistem nas cores primárias desse sistema, permitindo que qualquer cor possa ser decomposta em um espaço linear. Existem outras bases de cores, porém o termo RGB é utilizado para se referir aos sistemas de luz branca visível. Os sistemas RGB podem usar câmeras simples comerciais, sendo portáteis, de baixo custo e fáceis de usar, para capturar marcadores específicos de cor, forma ou código. No entanto, possuem a desvantagem quanto à possibilidade de serem influenciados pelas características do ambiente, tais como variações de luz, pois a iluminação altera os valores RGB percebidos pelo computador e, assim, confundindo com outros elementos, podendo resultar na captura de dados falsos ou mesmo a perda temporária do marcador (GOMIDE e colab., 2009; KURIHARA e colab., 2002).

2.2.1.2 Marcadores Ativos

Os sistemas ativos são aqueles formados pelo conjunto de marcadores que emitem a luz de forma ativa para uma câmera que coleta os dados (GIL e MOURA, 2015; GOMIDE e colab., 2009). Nos sistemas ativos os marcadores são formados por luzes de LEDs fixadas nas articulações do corpo do indivíduo filmado. Esse método é utilizado em sistemas RGB de modo a diminuir a interferência luminosa que o ambiente pudesse ocasionar em sistemas passivos. De modo geral,

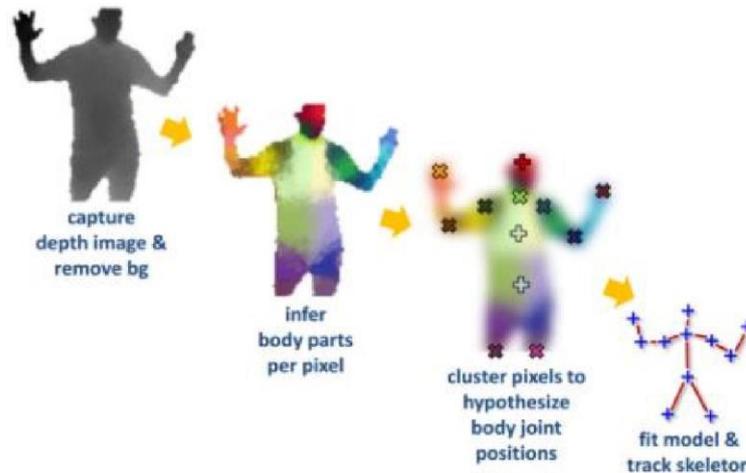
apresentam características semelhantes aos sistemas passivos RGB com a melhoria da sensibilidade à luz, entretanto, apresentam o mesmo problema de fixação dos marcadores, bem como aqueles referentes aos sistemas magnéticos, uma vez que o tempo de captura é diretamente proporcional ao tempo da bateria dos marcadores luminosos (GIL e MOURA, 2015; KURIHARA e colab., 2002).

2.2.1.3 Sem Marcadores

Pela definição sobre os elementos que compõem os sistemas ópticos, os sistemas sem marcadores apresentam a mesma definição que os passivos, embora os elementos reflexivos sejam o próprio corpo do indivíduo filmado, para isso em paralelo o software segmenta o corpo dos elementos de fundo da imagem. A grande evolução desses sistemas consiste na ausência de preparação do indivíduo a ser filmado, uma vez que não há equipamentos adicionais ou especiais. As câmeras são o único equipamento necessário, pois o *software* computacional interpreta as imagens e fornece os dados (GIL e MOURA, 2015; HASLER e colab., 2009; LIU e colab., 2013).

Esse sistema funciona com os mesmos princípios aplicados aos sistemas de marcadores passivos (IR e RGB). Através de visão computacional, as imagens são convertidas em estimativas de esqueletos, na maioria das vezes processados em tempo real. Embora não precisem de um tempo de preparação, o ambiente de filmagem para modelos de câmeras de cor convencional é mais controlado, uma vez que são monocromáticos e fechados a fim de facilitar a segmentação (GIL e MOURA, 2015; GOMIDE e colab., 2009; LIU e colab., 2013). Os métodos que utilizam luz infravermelha têm a vantagem de conseguir realizar a identificação com apenas uma câmera. Sistemas como o *Kinect* da *Microsoft* são exemplos da aplicação dessa tecnologia no ramo comercial do entretenimento, a Figura 2 expressa o *pipeline* das etapas de rastreamento realizado pelo *Kinect*, explicado posteriormente na seção 2.3 (GIL e MOURA, 2015; MOTTA, 2016).

Figura 2 - Pipeline do método de MoCap usado no Kinect.



Fonte: Motta (2016).

2.3 MICROSOFT KINECT

Em 2010, a *Microsoft* lançou um sensor, o Microsoft Kinect V1, em conjunto a um SDK, que utiliza uma inteligência artificial para efetuar cálculos em tempo real e reconhecimento de gestos (ZENNARO e colab., 2015). Uma das características apresentadas foi o rastreamento do corpo, fornecendo um esqueleto definido por 20 articulações a uma taxa de 30 quadros por segundo (fps). Como a *Microsoft* criou este dispositivo para o Xbox 360, os programadores centraram a sua utilização no fornecimento de uma boa jogabilidade e experiência de entretenimento. No entanto, para além das funcionalidades em jogos eletrônicos, muitas aplicações de realidade virtual e aumentada começaram a surgir (LUN e ZHAO, 2015).

Após quatro anos, em 2014, a *Microsoft* anunciou a nova versão do sensor, o *Kinect V2*, com maior resolução, maior precisão (SAMIR e colab., 2016), maior campo de visão (abertura de câmera), e um esqueleto definido por 25 articulações. Estas melhorias significam mais informação sobre a pose estimada e o ambiente, devido à nova técnica utilizada para capturar profundidade que passou da luz estruturada para o *Time-of-Flight* (ToF) (ZENNARO e colab., 2015).

Alguns anos depois do *Kinect V2* ter sido descontinuado, a *Microsoft* lançou o *Azure Kinect* em 2019. No entanto, ao contrário das versões anteriores, o *Azure Kinect* foi concebido e construído para o desenvolvimento em vez de se concentrar no mercado dos jogos. A *Microsoft* o rotulou como um kit de desenvolvimento com sensores avançados de IA para a construção de modelos computacionais de visão e

fala (TÖLGYESSY e colab., 2021). Tal como a versão anterior, o *Azure Kinect* vem com uma câmera de profundidade ToF e uma câmera RGB que suporta uma resolução de até 3840 × 2160 pixels. Contudo, o *Azure Kinect* também apresenta dois modos de detecção de profundidade (campo de visão estreito e campo de visão largo), cada um com duas opções operacionais: *binned* e *unbinned* (TÖLGYESSY e colab., 2021).

A diferença entre as três versões do *Kinect* previamente mencionadas é demonstrada na Tabela 1. É possível ver que as resoluções globais da câmera e o campo de visão têm sido melhorados ao longo do tempo.

Tabela 1 - Especificações das características dos Kinects.

	Kinect V1	Kinect V2	Azure Kinect
Resolução da câmera de cor	1280 × 720 px @ 12 fps 640 × 480 px @ 30 fps	1920 × 1080 px @ 30 fps	3840 × 2160 px @ 30 fps
Resolução da câmera de profundidade	320 × 240 px @ 30 fps	512 × 424 px @ 30 fps	NFOV unbinned—640 × 576 @ 30 fps NFOV binned—320 × 288 @ 30 fps WFOV unbinned—1024 × 1024 @ 15 fps WFOV binned—512 × 512 @ 30 fps
Tecnologia do sensor de profundidade	Structured light-pattern	ToF (Time-of-Flight)	ToF (Time-of-Flight)
Campo de visão (imagem de profundidade)	57° H, 43° V alt. 58.5° H, 46.6°	70° H, 60° V alt 70.6° H, 60°	NFOV unbinned—75° × 65° NFOV binned—75° × 65° WFOV unbinned—120° × 120° WFOV binned—120° × 120°
Especificações da medição de profundidade	0.4–4 m @ 30 fps	0.5–4.5 m	NFOV unbinned—0.5–3.86 m NFOV binned—0.5–5.46 m WFOV unbinned—0.25–2.21 m WFOV binned—0.25–2.88 m
Número de articulações	20	25	32

Fonte: Adaptado de Tölgyessy (2021).

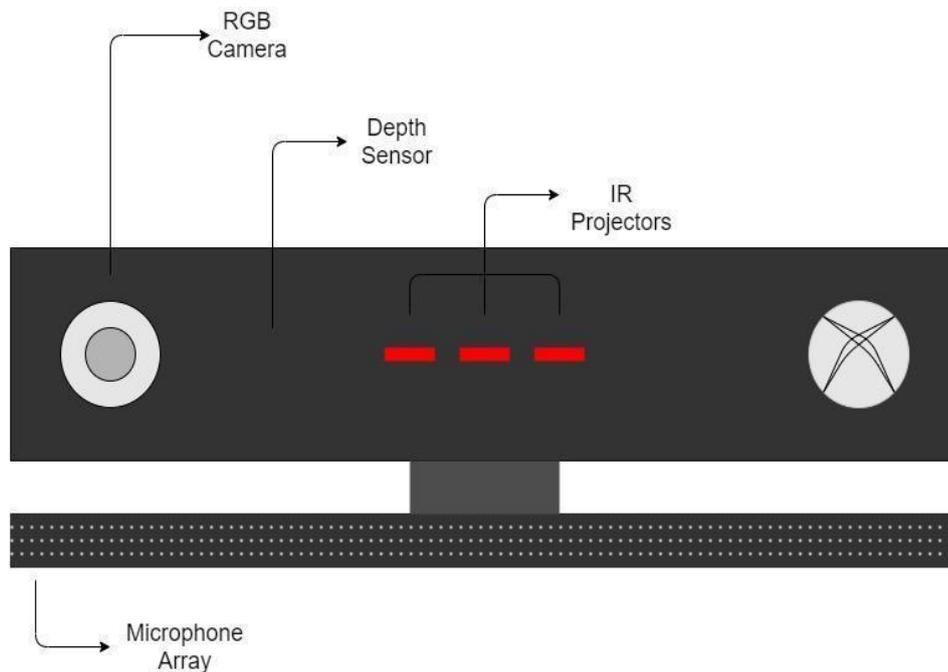
Desde que o *Azure Kinect* foi lançado, não tem sido muito utilizado em pesquisas científicas devido às limitações de compra territorial, à baixa produção e aos níveis de estoque, em comparação com suas versões anteriores. Além disso, considerando as especificações do *Kinect V2* e os resultados de estudos anteriores, o sensor é capaz de desempenhar bem várias tarefas, como o seguimento de

objetos em movimento (PHAM e colab., 2017) e a monitorização do espaço de trabalho (MOHAN e ASHOK, 2018). Considerando essas condições e características, optou-se por fundamentar neste último o sistema apresentado no presente estudo, cujas funcionalidades são descritas em sequência.

2.3.1 Kinect V2: Especificações Técnicas e Funcionalidades

O *Kinect V2* é um dispositivo composto por uma câmera RGB, sensor de profundidade, emissor de infravermelho (IR), e um microfone, como mostra a Figura 3. A informação de profundidade é obtida medindo o tempo das ondas IR para deixar os emissores IR e regressar ao sensor de profundidade (ToF), produzindo um mapa de profundidade.

Figura 3 - Vista frontal do Kinect V2 ilustrando a posição da câmera e dos sensores.



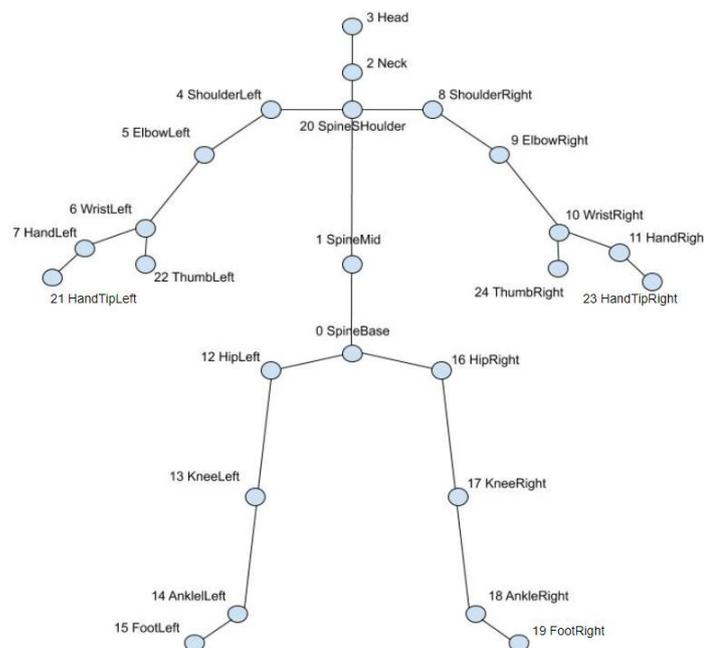
Fonte: O Autor (2022).

A resolução e detalhes dos componentes anteriormente mencionados são mostrados na Tabela 1.

A informação bruta é dada pelo *Kinect V2*, tal como a imagem colorida e de profundidade. No entanto, estes dados são também utilizados nos cálculos para

segmentação corporal por forma e rastreamento baseado no esqueleto de até seis pessoas simultaneamente, tanto em espaços bidimensionais como tridimensionais. Portanto, pode-se distinguir partes do corpo de pessoas diferentes e obter a estrutura do corpo dos esqueletos definidos por 25 articulações, que são apresentadas na Figura 4. Para isso, o algoritmo do Kinect treinou um *Deep Randomized Forest Decision Tree* a partir de imagens de profundidade realistas sintéticas de humanos de muitas formas e tamanhos (corpos segmentados) em poses variadas amostradas a partir de uma grande base de dados de captura de movimento (SHOTTON e colab., 2013).

Figura 4 - Modelo do esqueleto utilizado pelo Kinect V2.



Fonte: Burle (2020).

Em contraste, o *Kinect V2* tem dificuldade para reconhecer os biotipos que não se aproximam das formas corporais padrão na base de dados de treino, como por exemplo discrepância de membros e amputações (LAFAYETTE e colab., 2019). Outra limitação do *Kinect V2* é a influência da temperatura, por ser uma câmera ToF, é necessário um tempo substancial de pré-aquecimento (LACHAT e colab., 2015) e a temperatura tem uma forte correlação com a medição da distância (WASENMÜLLER e STRICKER, 2017).

2.3.2 Kinect V2 Software Development Kit (SDK)

Juntamente com o *Kinect V2*, a *Microsoft* também lançou um *Software Development Kit* (SDK) para ajudar os programadores a criar as suas aplicações utilizando o sensor. São fornecidos mais de 10 códigos de exemplo em C# e C++ ilustrando dados como profundidade, cor, articulações do esqueleto, por exemplo. Além disso, algumas ferramentas estão incluídas no SDK, como o *Kinect Studio*, que pode ser utilizado para gravar, monitorar e reproduzir dados do sensor, permitindo melhores testes e depuração de uma determinada aplicação sem a necessidade de gravações em tempo real e disponibilidade do sensor.

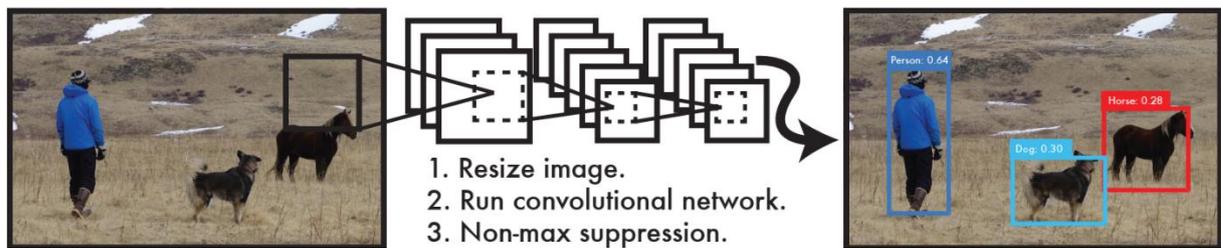
2.4 SOLUÇÕES DE ESTIMATIVA/RECONHECIMENTO DE POSE

A estimativa e reconhecimento da pose humana tem sido utilizada em várias aplicações, desde a fisioterapia em clínicas até ao ajustamento da postura nas indústrias. Para este fim, vários algoritmos de processamento de imagem utilizam as CNNs (*Convolutional neural network*) para identificar pontos-chave. Algumas das tecnologias disponíveis são *YOLO*, *OpenPose* e *MediaPipe*, que serão descritas nas subsecções seguintes.

2.4.1 Yolo

Lançado em 2016, *You Only Look Once* (YOLO) é um método para detectar objetos numa imagem (PLASTIRAS e colab., 2018). Pode detectar objetos como um único problema de regressão, diretamente desde os pixels da imagem até as coordenadas da caixa delimitadora e probabilidades de classe. Para isso, uma única rede convolucional prevê simultaneamente múltiplas caixas delimitadoras e probabilidades de classe para estas caixas e otimiza diretamente o desempenho de detecção, como se mostra na Figura 5.

Figura 5 - O processamento de imagens com YOLO é simples e direto. (1) redimensiona a imagem de entrada para 448×448 , (2) executa uma única rede convolucional na imagem, e (3) limita as detecções resultantes pela confiança do modelo.

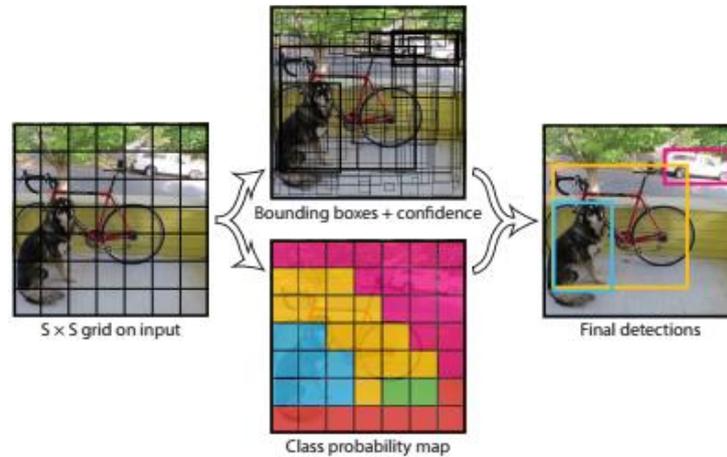


Fonte: Plastiras e colab. (2018)

A visto disso, YOLO é extremamente rápido, a sua rede de base funciona a 45 quadros por segundo (sem processamento em lote) numa GPU Titan X e a versão rápida funciona a mais de 150 fps. Assim, é possível processar vídeo em tempo real com menos de 25 milissegundos de latência. Além disso, o YOLO comete menos de metade dos erros envolvendo o fundo da imagem em comparação com outros métodos que utilizam um método com R-CNN rápido. Uma vez que YOLO é altamente generalizável, é menos provável que apresente problemas de identificação quando aplicado a novos ambientes ou entradas inesperadas (PLASTIRAS e colab., 2018).

A forma como o YOLO foi projetada permite o treinamento ponta a ponta e processamento em tempo real ao mesmo tempo em que mantém uma alta precisão na média. Entende-se por treinamento ponta a ponta a seleção das imagens a serem reconhecidas e o resultado do reconhecimento usando a rede genérica do mesmo, sem a necessidade de modificar os passos intermediários do reconhecimento. Para tal, YOLO divide a imagem de entrada numa região de interesse $S \times S$. Se o centro de um objeto cair numa ROI (*Region of Interest*), essa região é responsável pela detecção desse objeto. Cada ROI prevê B delimitando caixas e pontuações de confiança para essas caixas. Estas escalas de confiança refletem quão preciso é a estimativa do objeto contido no ROI, como também a própria posição da caixa (PLASTIRAS e colab., 2018). A Figura 6 ilustra o fluxo do processo realizado pelo YOLO.

Figura 6 - A detecção de modelos YOLO como um problema de regressão. Divide a imagem numa ROI $S \times S$ e para cada célula da ROI prediz caixas de delimitação B , confiança para essas caixas, e probabilidades de classe C . Estas previsões são codificadas como um tensor $S \times S \times (B \times 5 + C)$.



Fonte: Plastiras e colab. (2018)

A partir da capacidade de detecção do YOLO, torna-se possível a sua aplicação para estimativa de articulações (HUANG e colab., 2021). Tekin et al. propuseram uma estrutura YOLO para a estimativa de 3D de objetos de mão, reconhecimento de objetos, e classificação de atividade num único processo de alimentação (TEKIN e colab., 2019). Cheng et al. utilizam o algoritmo de estimativa de pose CPM baseado na aprendizagem profunda para reconhecer as articulações humanas. O algoritmo YOLOv3 é inicialmente aplicado para localizar e avaliar as ferramentas de montagem e reconhecer a ação de montagem dos trabalhadores (CHEN e colab., 2020).

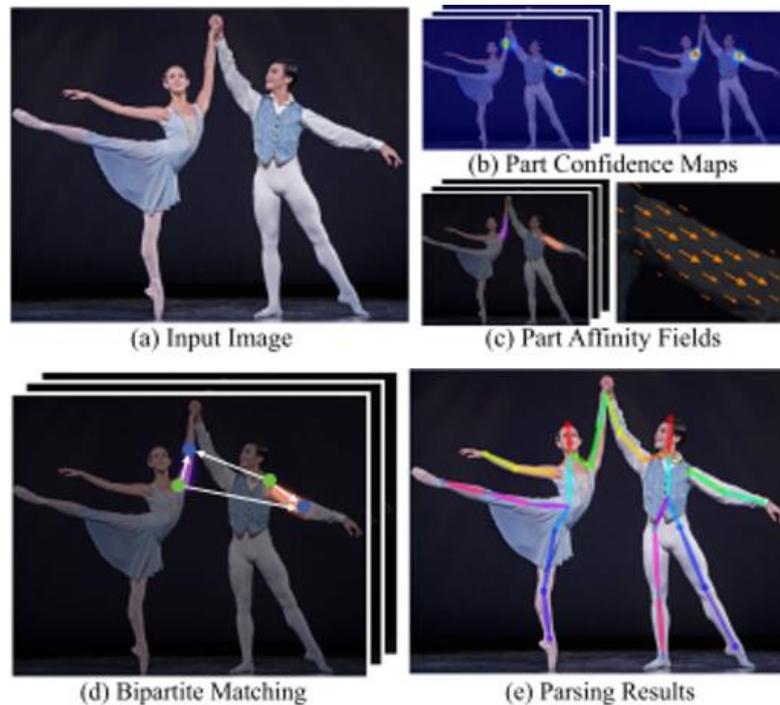
2.4.2 OpenPose

OpenPose é um método centrado na estimativa da pose humana 2D lançado em 2019. O que torna *OpenPose* diferente é a sua capacidade de *multi-tracking* de números indeterminados de indivíduos, incluindo os desafios de posições inesperadas, escalas diferentes, interações entre pessoas induzindo interferências espaciais complexas devido ao contato, oclusão, ou articulações (CAO e colab., 2021).

A abordagem do *OpenPose* apresenta uma representação bottom-up para os escores de associação usando *Part Affinity Fields* (PAFs), criando uma configuração

de vetores planares que determinam a posição e orientação dos membros na região da imagem, como mostra a Figura 7.

Figura 7 - Pipeline global do OpenPose. (a) OpenPose toma a imagem inteira como entrada para uma CNN prever em conjunto (b) mapas de confiança para detecção de partes do corpo e (c) PAFs para associação de partes. (d) A etapa de análise realiza um conjunto de comparações bipartidas para correlacionar candidatos a partes de corpo. (e) OpenPose finalmente reúne-os em poses de corpo inteiro para todas as pessoas na imagem.



Fonte: CAO e colab. (2021)

O sistema recebe uma imagem RGB como entrada e utiliza uma rede de alimentação para estimar um mapa de confiança 2D da localização das partes do corpo. Depois disso, o mapa de confiança (e os PAFs) é analisado pela inferência *greedy*, criando uma saída com pontos-chave 2D para todas as pessoas da imagem.

Devido à eficiência e facilidade de utilização do algoritmo, *OpenPose* faz parte de muitas aplicações. Chen et al. utilizam *OpenPose* integrado com YOLO para reconhecer ações de montagem repetitivas e estimar os seus tempos de funcionamento (CHEN e colab., 2020). Como Chen, Han et al. combinaram YOLO e *OpenPose* para detectar o comportamento humano (LIN e DING, 2020). Viswakumar et al. aplicam *OpenPose* para analisar a biomecânica da marcha, avaliando o deslocamento dos joelhos (VISWAKUMAR e colab., 2019).

Embora as informações do *OpenPose* sejam planares, é possível obter informações tridimensionais utilizando múltiplas câmeras, bem como métodos de

estereografia. Por esta perspectiva, D'Antonio et al. realizaram um estudo de marcha sobre esteira com o *OpenPose* gerando uma análise 3D das articulações capturadas e conseguiram identificar diferentes *setups* mais propícios para a combinação de 2 câmeras (D'ANTONIO e colab., 2021).

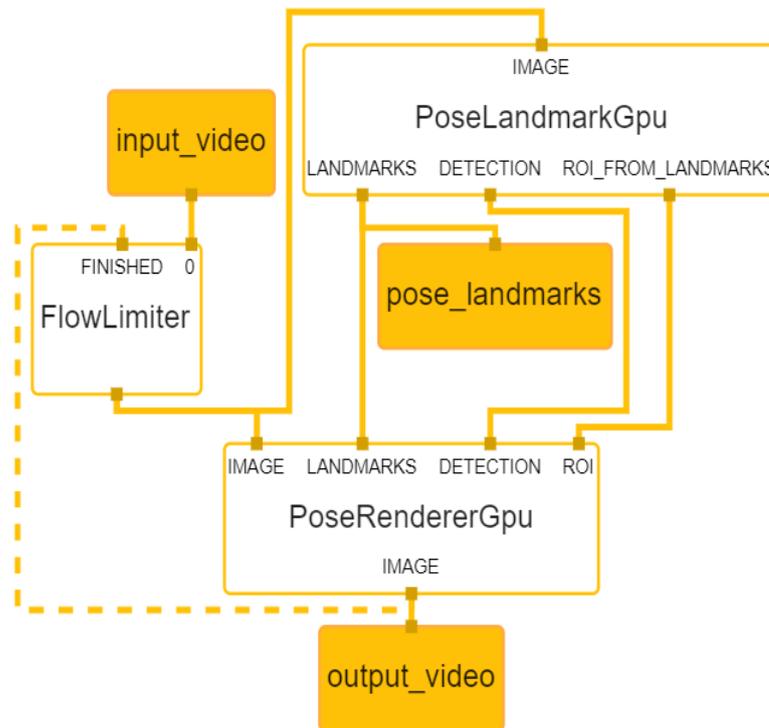
Nesse viés, a aplicação de *OpenPose* pode ser realizada junto a um equipamento RGB-D para em conjunto fornecer informações 3D com apenas um único sensor. Li et al. utilizou *Kinect v1* junto ao *OpenPose* para estimar a postura de crianças na posição de decúbito dorsal (LI e colab., 2021). Wu et al. utilizou *OpenPose* em associação com *Kinect v1* para desenvolver um método de reconhecimento de paralisia cerebral de forma não invasiva em crianças recém nascidas (WU e colab., 2021).

2.4.3 Google MediaPipe

MediaPipe é uma estrutura de código-fonte aberto transversal (por exemplo compatível com Android, iOS, web, desktop, cloud, e IoT) desenvolvida pela Google para construir fluxos de inferências sobre dados multimodais (por exemplo, vídeo, áudio, dados de séries cronológicas) (LUGARESI e colab., 2019). Foi concebido para aqueles que trabalham com a aprendizagem de máquinas, tais como pesquisadores, estudantes, programadores de *software etc.* O *MediaPipe Pose* é uma solução de aprendizagem de máquina para o rastreamento de pose corporal de alta fidelidade, inferindo 33 pontos de referência 3D em todo o corpo a partir de frames de vídeo RGB (ZHANG e colab., 2020).

MediaPipe cria um pipeline de percepção que é construído como um Gráfico (ver na Figura 8), composto de componentes modulares, tais como modelos de inferência (por exemplo, TensorFlow (“TensorFlow”, [S.d.]), TFLite (“GitHub - tensorflow/tflite-support”, [S.d.])). A principal utilização do *MediaPipe* é conceber protótipos de aplicações que fazem uso de modelos de inferência articulares e outros componentes reutilizáveis.

Figura 8 - Fluxograma do funcionamento do MediaPipe.



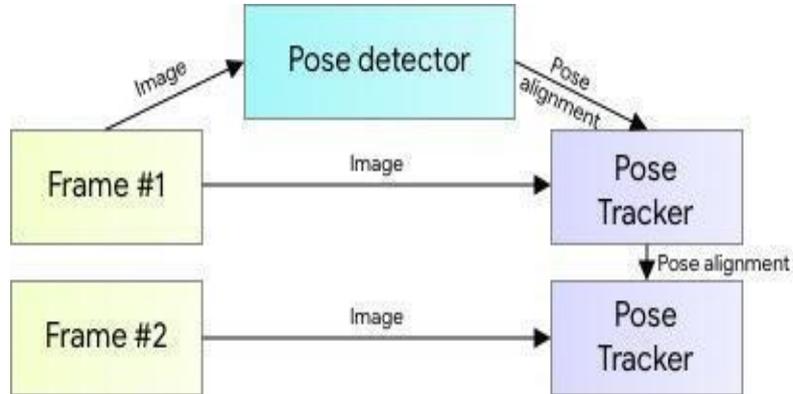
Fonte: Bazarevsky e colab. (2020)

Esta aplicação realiza a captura de diferentes perspectivas, uma vez que é necessário adicionar algumas etapas de processamento e até mesmo componentes de inferência. Além disso, o desenvolvimento da mesma aplicação para plataformas diferentes pode ser ainda mais difícil, considerando a necessidade de otimização de algumas etapas e alteração do processamento para se adequar ao dispositivo alvo. Para lidar com isto, o MediaPipe cria fluxos para abstrair e ligar os diferentes modelos de inferência, que já se encontram otimizados. O mesmo pipeline pode ser utilizado em várias plataformas com o mesmo comportamento, permitindo ao programador gerar uma aplicação num ambiente de trabalho e implementá-la em um telefone celular.

A estimativa de pose para aplicações de saúde e bem-estar (por exemplo, ginásio, dança, etc.) é particularmente desafiadora, em razão da grande variedade de poses possíveis, numerosos graus de liberdade, oclusões, e uma variedade de características corporais. O BlazePose (BAZAREVSKY e colab., 2020) utiliza dois passos, que são um detector e um rastreador para a estimativa da pose. O primeiro (detector de pose na Figura 9), BlazePose Detector, é um detector que localiza a

região de interesse da pose (ROI) dentro da imagem. O segundo (detector de postura na Figura 9), BlazePose GHUM 3D, é utilizado para inferir 33 pontos de referência de pose em todo o corpo neste ROI.

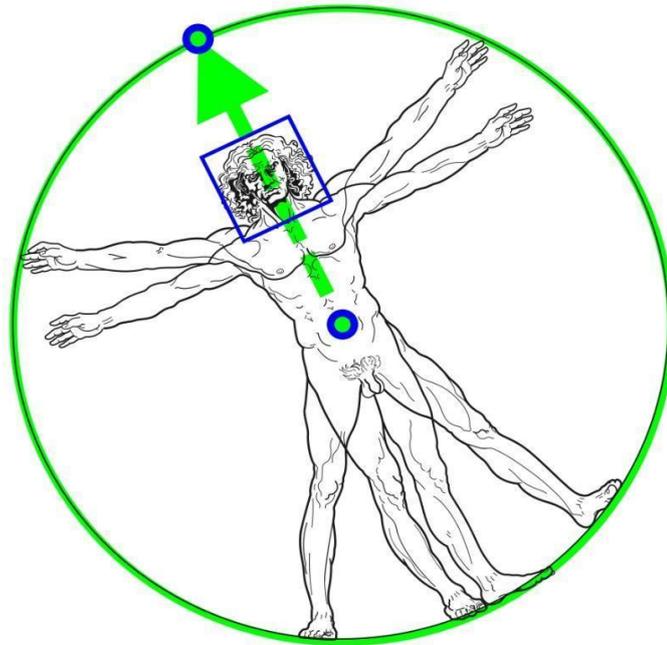
Figura 9 - O fluxograma de estimativa da pose humana do MediaPipe.



Fonte: Bazarevsky (2020).

O Detector BlazePose baseia-se no Modelo BlazeFace (BAZAREVSKY e colab., 2019), que é utilizado na Detecção Facial do MediaPipe. O BlazePose escolheu este detector, uma vez que o sinal mais forte para a Rede Neural sobre a posição do tronco da pessoa é o rosto da pessoa (devido às suas características de alto contraste e variações comparativamente baixas na aparência). O Detector BlazePose prevê dois pontos-chave virtuais que descrevem o centro do corpo humano, a rotação e a escala como um círculo. Inspirado no Homem Vitruviano de Leonardo di Ser Piero da Vinci, o modelo prevê o ponto médio do quadril de uma pessoa, o raio de um círculo que circunscreve a pessoa inteira, e o ângulo inclinado da linha que liga os pontos médios do ombro e do quadril (ver Figura 10 para mais detalhes).

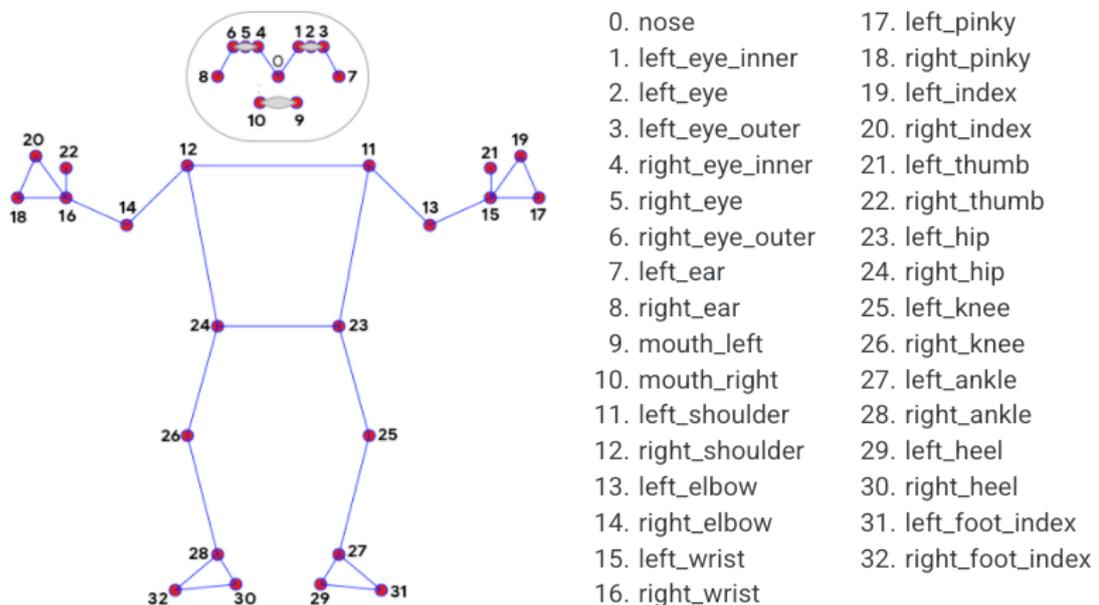
Figura 10 - Homem vitruviano alinhado com dois pontos-chave virtuais previstos pelo modelo BlazePose.



Fonte: Bazarevsky e colab. (2020)

O *Pose Landmark Model* (PLM) permite o rastreamento da pose humana empregando a aprendizagem da máquina, do inglês, *Machine Learning* (ML) para inferir 33 articulações, como demonstrado na Figura 11, marcos 2D de um corpo a partir de um único quadro com 3 graus de liberdade (x, y e visibilidade) mais os dois *keypoints* virtuais anteriormente mencionados.

Figura 11 - Pontos de referência da MediaPipe.



Fonte: Bazarevsky e colab. (2020)

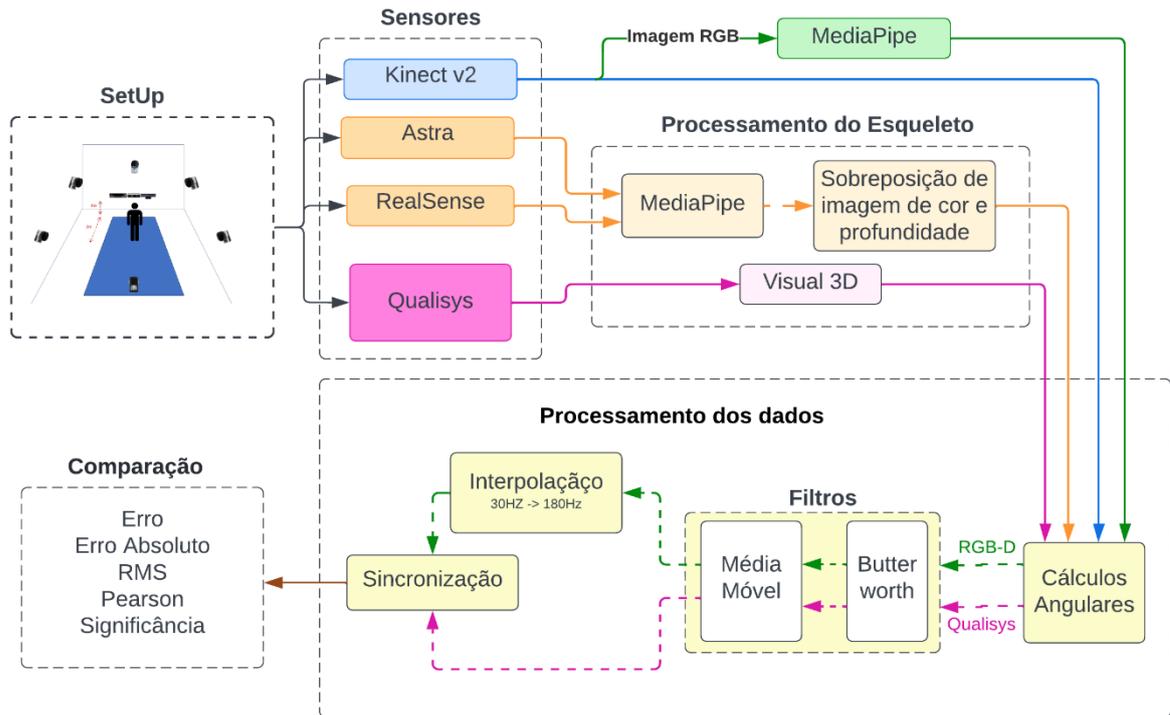
O PLM gera duas possíveis saídas: uma lista de pontos de referência, cada uma consistindo coordenadas normalizadas pela largura e altura da imagem respectivamente (x,y) e uma coordenada z, que representa a profundidade do ponto de referência, sendo a profundidade no ponto médio do quadril a origem (quanto menor o valor, mais próximo o marco está da câmera). Além disso, tem um valor de visibilidade entre 0,0 e 1,0, indicando a probabilidade de o ponto de referência estar visível. A outra saída é uma lista de pontos de referência de pose em coordenadas globais, cada uma contendo x, y e z coordenadas 3D do Mundo Real em metros com a origem no ponto médio do quadril e o mesmo valor de visibilidade que antes.

3 METODOLOGIA

Essa seção introduz os aspectos metodológicos utilizados na confecção deste trabalho, desde a captação de voluntários até a análise final dos dados, conforme demonstrado pela Figura 12 quanto ao processo geral da pesquisa. O processo do experimento consiste na gravação de execução de movimentos biomecânicos monoplanares, inicialmente o voluntário é preparado com a fixação dos marcadores reflexos e posicionado em frente aos sensores para a gravação simultânea. Após a gravação, é necessária a extração dos esqueletos das gravações, o SDK do Kinect V2 fornece a estimativa articular do corpo do voluntário, entretanto o Astra e o RealSense não disponibilizam de modo gratuito o SDK para extração, dessa forma foi necessário uma etapa intermediária de extração utilizando MediaPipe para identificar e rastrear as articulações nas imagens RGB 2D e subseqüentemente realizar a sobreposição com as imagens de profundidade e converter em coordenadas 3D. Por outro lado, para as gravações do Qualisys foi utilizado o software licenciado da empresa Visual 3D para a partir dos marcadores determinar as posições as articulações desejadas.

Ademais, após a obtenção de todos os esqueletos foram determinadas as variações de movimento das articulações de interesse. No cálculo angular, foram utilizados os padrões de referência da Sociedade Internacional de Biomecânica, em inglês, *International Society of Biomachanichs* (ISB) para movimento articular. Durante a obtenção dos dados, é passível a captação de interferências e de ruídos, devido a esse problema dois filtros foram aplicados antes da sincronização dos dados. Os arquivos filtrados apresentavam diferenças de tamanhos, principalmente os provenientes dos Qualisys devido a sua maior taxa de aquisição, por isso foi preciso realizar uma interpolação dos arquivos dos sensores RGB-D e RGB. Para finalizar a preparação dos dados ocorreu a sincronização deles colocando todos em fase. Desse ponto foi possível calcular as variáveis de erro e correlação ao padrão ouro assim como a análise estatística dos dados obtidos.

Figura 12 - Fluxograma geral das etapas e processos realizados na pesquisa.



Fonte: O Autor (2022).

Para a realização desta pesquisa, foram selecionados os equipamentos Microsoft Kinect v2, Orbbec Astra e Intel RealSense como dispositivos de captura RGB-D. Para aplicação RGB foi escolhida a API do Google MediaPipe devido ao seu diferencial de estimar profundidade apenas com uma câmera RGB, além de ser um *framework* novo e com poucos artigos para sua aplicação em rastreamento de articulações. A referência adotada como padrão ouro para comparação das medidas angulares nesse trabalho foi obtida utilizando o sistema de captura de movimento Qualisys.

Como última avaliação de aplicabilidade foi utilizado o MediaPipe para simular a aquisição de dados proveniente do Kinect V2 e enviado para um script já desenvolvido, utilizando um script como máscara que molda os dados do MediaPipe no formato do Kinect.

3.1 SUJEITOS

Foi realizado um estudo transversal experimental no Laboratório de Cinesiologia e Avaliação Funcional (LACAF) da Universidade Federal de Pernambuco

(UFPE). Os critérios de inclusão foram jovens acima de 18 anos e estudantes da UFPE. O único critério de exclusão para a seleção dos sujeitos foi a existência de qualquer doença ou lesão que afetasse sua coordenação e amplitude de movimento, avaliado previamente antes da preparação de gravação por um estudante de fisioterapia, não foram impostas restrições sobre idade, peso, altura e gênero. Os participantes foram recrutados no final do mês de dezembro de 2021. Os voluntários são estudantes do departamento de fisioterapia da UFPE, ao final foram cooptados 5 sujeitos saudáveis: 4 homens e 1 mulher. Os dados demográficos associados são apresentados na Tabela 2. O experimento foi aprovado pelo Comitê de Ética da UFPE, que foi aprovado com o parecer: 3.225.381 (CAAE:03508918.9.0000.5208).

Tabela 2 - Características dos voluntários.

	Média	Desvio Padrão	Mínimo	Máximo
Idade (anos)	21,5	1,37	19	23
Altura (m)	1,69	0,12	1,65	1,96
Peso (kg)	63,65	7,33	55	73
IMC	21,17	2,8	18,22	25,26

Fonte: Autor (2022).

3.2 MATERIAIS E SETUP

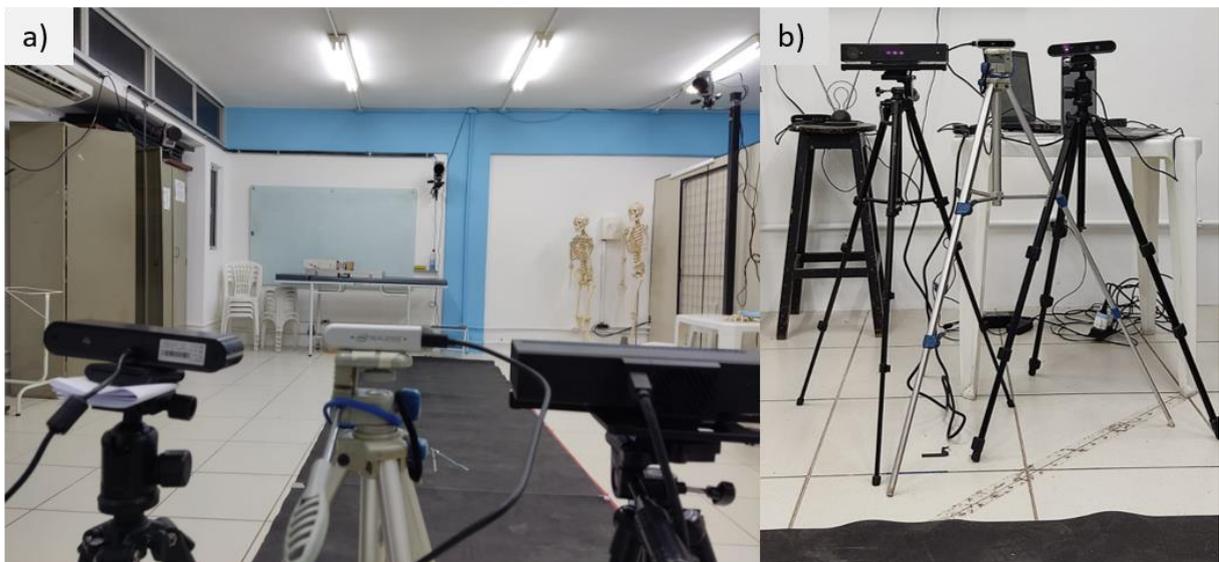
Foram utilizados 3 sistemas Mocap RGB-D: Microsoft Kinect v2, Intel RealSense e Orbbec Astra, cujas especificações técnicas estão descritas na Tabela 3. Além disso, foi utilizado o *framework* Google MediaPipe para rastreamento e estimativa de articulações em imagens RGB. Os equipamentos de captura ficaram dispostos em frente a área de coleta posicionados um ao lado do outro como demonstrado na Figura 13, há uma elevação do solo de cerca de 1.20m.

Tabela 3 - Especificação dos Equipamentos utilizados.

vPropriedades	Kinect v2	Astra	RealSense
Resolução Câmera RGB	1920x1080 px	640x480 px	1920x1080 px
Resolução Câmera de Profundidade	512x424 px	640x480 px	1280x720 px
Taxa de captura	30 fps	30 fps	30 fps
Campo de Visão RGB	70x60 graus	60x49,5 graus	69x42 graus
Campo de Visão de Profundidade	84x53 graus	60x49,5 graus	69x42 graus

Fonte: Autor (2022).

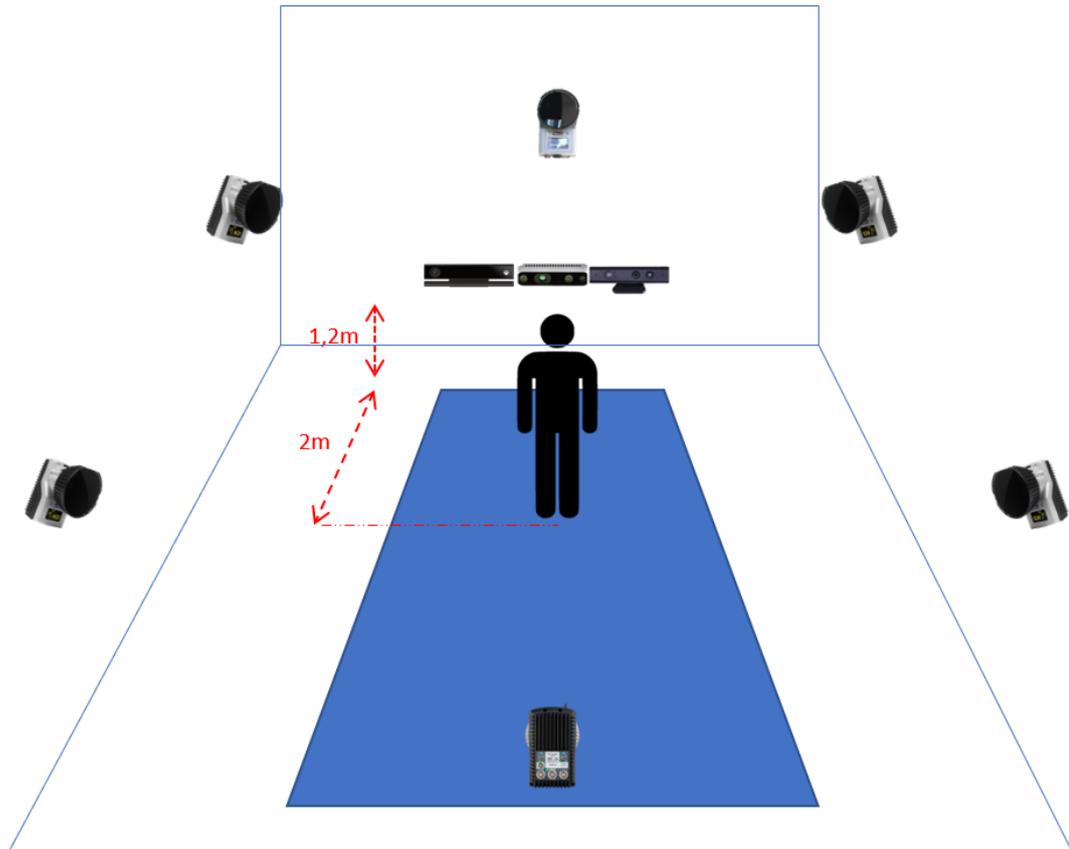
Figura 13 - Vista do equipamentos RGB-D. a) Vista traseira dos equipamentos alinhados com a pista de gravação. b) Visão frontal dos equipamentos e tripes de elevação.



Fonte: O Autor (2022).

Como parâmetro de referência da variação angular foi utilizado o equipamento padrão ouro Qualisys com 6 câmeras. A disposição dos equipamentos foi representada no esquemático na Figura 14.

Figura 14 - Representação esquemática da disposição dos equipamentos (sensores RGB-D) e 6 câmeras da Qualisys.



Fonte: O Autor (2022).

As gravações ocorreram simultaneamente para todos os equipamentos. As frequências de gravação foram diferentes entre os equipamentos RGB-D/RGB e o padrão ouro, sendo respectivamente 30Hz e 180Hz. Ademais, para a preparação dos voluntários para a captura do Qualisys, foram utilizados 24 marcadores reflexivos em duas configurações principais (para membro superior e membro inferior). Para o posicionamento dos marcadores foi seguido o guia disponibilizado site da C-Motion Wiki (“Marker Set Guidelines - Visual3D Wiki Documentation”, [S.d.]). Cada membro precisa ser marcado de modo que cada segmento se complete com o próximo. C-Motion define 4 tipos diferentes de marcadores para o Visual 3D:

- Vermelho: utilizado para definição e rastreamento de segmento
- Verde: apenas rastreamento
- Azul: apenas para definição
- Preto: marcador virtual tanto para definição quanto rastreamento

Com as definições dos tipos de marcadores, foi escolhido um método de identificação para MMII de modo a viabilizar e facilitar o experimento, focando na praticidade sem a perda de qualidade da gravação. Dessa maneira os membros inferiores foram separados em: pelve, coxa, canela e pés. O modelo de pelve utilizado foi o Coda Pelvis apresentando 4 marcações (Espinha Íliaca Antero Superior esquerda e direita, Espinha Íliaca Pósterio Superior esquerda e direita), como demonstrado na Figura 15. As coxas utilizaram o modelo de *Thigh 3* descrito por C-Motion devido à baixa quantidade de marcadores, assim como a presença de marcadores de rastreamento e identificação. Os pontos anatômicos marcados foram: trocanter maior do fêmur, epicôndilo lateral do fêmur e epicôndilo medial do fêmur), como ilustrado na Figura 16.

Figura 15 - Esquemático de posição de marcadores Coda Pelvis

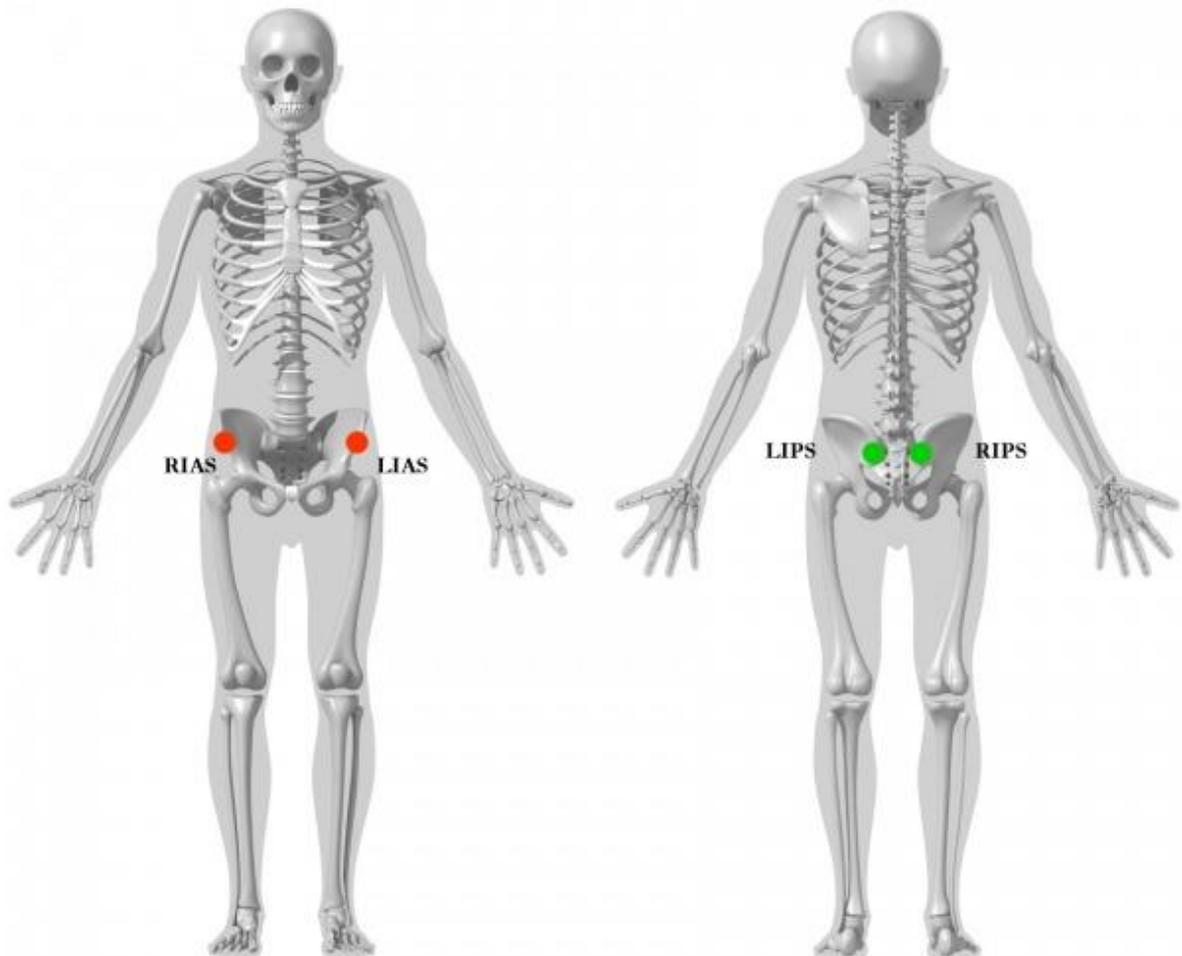
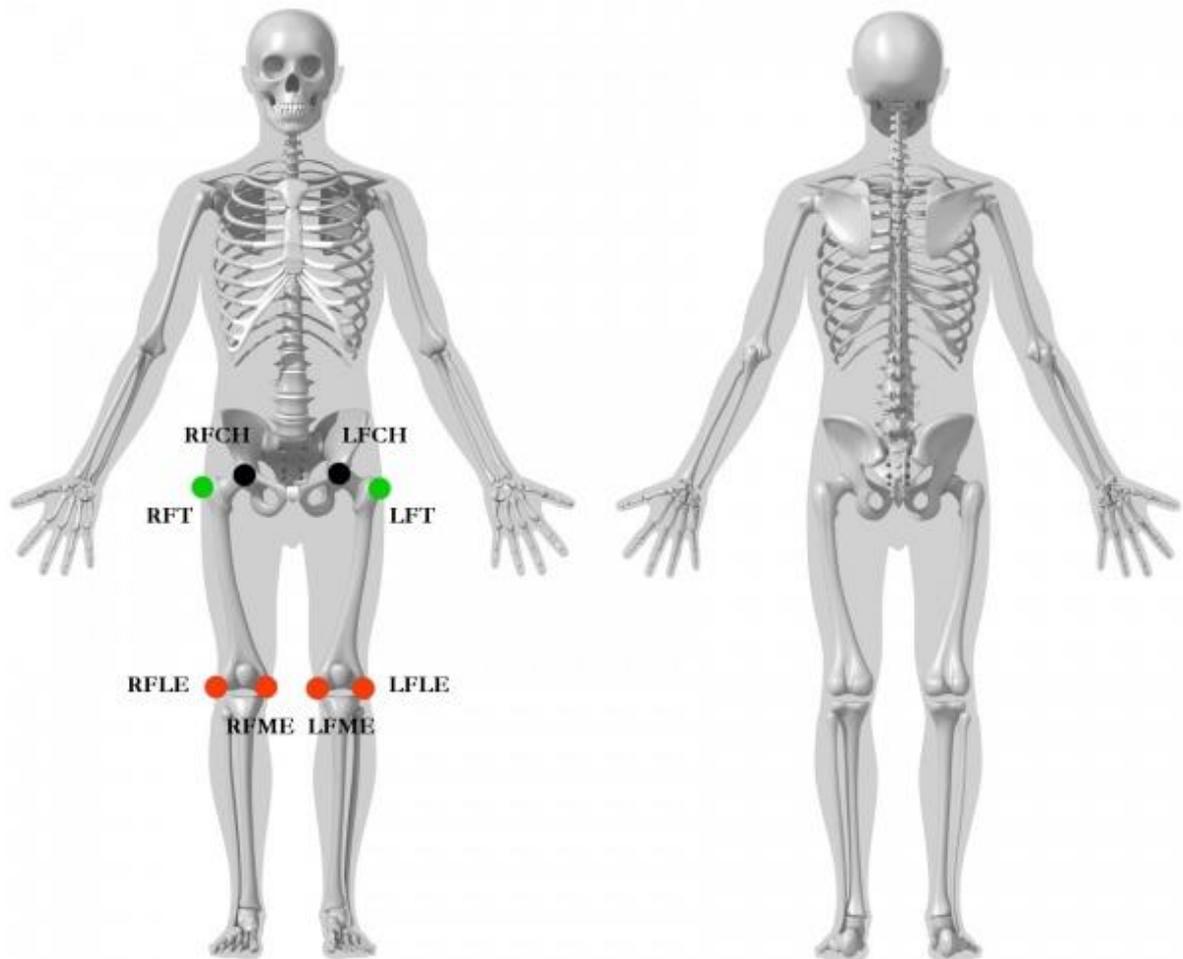


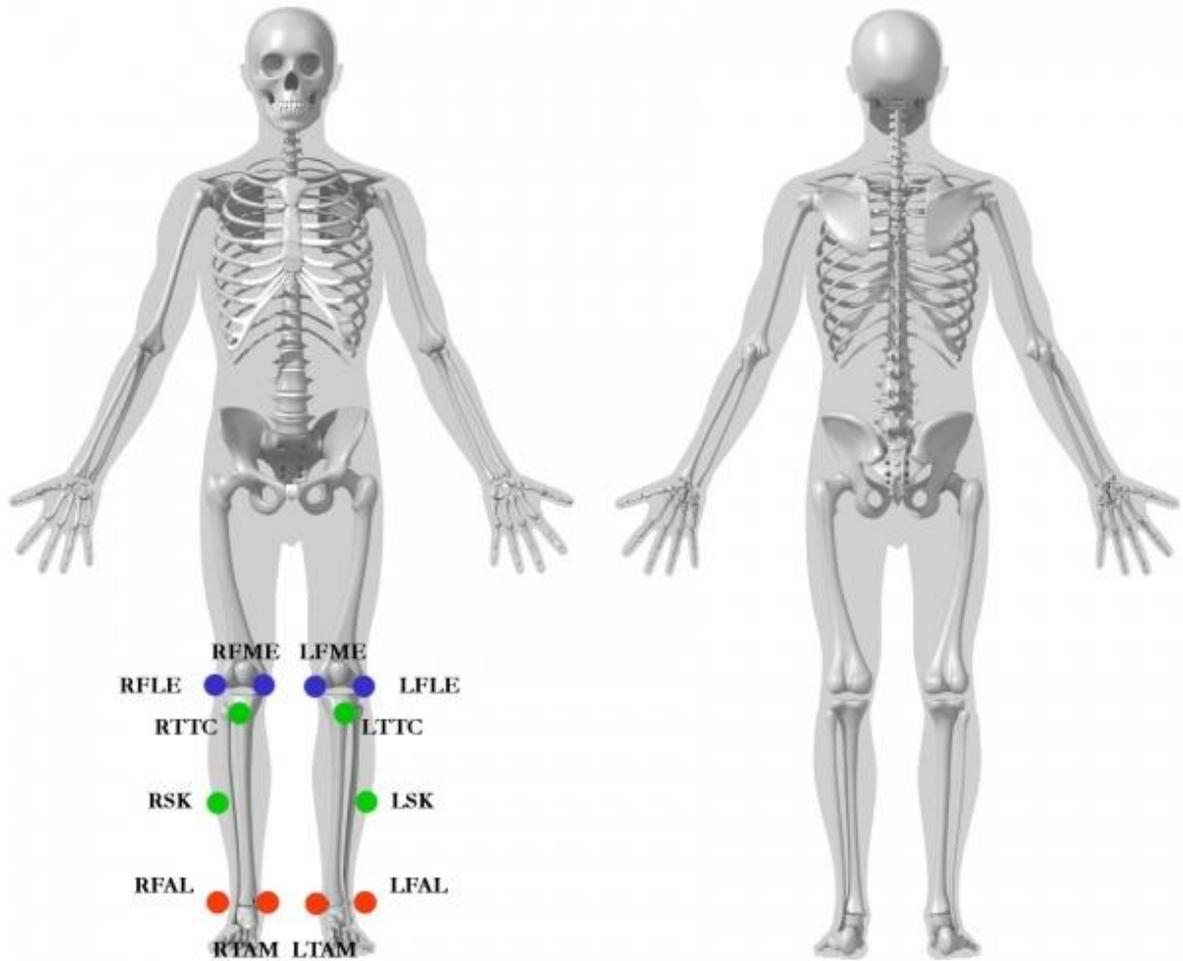
Figura 16 - Esquemático de posição de marcadores Thigh 3



Fonte: Marker Set Guidelines (2022).

A canela foi marcada seguindo o modelo de *Shank 4* integrando as marcações dos epicôndilos do fêmur. Assim, foram marcados os seguintes pontos anatômicos: tuberosidade da tíbia, a metade do comprimento da fíbula (posição estimada por medição com fita métrica – marcador de rastreamento apenas), maléolo lateral da fíbula, maléolo medial da tíbia, como ilustrado na Figura 17. Por último, para as marcações relativas aos pés, foi aplicada a marcação mais simples (*One Segment Feet*), visto que não era o escopo da pesquisa analisar movimentos finos dos pés. Os pontos anatômicos marcados foram: superfície posterior do calcâneo, cabeça do quinto metatarso e cabeça do segundo metatarso, Figura 18.

Figura 17 - Esquemático de posição de marcadores Shank 4



Fonte: Marker Set Guidelines (2022).

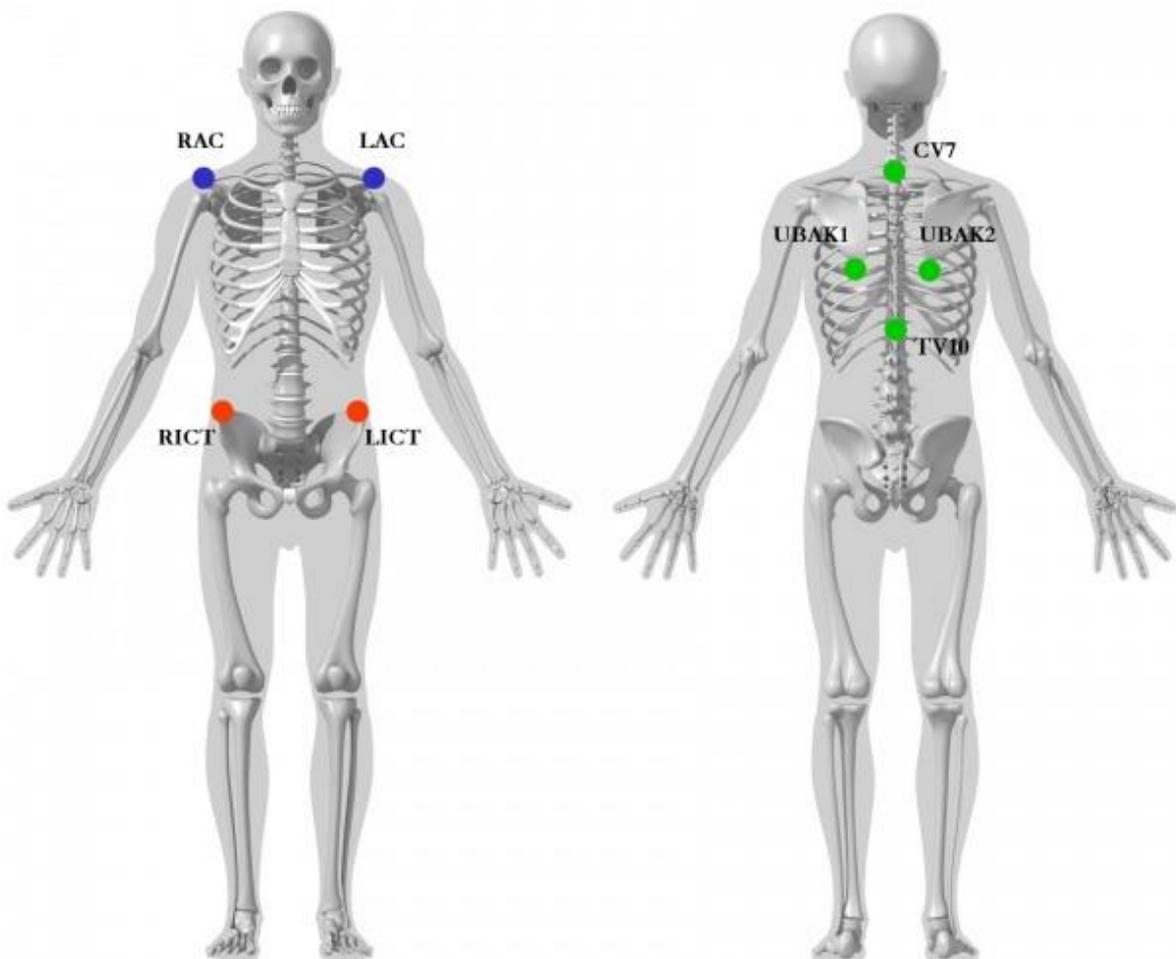
Figura 18 - Esquemático de posição de marcadores One Segment Feet



Fonte: Marker Set Guidelines (2022).

Para as gravações de membros superiores é necessário identificar com marcadores a estrutura do tórax também. Primeiramente, os segmentos do tórax foram marcados segundo as definições de *Thorax/Abdomen Model 2*, utilizando-se como referência os pontos anatômicos: sétima vértebra cervical, décima vértebra torácica, ápice do ângulo inferior da escápula direita e esquerda, acrômio direito e esquerdo e tubérculo da crista íliaca direita e esquerda, como ilustrado na Figura 19.

Figura 19 - Esquemático de posição de marcadores Thorax/Abdomen Model 2

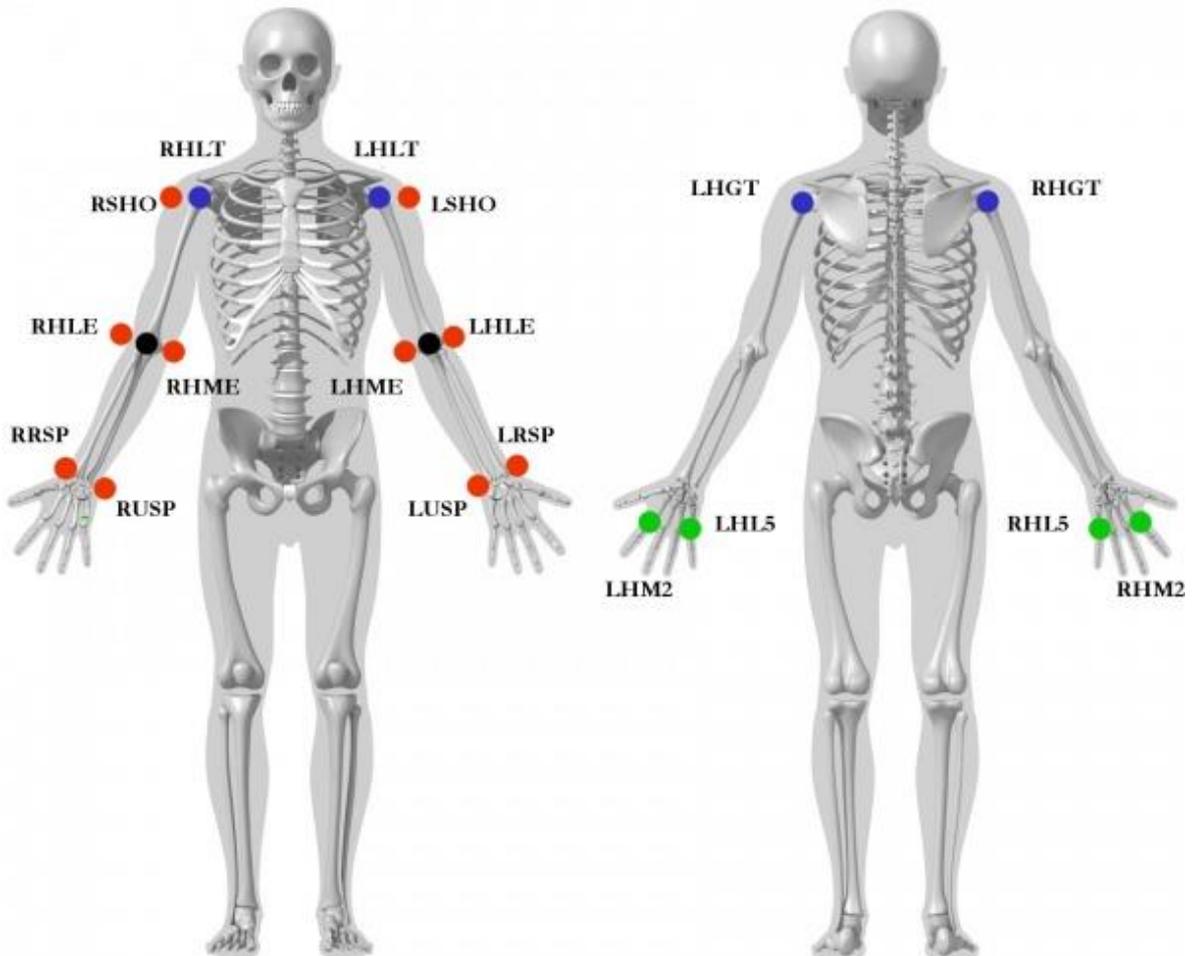


Fonte: Marker Set Guidelines (2022).

O braço foi marcado segundo o *Upper Arm and Lower Arm Segment Model 1*, sendo os pontos anatômicos referenciados: tubérculo maior do úmero, tubérculo menor do úmero, ombro (centro da porção lateral do deltoide), epicôndilo lateral do úmero, epicôndilo medial do úmero, processo estiloide do rádio, processo estiloide

da ulna, cabeça do segundo e quinto metacarpo, como foi demonstrado na Figura 20.

Figura 20 - Esquemático de posição de marcadores Upper Arm and Lower Arm Segment Model 1



Fonte: Marker Set Guidelines (2022).

Devido à quantidade limitada de marcadores com uma alta taxa e reflexão foi necessário gravar separadamente os movimentos de membro inferior e superior, pois a gravação de ambos os membros e bilaterais resultaria em um total de 50 marcadores.

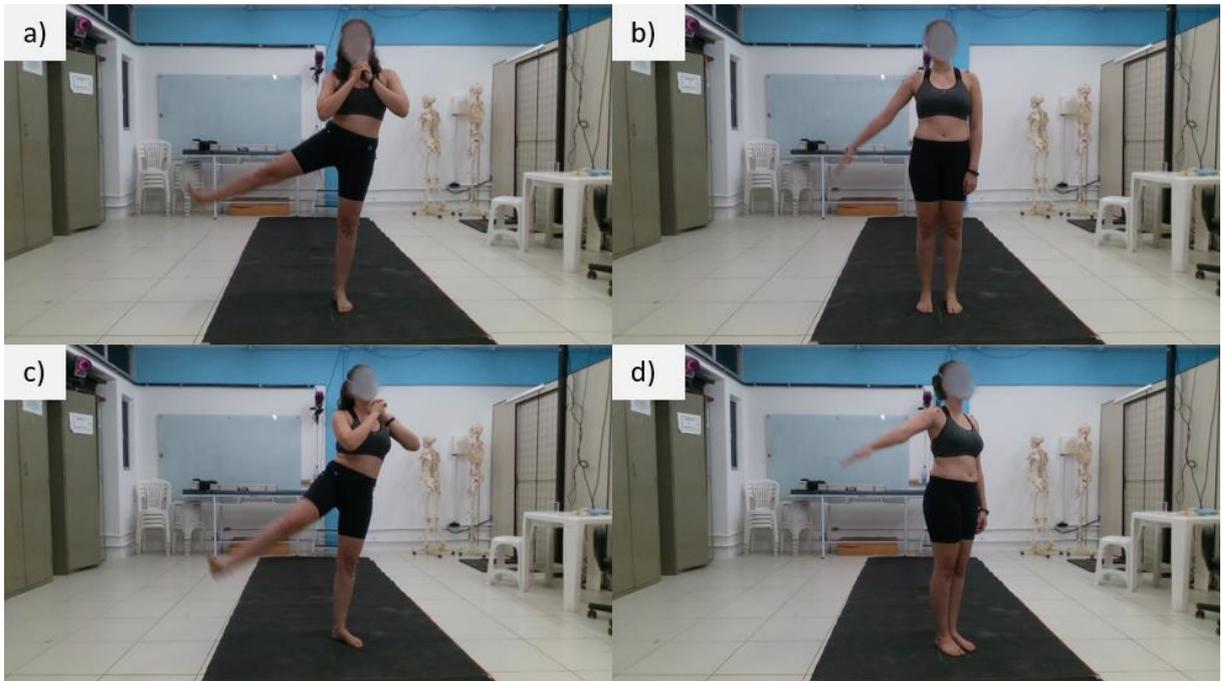
3.3 PROTOCOLO EXPERIMENTAL

As gravações ocorreram em momentos distintos para membros inferiores e superiores a fim de facilitar a identificação dos marcadores pós gravações e organização dos dados salvos. Foram realizados movimentos em duas articulações para cada membro realizando os seguintes movimentos monoplanares:

- Membro Inferior:
 - Joelho:
 - Flexão e extensão
 - Quadril:
 - Abdução e adução
 - Flexão e extensão
- Membro Superior:
 - Cotovelo:
 - Flexão e extensão
 - Ombro
 - Abdução e adução
 - Flexão e extensão

Para cada movimento foram realizadas duas gravações com diferentes inclinações em relação aos sensores. Nas primeiras gravações, o voluntário permanecia em frente ao equipamento e, posteriormente, repetiria os movimentos com uma rotação à esquerda de 30° em relação ao sensor, como demonstrado na Figura 21.

Figura 21 - Movimentos e posicionamento dos voluntários durante os exercícios. a) Movimento de MI de frente aos sensores. b) Movimento de MS de frente aos sensores. c) Movimento de MI inclinado aos sensores. d) Movimento de MS inclinado aos sensores.



Fonte: O Autor (2022).

Entretanto, o movimento de flexão de joelho foi realizado com o voluntário na perpendicular dos sensores de modo que o movimento ocorresse no plano frontal dos mesmos. Essa mudança específica foi realizada para viabilizar a captura do movimento, uma vez que se o voluntário estivesse de frente para os sensores ao longo do movimento a perna ficaria ocluída pela coxa, perdendo a referência de rastreamento das articulações. Os seguimentos móveis foram sempre os membros direito, como demonstrado na Figura 21. No total, foram realizadas 120 repetições por voluntário sendo analisados separadamente para cada sensor.

3.4 PROCESSAMENTO E TRATAMENTO DOS DADOS

É necessário estruturar as informações de modo que seja realizado um pré-processamento dos dados, visto que os dados brutos coletados são: informações de profundidade, posição espacial dos marcadores, imagens de cor e estimativas articulares. É necessário ordenar e agrupar as informações e iniciar os tratamentos para se obter a variação angular dos movimentos.

Nesse viés, o Kinect V2 fornece o rastreamento das articulações já incluído em seu SDK, não precisando de pré-processamento. Por outro lado, os sensores Astra e RealSense apresentam a possibilidade de rastreamento articular mediante a uma inscrição mensal (dificultando a utilização e aumentando os custos de pesquisa). Devido a isto, para realizar a identificação da posição das articulações foi utilizado o MediaPipe, e em sequência é realizada a sobreposição das imagens de profundidade e RGB para obtenção da coordenada 3D das articulações, como explicado no tópico a seguir, pois a partir das articulações 3D é possível determinar a variação angular dos segmentos corporais analisados.

3.4.1 Sobreposição da imagem RGB e profundidade

Devido à ausência do SDK gratuito para os sensores Astra Orbbec e Intel RealSense foi necessário utilizar um método híbrido para a inferência das articulações. Foi utilizado o *framework* do MediaPipe (BlazePose) para a detecção e rastreamento das articulações dos voluntários para os equipamentos Astra e RealSense. Desse modo, foi utilizado o BlazePose nas imagens RGB para determinar a posição articular e em seguida sobrepor a imagem de profundidade e reconstruir a coordenada 3D esperada.

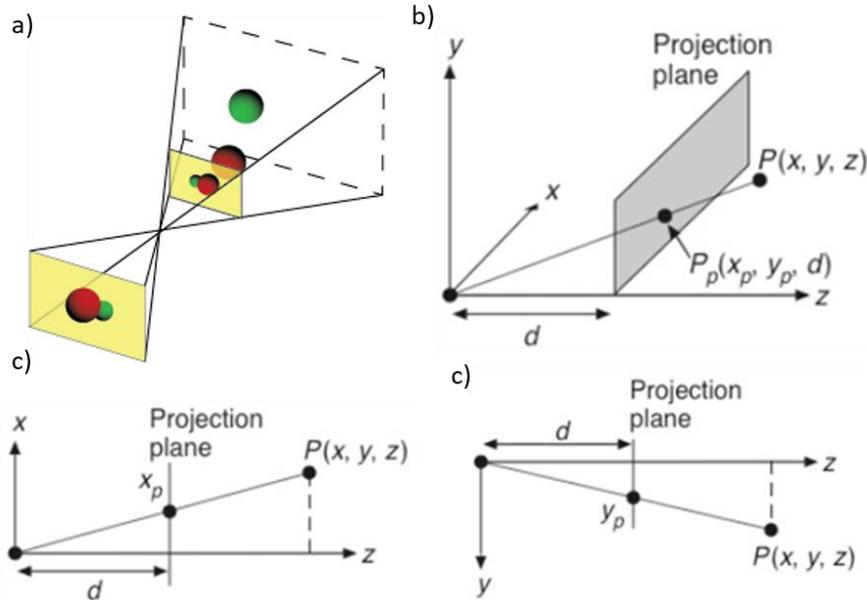
As informações da imagem RGB são coordenadas planares que devem ser associadas com a imagem de profundidade para se obter a representação 3D das articulações. Para isso, utilizou-se o método descrito por Knust et. al. (KUNST e colab., [S.d.]) de projeção e relações de semelhanças de distâncias.

Knust et. al. (KUNST e colab., [S.d.]) utilizaram semelhança de triângulos para correlacionar a medida planar com (x,y) e transformar em coordenadas de mundo 3D. Dadas as relações expressas em Equação 1 e 2 e na Figura 22.

$$\frac{x_p}{d} = \frac{x}{z} \quad (1)$$

$$\frac{y_p}{d} = \frac{y}{z} \quad (2)$$

Figura 22 - Representação do processo de semelhança angular. a) Projeção da frente da câmera. b) Representação sem considerar pontos principais. c) Vista do Eixo X. d) Vista do Eixo Y.



Fonte: Adaptado de Knust (2021).

Como o sensor oferece a informação de profundidade (z), as coordenadas de tela agora podem ser transformadas em coordenadas de mundo:

$$(x, y, z) = \left(\frac{x_p * z}{d}, \frac{y_p * z}{d}, z \right) \quad (3)$$

Reescrevendo obtém-se:

$$(x^w, y^w, z^w) = \left(\frac{x_p * z}{f}, \frac{y_p * z}{f}, z \right) \quad (4)$$

Considerando o deslocamento da origem de referência:

$$(x^w, y^w, z^w) = \left(\frac{(x_p - x_o) * z}{f}, \frac{(y_p - y_o) * z}{f}, z \right) \quad (5)$$

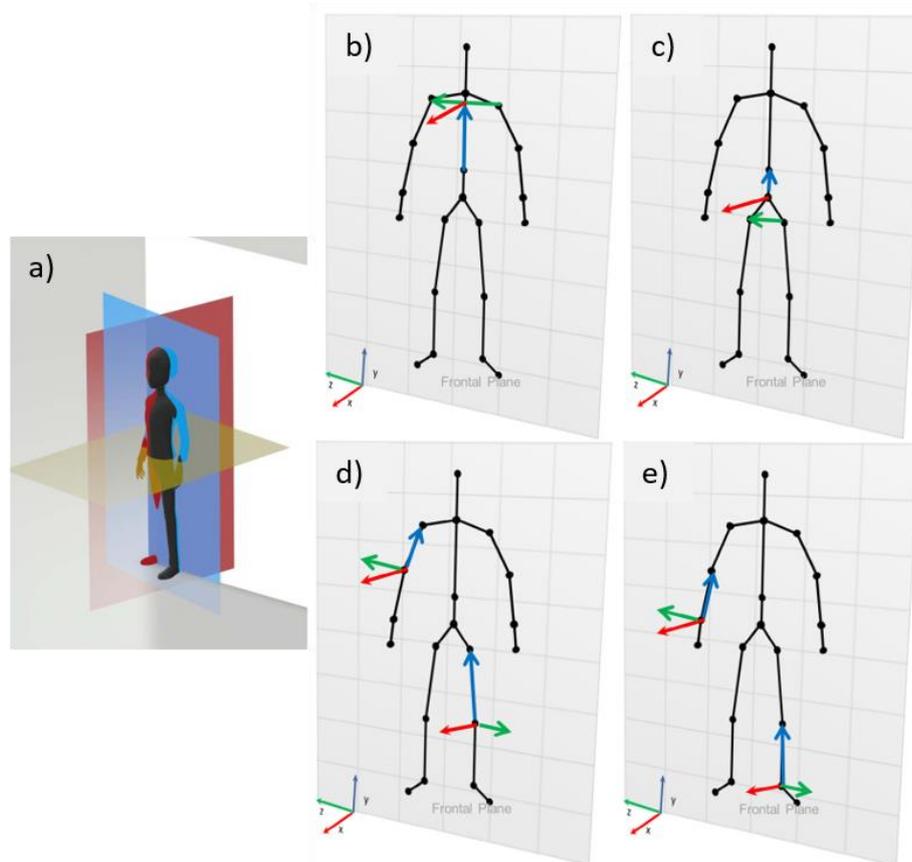
Para os cálculos de transformação 2D para 3D foi implementado uma biblioteca em C++, com *wrapper* para Python, empregando PyBind. Dessa forma foi utilizado C++ para o Astra, visto que é a linguagem padrão empregada na biblioteca da Orbbec. Para o RealSense foi utilizada sua biblioteca disponível em Python que continha os comandos básicos de operação e funcionamento do equipamento, embora não contivesse o SDK de inferência articular.

3.4.2 Cálculos angulares

Após a estimativa do posicionamento dos esqueletos, foi obtido um arquivo para cada movimento e cada sensor organizando as articulações. De modo geral, os arquivos contêm a seguinte organização de dados, $\{[\alpha]_i\}_j$, no qual α é um tupla (x, y, z) das coordenadas, i é o índice de cada articulação e j é o frame correspondente da gravação.

Para os cálculos angulares foi adotado as referências da ISB para o Sistema de Coordenadas Articulares, do inglês, *Joint Coordinate System (JCS)*, descrito por Da Gama (DA GAMA, Alana Elza Fontes e colab., 2016b), utilizando-se os três planos (Frontal, Sagital e Horizontal) anatômicos como padrão, como demonstrado na Figura 23. O JCS define o sistema de eixos coordenados individuais para cada articulação, sendo os eixos definidos como: \hat{x} eixo normal ao plano Frontal, \hat{y} eixo normal ao plano Horizontal e \hat{z} eixo normal ao plano Sagital.

Figura 23 - Representações das definições da ISB. a) Planos anatômicos. Referências utilizadas para sistema de coordenada: b) ombro e cervical; c) pélvis e lombar; d) cotovelo e joelho; e) punho e tornozelo.



Fonte: Adaptado de Burle (2020) e Da Gama (2016)

Dessa forma, para calcular o ângulo do movimento, três articulações precisam ser correlacionadas duas a duas, sendo a ângulo da articulação desejada equivalente à articulação central. Duas articulações sucessivas são conectadas para representar os segmentos do corpo, tendo como ponto de origem a articulação proximal e final a articulação distal.

$$\overrightarrow{Seg}_n = (x_d - x_p, y_d - y_p, z_d - z_p) \quad (6)$$

A partir dos esqueletos, foi utilizada a biblioteca BioAnalysis (KUNST e colab., [S.d.]), uma biblioteca escrita em Haxe, que utiliza o padrão ISB para calcular a angulação de uma determinada articulação ou vetor em relação a um dos planos do corpo (DA GAMA, A.E.F. e colab., 2014). Para cada *frame*, o esqueleto é passado para a biblioteca, além da articulação, do plano e da tolerância que serão analisados (valor máximo que o movimento poderá ocorrer fora do plano). Com isto, o BioAnalysis computa o ângulo entre a normal do plano e o vetor que está se movendo (caso tenha recebido uma articulação como entrada, o BioAnalysis converte para o vetor que representa a articulação).

Na prática, primeiro o esqueleto é convertido para coordenadas de corpo, i. e., as coordenadas são dadas em relação ao sistema de coordenadas de cada articulação. Para isto, para cada vetor realiza uma mudança de base em relação ao seu sistema de coordenadas, por exemplo: para membros superiores é utilizado o vetor do meio da coluna para o topo como eixo y, o produto vetorial do vetor criado a partir do quadril esquerdo e direito e o eixo y e o produto vetorial de x e y para formar o eixo z (o mesmo raciocínio, porém vetores diferentes são feitos para a clavícula, membros inferiores, tronco, cotovelos esquerdo e direito, pulsos esquerdo e direito, joelhos esquerdo e direito, tornozelos direito e esquerdo).

$$\begin{aligned} \vec{u} \times \vec{v} &= (a, b, c) \times (d, e, f) \\ \vec{u} \times \vec{v} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a & b & c \\ d & e & f \end{vmatrix} \\ \vec{u} \times \vec{v} &= (bf\vec{i} + cd\vec{j} + ae\vec{k} - bd\vec{k} - af\vec{j} - ce\vec{i}) \\ \vec{u} \times \vec{v} &= ((bf - ce)\vec{i} + (cd - af)\vec{j} + (ae - bd)\vec{k}) \\ \vec{u} \times \vec{v} &= ((bf - ce), (cd - af), (ae - bd)) \end{aligned} \quad (7)$$

Figura 24 - Código da criação dos vetores. var y definida como o vetor de origem na espinha do ombro e final na espinha média. Var x produto vetorial do vetor y e o vetor de origem no ombro esquerdo e final no ombro direito (vetor normal ao plano horizontal). Var z produto vetorial entre x e y

```
var y = IVector3.sub(esqueleto[Joins.SPINESHOULDER], esqueleto[Joins.SPINEMID]);
var x = IVector3.cross(IVector3.sub(esqueleto[Joins.SHOULDERLEFT], esqueleto[Joins.SHOULDERRIGHT]), y);
var z = IVector3.cross(x, y);
```

Fonte: O Autor (2022).

O sistema de coordenadas é normalizado e definido pelos vetores \vec{x} , \vec{y} e \vec{z} definidos acima. Com isso, a matriz de coordenadas do espaço é definida como:

$$[B]^{cc} = \begin{bmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & b_z \end{bmatrix} \quad (8)$$

Para a conversão do sistema de coordenadas dos sensores para um sistema de coordenadas para a articulação de análise foi construída a matriz de mudança de base. A expressão a seguir constitui a definição da operação de mudança de base:

$$[v]_{cc} = [I]_{cs}^{cc} [v]_{cs} \quad (9)$$

Na qual $[I]_{cs}^{cc}$ é a matriz de transformação de cs (coordenadas do sensor) para cc (coordenadas de corpo). Dessa forma tem-se:

$$[I]_{cs}^{cc} = \{[B]^{cc}\}^{-1} = \begin{bmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & b_z \end{bmatrix}^{-1} \quad (10)$$

Como os vetores em coordenadas de corpo, temos \vec{x} , \vec{y} e \vec{z} como os vetores (1,0,0), (0,1,0) e (0,0,1), respectivamente. Para calcular a tolerância, é analisado, primeiramente, o ângulo entre o vetor que está se movendo e a normal do plano em que o movimento está sendo realizado, para isso é utilizada a relação matemática do cos entre vetores:

$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| |\vec{v}|} \quad (11)$$

Após verificar se o vetor pertence ao plano, i.e., o vetor que está se movendo deve ter um ângulo de 90° em relação ao vetor normal do plano. Subsequentemente, considerando a tolerância dada, é realizado o cálculo do ângulo entre o vetor que está se movendo com um dos vetores perpendiculares que pertencem ao plano, por exemplo, para o plano frontal, será realizado o cálculo entre o vetor que está se movendo com o eixo Y. Caso o vetor não pertença ao plano, é retornado um erro.

3.4.3 Tratamento dos Dados

Após a computação dos ângulos, os sinais dos movimentos precisam ser suavizados e sincronizados. Com intuito de facilitar e tornar mais acessível, foi construído um *script* em Python utilizando as bibliotecas NumPy (“NumPy”, [S.d.]), SciPy (“SciPy”, [S.d.]) e Matplotlib (“Matplotlib — Visualization with Python”, [S.d.]) por serem gratuitas e permitirem facilmente ajustes e a possibilidade de compartilhar o algoritmo.

No fluxo de tratamento dos dados, primeiramente é realizada uma filtragem com o filtro de Butterworth passa-baixa de sétima ordem com frequência de corte de 5Hz e 30Hz para os sinais oriundos dos sensores RGB-D e Qualisys, respectivamente, para retirar os ruídos/interferência durante a captura, sendo sua forma genérica expressa abaixo:

$$H_{j\omega} = \frac{1}{\sqrt{1 + \varepsilon^2 \left(\frac{\omega}{\omega_p}\right)^{2n}}} \quad (12)$$

Em seguida é aplicado um filtro de média móvel para suavizar o formato da curva após o filtro Butterworth. O filtro de média móvel aplicado possuía um *kernel* de tamanho 7.

$$\bar{x}_n = \frac{1}{N} \sum_{k=0}^{N-1} x_{n-k} \quad (13)$$

Com os sinais devidamente filtrados e melhor suavizados é necessário torná-los de mesmo tamanho, visto que os sinais provenientes do Qualisys apresentam mais de 6 vezes a frequência de captura dos sensores RGB-D (200Hz / 30Hz). O processo de interpolação foi realizado simultaneamente ao processo de sincronização. Os sinais foram divididos em três partes:

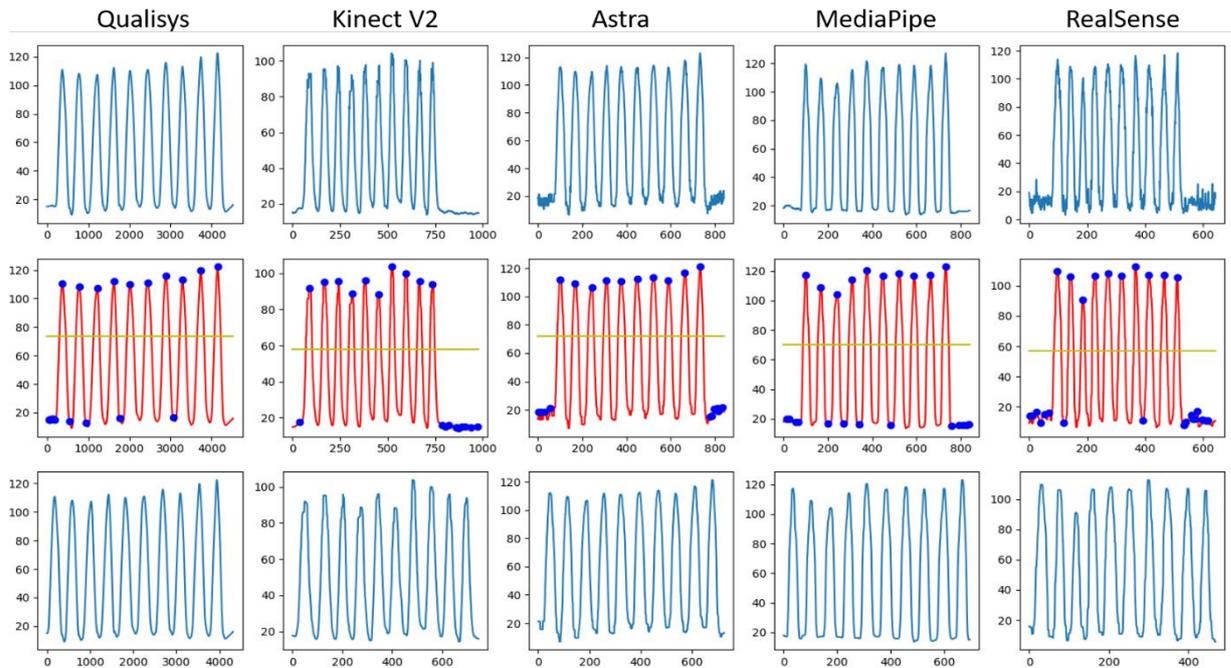
- Início ao primeiro pico
- Primeiro pico ao último pico
- Último pico

Os picos dos sinais representam o mesmo instante de tempo, o qual é o valor máximo de deslocamento angular do movimento em análise. Primeiramente foi identificada a distância entre a quantidade de *frames* entre as amostras em relação à referência de interpolação. Tendo essas secções e a quantidade de pontos necessários a ser interpolado, utilizou-se uma função de interpolação polinomial de segunda ordem. Desse modo, após reestruturar os dados, todos os sinais apresentavam o mesmo tamanho e estavam em fase.

O primeiro processo foi realizar a identificação dos picos para identificar o início dos movimentos. A Figura 25 exibe a mudança de cada dado ao longo do processo. A primeira linha contém os dados puros sem nenhum tipo de tratamento, ao passo que a segunda linha representa os dados após a filtragem com os filtros de média móvel e Butterworth, suavizando as irregularidades e facilitando a identificação de picos (máximos locais).

A partir da identificação dos picos que estavam acima do valor médio da amostra, foi possível determinar quais significavam a execução propriamente e não interferência, sendo o resultado visualizado na segunda linha de gráficos da Figura 25. Por fim, os dados foram restritos apenas ao intervalo de tempo que compreendia ao movimento, resultando em um conjunto limpo de informações do movimento (terceira linha da Figura 25).

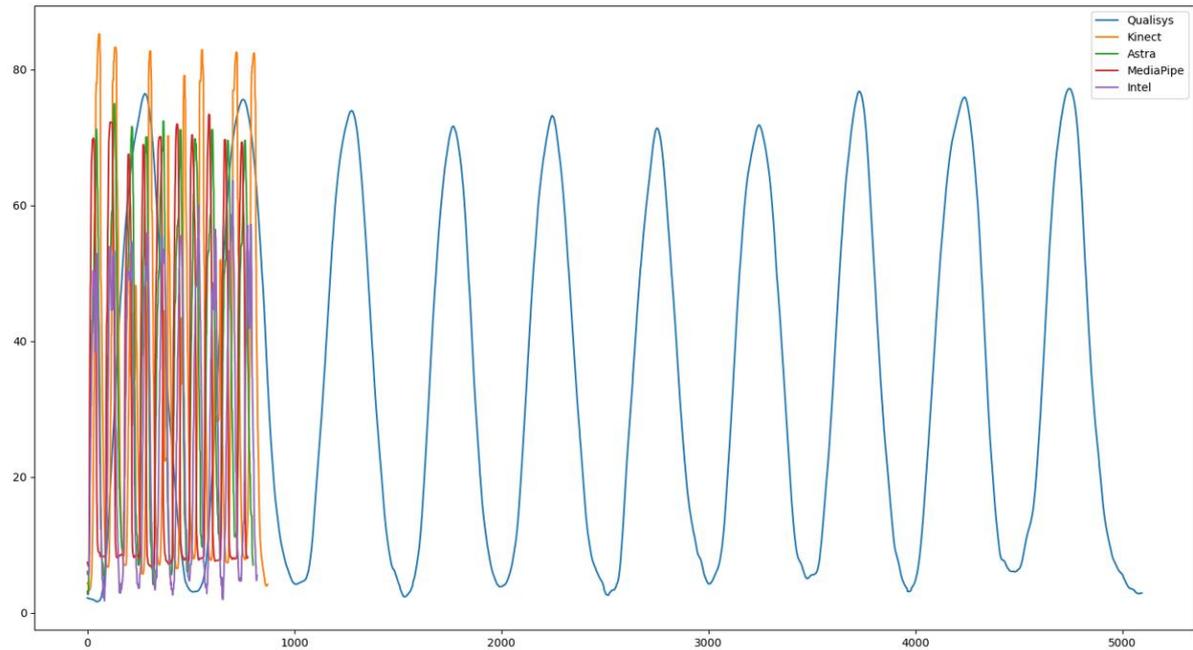
Figura 25 - Gráficos do processo de tratamento dos dados.



Fonte: O Autor (2022).

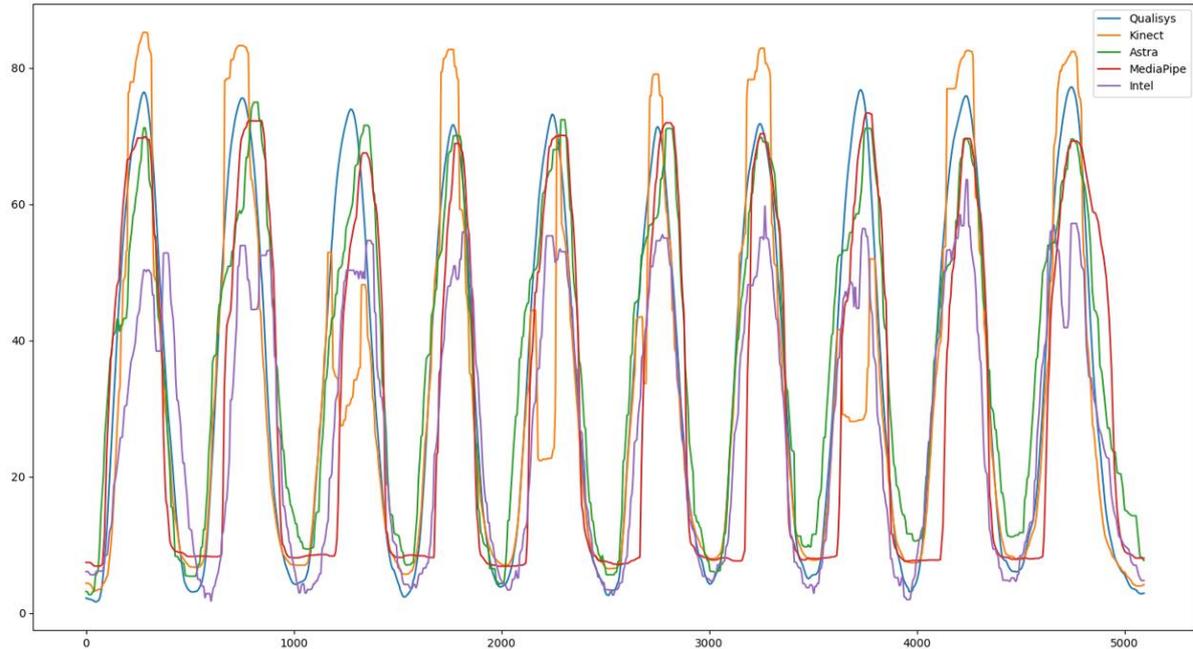
O procedimento de sincronização e interpolação foi realizado simultaneamente como descrito na seção Tratamento dos Dados. A Figura 26 exibe a diferença de tamanho entre as amostras resultado da discrepância entre as frequências amostrais dos equipamentos. Após a interpolação e a sincronização, todos os dados apresentaram a mesma fase (como demonstra a Figura 27). Ambas as imagens abaixo representam o mesmo conjunto de dados. É importante salientar que, mesmo com os filtros, as curvas dos equipamentos RGB-D não apresentaram a mesma suavidade que as do Qualisys.

Figura 26 - Dados não sincronizados e interpolados.



Fonte: O Autor (2022).

Figura 27 - Dados sincronizados e interpolados.



Fonte: O Autor (2022).

Um código alfabético foi determinado para simplificar o registro e identificação dos movimentos, seguindo a regra: Articulação – movimento- posição em relação ao sensor- índice da gravação, por exemplo J-FLEX-FR-1 representa joelho-flexão-

frontal-gravação 1. Dessa forma, as combinações de códigos de identificação formados foi:

- J-FLEX-FR
- J-FLEX-INL
- Q-ABD-FR
- Q-ABD-INL
- Q-FLEX-FR
- Q-FLEX-INL
- C-FLEX-FR
- C-FLEX-INL
- O-ABD-FR
- O-ABD-INL
- O-FLEX-FR
- O-FLEX-INL

3.4.4 Variáveis de validação e estatística

Com os dados devidamente interpolados e sincronizados foi possível realizar os testes de comparação. Como o objetivo desse trabalho foi avaliar o desempenho de diferentes sensores em frente ao padrão ouro, calculou-se o erro relativo e erro absoluto (E_A) dos sensores e sua média quadrática (RMS) para se entender a variação do erro e dispersão das amostras.

$$e_r = \theta_s - \theta_Q \quad (14)$$

$$E_A = |\theta_s - \theta_Q| \quad (15)$$

$$x_{RMS} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \dots + x_n^2)} \quad (16)$$

Em adicional, foi utilizada a Correlação de Pearson entre os valores angulares, pois a correlação indica a proximidade relativa entre duas variáveis, a

força de sua relação linear. Ademais, o coeficiente de correlação de concordância indica o acordo absoluto entre as variáveis, medindo não apenas a distância que cada ponto se afasta da linha que se ajusta aos dados (precisão), mas também a distância que esta linha se afasta da linha de 45 graus da origem (acurácia)(VILAS-BOAS, Maria Do Carmo e ROCHA e colab., 2019).

Para finalizar a análise, determinou o comportamento dos dados em vista a distribuição normal. Foi concluído pelo teste paramétricos que os dados não seguem uma distribuição normal, dessa forma, os dados são dependentes não paramétricos. Por isso, foi utilizado o teste de significância das diferenças (teste Wilcoxon não paramétrico pareado) entre cada par de sensores, onde dois níveis de significância foram definidos e relatados como (*) para $p < 0,05$ e (**) para $p < 0,001$ (ALBERT e colab., 2020). Outra análise post hoc utilizada foi a de rank-sum de Friedman para comparação múltipla dos sensores, com os mesmos níveis para $p < 0,05$ e (**) para $p < 0,001$. Utilizando a escala de correlação de Pearson estabelecida por Portney e Watkins (PORTNEY e WATKINS, 2015), tem-se que os limites de correlação serão classificados como: Pobre ($<0,5$); Moderado ($\geq 0,5$ e $<0,75$); Bom ($\geq 0,75$ e $<0,9$) e Excelente ($\geq 0,9$), utilizando a mesma sinalização para p.

Foi avaliado o erro absoluto, RMS e correlação de Pearson entre o padrão ouro e os sensores para o conjunto total de dados, bem como para cada movimento analisado. Para os cálculos estatísticos foi utilizado o *software* R, por ser de uso aberto e gratuito. Foi escolhido o gráfico de caixa e de barras para melhor representação e visualização dos valores comparativos.

3.5 VIRTUAL KINECT

O Virtual Kinect (VK) é uma solução que simula o comportamento do Kinect V2, permitindo ao usuário testar e desenvolver aplicações de rastreamento corporal sem ter um sensor físico Kinect V2 disponível (uma câmera RGB comum é usada em seu lugar). Além disso, permite ao programador mudar do dispositivo Kinect real para o VK mudando apenas os arquivos DLL usados, mantendo o código fonte intacto. Nesta primeira versão do VK, duas características do Kinect V2 estão disponíveis: acesso aos dados fornecidos pela câmera RGB e coordenadas com articulação de cada quadro. Entretanto, para que o VK funcione da mesma forma

que o Kinect V2, as capacidades de segmentação do corpo e as informações de profundidade da imagem ainda têm que ser adicionadas ao projeto.

Devido à diferença entre os esqueletos do Kinect V2 e do MediaPipe, como o posicionamento das articulações e o número de articulações, foi realizada uma correlação matemática das articulações do MediaPipe para se adequar ao esqueleto do Kinect. As equivalências das articulações são mostradas na Tabela 4.

Tabela 4 - Correlação entre as articulações do Kinect e do MediaPipe.

Nome da Articulação	Índice do Kinect	Índice do Virtual Kinect
Base da Coluna Vertebral	0	$(24+23)/2$
Coluna Vertebral	1	$(11+12+23+24)/4$
Pescoço	2	$(9+10+11+12)/4$
Cabeça	3	0
Ombro esquerdo	4	11
Cotovelo esquerdo	5	13
Pulso esquerdo	6	15
Mão esquerda	7	$(19+17+15)/3$
Ombro Direito	8	12
Cotovelo Direito	9	14
Pulso Direito	10	16
Mão Direita	11	$(16+18+20)/3$
Quadril esquerdo	12	23
Joelho Esquerdo	13	25
Tornozelo Esquerdo	14	27
Pé Esquerdo	15	31
Quadril Direito	16	24
Joelho Direito	17	26
Tornozelo Direito	18	28
Pé Direito	19	32
Coluna Vertebral C7	20	$(11+12)/2$
Ponta da mão esquerda	21	$(17+19)/2$
Polegar Esquerdo	22	21
Ponta da Mão Direita	23	$(18+20)/2$
Polegar Direito	24	22

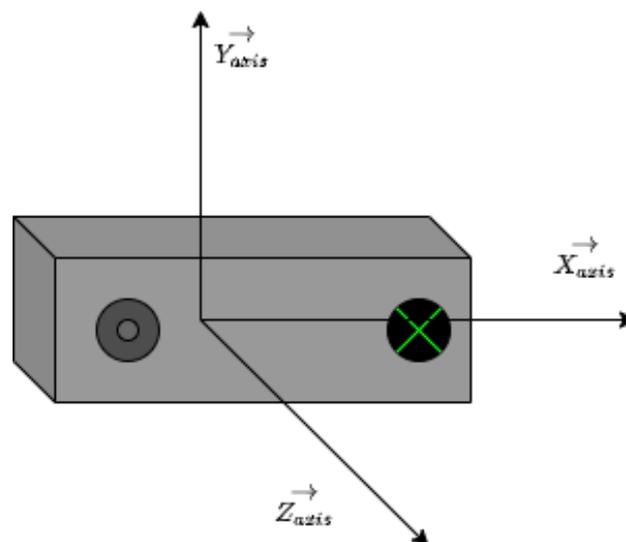
Fonte: O Autor (2022).

Como o esqueleto do MediaPipe tem mais oito articulações do que o do Kinect, nem todas as articulações têm uma equivalência direta, portanto, o calcanhar e a maioria das articulações do rosto não foram utilizados. Considerando que

algumas articulações estão presentes em ambos os esqueletos, apenas utilizamos as coordenadas fornecidas pelo MediaPipe diretamente na junta correspondente no esqueleto do Kinect V2. Entretanto, existem articulações no esqueleto do Kinect que o MediaPipe não representa, de modo que estas articulações têm suas coordenadas calculadas com base em duas ou mais articulações pela média de suas posições. A coordenada da base da coluna vertebral, por exemplo, é obtida calculando a média entre as coordenadas do quadril esquerdo e direito a partir do MediaPipe.

Apesar de Kinect e MediaPipe fornecerem informações de coordenadas de cada junta, há uma diferença em como os valores do eixo Z são calculados. O sistema cartesiano adotado pelo Kinect tem sua origem localizada no centro do Kinect, como mostrado na Figura 28, o que significa que a coordenada Z não pode assumir valores negativos e, à medida que um objeto se move mais, seus valores do eixo Z aumentam (em metros). Por outro lado, o MediaPipe tem duas formas diferentes de representar as coordenadas do eixo Z, ambas com a profundidade no ponto médio entre os quadris sendo a origem. A magnitude de Z usa aproximadamente a mesma escala que x e as coordenadas da origem são normalizadas para [0,0, 1,0] pela largura e altura da imagem, respectivamente. A outra forma é utilizando coordenadas 3D do mundo real, nas quais os valores de uma posição em cada eixo são a distância em metros da origem do cartesiano localizado entre os quadris.

Figura 28 - Sistema cartesiano Kinect V2 utilizado.



Fonte: O Autor (2022).

MediaPipe tem o parâmetro de visibilidade, com valores que vão de 0,0 até 1,0, indicando a probabilidade de que o marco seja visível (presente e não ocluído na imagem). Kinect tem um parâmetro análogo chamado Tracking State que pode assumir três valores diferentes: 0 (Não Rastreado), 1 (Inferido) e 2 (Rastreado). De acordo com a documentação do MediaPipe, é possível executar a maioria de suas funcionalidades diretamente no Windows, através do Python. Como o Kinect SDK é implementado em C++, precisávamos de uma forma de realizar a comunicação entre Python e C++ que pudesse ser suficientemente eficiente para suportar a transferência em tempo real das informações dos quadros RGB, adquiridas pela webcam no lado Python do MediaPipe, juntamente com as coordenadas conjuntas calculadas por ele. A solução encontrada foi o uso de memória compartilhada entre as duas aplicações. Na parte em Python, usamos o pacote mmap para comunicação Inter processada. O código parcial mostrado na Figura 29 descreve como é criado o espaço de memória compartilhada (linha 3).

Figura 29 - Uso de memória compartilhada na parte em Python.

```
1 import mmap
2
3 shm = mmap.mmap(0, 640*480*3 + 33*4*4, "VKInfo")
4
5 while cap.isOpened():
6     success, image = cap.read()
7
8     shm.seek(0)
9     shm.write(bytes(image));
10
11     results = pose.process(image)
12
13     ... #pack the joint results into the joint_data byte
14         array
15     shm.write(joint_data)
```

Fonte: O Autor (2022).

No momento, estamos reservando espaço para armazenar uma imagem 640x480 RGB juntamente com informações (4 números de ponto flutuante) para cada uma das 33 juntas fornecidas pelo MediaPipe. Num futuro próximo, esperamos suportar diferentes resoluções de webcam juntamente com mais recursos (como

segmentação do player e informações de profundidade da imagem). Em cada frame lido da webcam (linha 6) devemos definir o ponteiro da memória compartilhada para o início da região mapeada (linha 8) e escrever tanto a imagem (linha 9) quanto as informações das coordenadas das articulações (linha 15) para a memória compartilhada. A cadeia VKInfo (linha 3) é o identificador que será usado no lado C++ para localizar a compartilhada região de memória.

Na parte executada em C++, mostrado na Figura 30, sempre que o sensor Kinect Virtual é aberto, a função `createFileMapping` é chamada (linha 7) para obter uma conexão para o arquivo de mapa e o `MapViewOfFile` é chamado (linha 15) em sequência para fornecer o ponteiro para o início da região de memória compartilhada. Como armazenamos as informações da imagem no início da memória compartilhada, o `pBuf` já aponta para o endereço correto. Para que as juntas apontem para o endereço adequado, devemos apontar para uma posição imediatamente após os bytes de imagem $640 \times 480 \times 3$ (linha 22). Este código só é necessário para executar uma única vez. Depois disso, ambas as informações de imagem e de juntas podem ser acessadas diretamente dos ponteiros `pBuf` e `juntas`, respectivamente.

Figura 30 - Uso de memória compartilhada no código em C++.

```

1 #define BUF_SIZE 640*480*3+33*4*4
2 TCHAR szName[] = TEXT("VKInfo");
3
4 HRESULT STDMETHODCALLTYPE VKKinectSensor::Open(void) {
5     HANDLE hMapFile;
6
7     hMapFile = CreateFileMapping(
8         INVALID_HANDLE_VALUE,    // use paging file
9         NULL,                    // default security
10        PAGE_READWRITE,         // read/write access
11        0,                       // maximum object size
12        BUF_SIZE,               // maximum object size
13        szName);                // name of mapping object
14
15    pBuf = (char*)MapViewOfFile(
16        hMapFile,                // handle to map object
17        FILE_MAP_ALL_ACCESS,     // read/write permission
18        0,
19        0,
20        BUF_SIZE);
21
22    joints = (float*)&pBuf[640 * 480 * 3];
23
24    return 0;
25 }

```

Fonte: O Autor (2022).

A fim de tornar o uso da VK o mais transparente possível para o usuário final (isto é, somente mudando as DLLs), tivemos que reimplementar a biblioteca do Kinect SDK. É importante notar que somente as funções relacionadas ao acesso ao RGB e informações das articulações foram implementadas em neste momento. Com a evolução do MediaPipe, novas funcionalidades serão acrescentadas. Isto significa que tivemos que criar as seguintes classes C++ para permitirem o acesso aos dados do MediaPipe, de acordo com as interfaces especificadas no SDK original do Kinect (mostrado na Tabela 5). Como este é um projeto de código aberto, a implementação completa está disponível em <https://github.com/voxarlabs/Virtual-Kinect>.

Tabela 5 - Classes do Virtual Kinect implementadas com base nas interfaces originais definidas pelo Kinect SDK.

Nome da Classe	Interface Original do SDK
VKKinectSensor	IKinectSensor
VKColorFrame	IColorFrame
VKColorFrameSource	IColorFrameSource
VKColorFrameReader	IColorFrameReader
VKBody	IBody
VKBodyFrame	IBodyFrame
VKBodyFrameReader	IBodyFrameReader
VKBodyFrameSource	IBodyFrameSource

Fonte: O Autor (2022).

4 RESULTADOS

Esta seção contém os resultados obtidos a partir dos processos descritos para tratamento dos dados. Foi realizada uma avaliação quantitativa e estatística das informações angulares, visando determinar a precisão dos equipamentos em frente ao padrão ouro. Para melhor visualização e entendimento dos resultados essa seção foi dividida em: perda de dados, resultados gerais, membros inferiores e membros superiores.

4.1 PERDA DE DADOS

Dos movimentos listados para as gravações, alguns não puderam ser comparados devido à inviabilidade de extração de articulações no Visual 3D e/ou consistências nas gravações dos sensores RGB-D. Isso ocorreu devido à perda de marcadores de definições articulares ao longo da gravação, de modo que não fosse possível interpolar de modo viável o deslocamento desses marcadores. Dessa forma, não foi possível definir os segmentos no Visual 3D para determinar a posição das articulações. Perdas dos sensores ocorreram devido à geração de dados absolutamente ruidosos ou com um *delay* de gravação das informações de profundidade assíncronas das imagens de profundidade. Quando a perda vinha de um sensor as outras gravações não eram avaliadas para garantir sempre a comparação dentre todos os sensores. A Tabela 6 expõe as gravações que foram e não foram analisadas, as gravações marcadas com “x” não foram avaliadas para determinado voluntário.

Tabela 6 - Tabela dos movimentos gravados, explicitando quais não foram analisados, identificados com "o" os positivos e "x" os negativos.

	Voluntário 1	Voluntário 2	Voluntário 3	Voluntário 4	Voluntário 5
J-FLEX-FR-1	o	x	o	o	o
J-FLEX-FR-2	o	x	o	o	o
J-FLEX-INL-1	o	o	o	o	x
J-FLEX-INL-2	o	o	o	o	o
Q-ABD-FR-1	o	o	x	o	o
Q-ABD-FR-2	o	o	o	o	o
Q-ABD-INL-1	o	x	o	o	x
Q-ABD-INL-2	o	x	o	o	o
Q-FLEX-FR-1	o	o	o	o	o
Q-FLEX-FR-2	o	o	x	o	o
Q-FLEX-INL-1	x	o	o	o	o
Q-FLEX-INL-2	x	o	o	x	o
C-FLEX-FR-1	o	o	o	o	x
C-FLEX-FR-2	o	o	o	o	o
C-FLEX-INL-1	o	o	o	x	o
C-FLEX-INL-2	x	o	o	o	x
O-ABD-FR-1	x	o	o	x	o
O-ABD-FR-2	o	x	o	o	o
O-ABD-INL-1	o	o	o	o	o
O-ABD-INL-2	x	o	o	o	o
O-FLEX-FR-1	o	o	o	o	o
O-FLEX-FR-2	o	o	o	o	o
O-FLEX-INL-1	o	o	o	o	o
O-FLEX-INL-2	o	o	o	o	o

Fonte: O Autor (2022).

Dos dados perdidos a fonte da perda foi variada, entretando foi evidenciado o sensor o qual impossibilitou a gravação para cada voluntário:

- Voluntário 1:
 - Kinect: Q-FLEX-INL-1 e 2 / O-ABD-FR-1
 - Qualisys: C-FLEX-INL-2/ O-ABD-INL-2
- Voluntário 2:
 - Astra: O-ABD-INL-1 e 2
 - Intel: O-ABD-FR-2
 - Kinect: J-FLEX-FR-1 e 2
- Voluntário 3:
 - Astra: Q-ABD-FR-1
 - Qualisys: Q-FLEX-FR-2

- Voluntário 4:
 - Intel: Q-FLEX-INL-2 / C-FLEX-INL-1
 - Qualisys: Q-FLEX-INL-2
- Voluntário 5:
 - Astra: C-FLEX-FR-1
 - Intel: C-FLEX-INL-2
 - Qualisys: J-FLEX-INL-1 / Q-ABD-INL-1

Das 120 gravações realizadas 19 não puderam ser avaliadas, uma perda de 15,83%. Os motivos para a perda pelo Qualisys foram a perda de marcadores de identificação articular ao longo da gravação. Para o Astra foi a perda de identificação das articulações ao longo do movimento gerando saltos de posição das articulações. As perdas provenientes do Intel RealSense ocorreu por causa de problema de buffer durante a coleta e a frequência das gravações ficaram inconstantes com perda da sincronia entre as imagens. Por outro lado, as perdas provenientes do Kinect foram resultados da perda de frames e do corrompimento de uma gravação, sem motivo determinado.

Ademais, as 101 gravações foram analisadas, totalizando 505 amostras contendo 10 repetições de movimento, resultando 5050 execuções. Desse modo primeiramente será exposto os dados gerais seguidos das avaliações por membros.

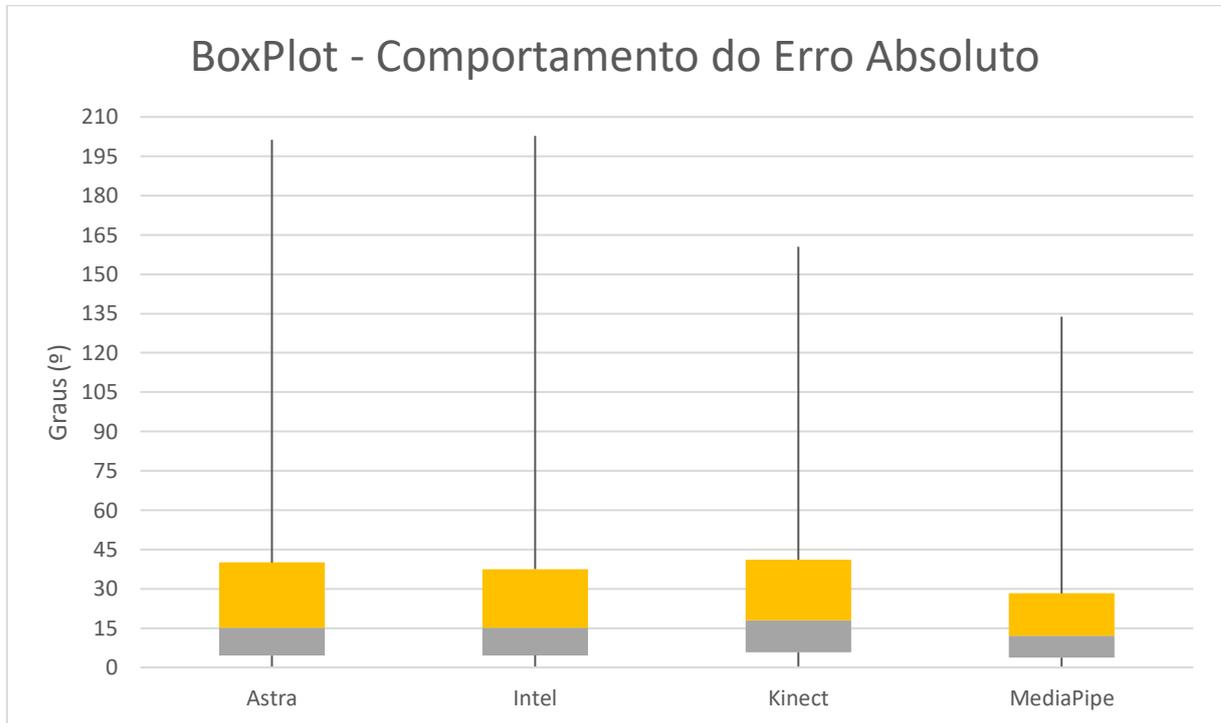
4.2 DADOS GERAIS

Os dados coletados não apresentavam comportamento de distribuição normal para nenhum dos dados coletados e calculados (variação angular e erro absoluto) pois apresentaram $p\text{-value} < 0.001$ no teste de normalidade. Sendo os dados não paramétricos foi conduzido um teste de Wilcoxon par a par entre o Qualisys com cada sensor, $p\text{-value}$ foi menor que $p < 0,001$, além disso foi realizado outro teste post-hoc, o teste de rank-sum de Friedman considerando todos os sensores para a comparação resultando em $p < 0,001$. Dessa forma percebe-se que há a diferença estatística dos dados coletados para cada sensor.

Das variáveis analisadas a primeira foi o Erro Absoluto da variação angular dos sensores frente ao Qualisys. Na Figura 31 é possível perceber a distribuição do erro. Dentre os sensores, MediaPipe apresentou a menor mediana (8,3°). O Kinect

apresentou o valor mais elevado de mediana, 12,2°. Astra e Intel apresentaram valores máximos e picos de discrepância maiores.

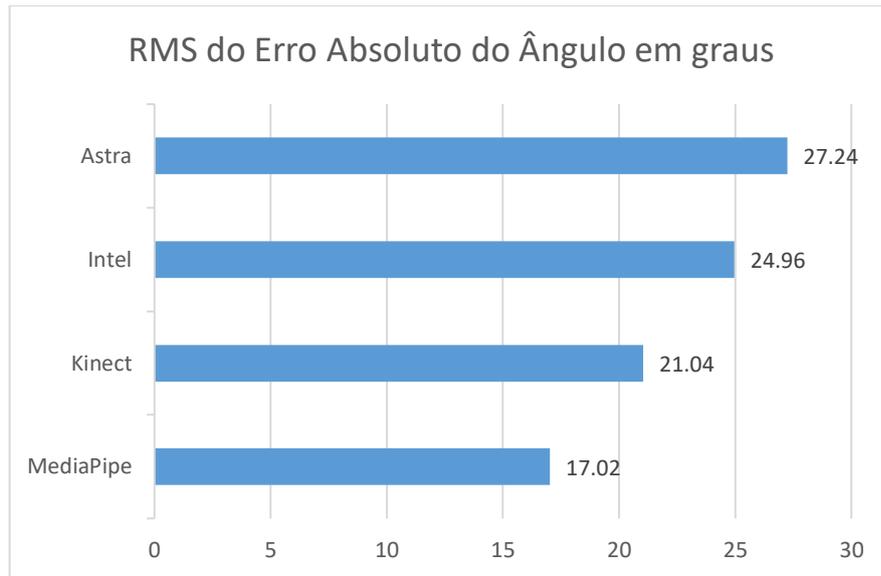
Figura 31 - Box Plox do Erro Absoluto dos dados gerais. Significância $p < 0,001$ para todos os sensores segundo o teste de rank-sum de Friedman.



Fonte: O Autor (2022).

Avaliando a Figura - 32 é possível entender que o comportamento do MediaPipe e do Kinect apresentaram valores menos dispersos. Essa dispersão pode ser avaliada pelos valores de RMS do Erro Absoluto. Diante disso é possível notar que a dispersão dos valores de erro para os sensores Kinect e MediaPipe obtiveram os menores valores.

Figura - 32 Gráfico de RMS do Erro Absoluto comparando todos em sensores quanto a dispersão do erro.



Fonte: O Autor (2022).

O último elemento avaliado foi a correlação entre os sensores e o Qualisys pelo método de Pearson. No teste de Matriz de correlação de Pearson todos os sensores apresentaram um $p < 0,001$. A *Tabela 7* exibe os valores da correlação de Pearson para os sensores, MediaPipe e Kinect apresentaram valores de Bom de correlação, enquanto Astra e Intel valores Moderados de correlação.

Tabela 7 - Correlação de Pearson para os dados de variação angular. Para todos os sensores $p < 0,001$ Azul ($>0,75$ e $<0,9$) e Amarelo ($>0,5$ e $<0,75$).

Astra	Intel	Kinect	MediaPipe
0,69	0,73	0,81	0,86

Fonte: O Autor (2022).

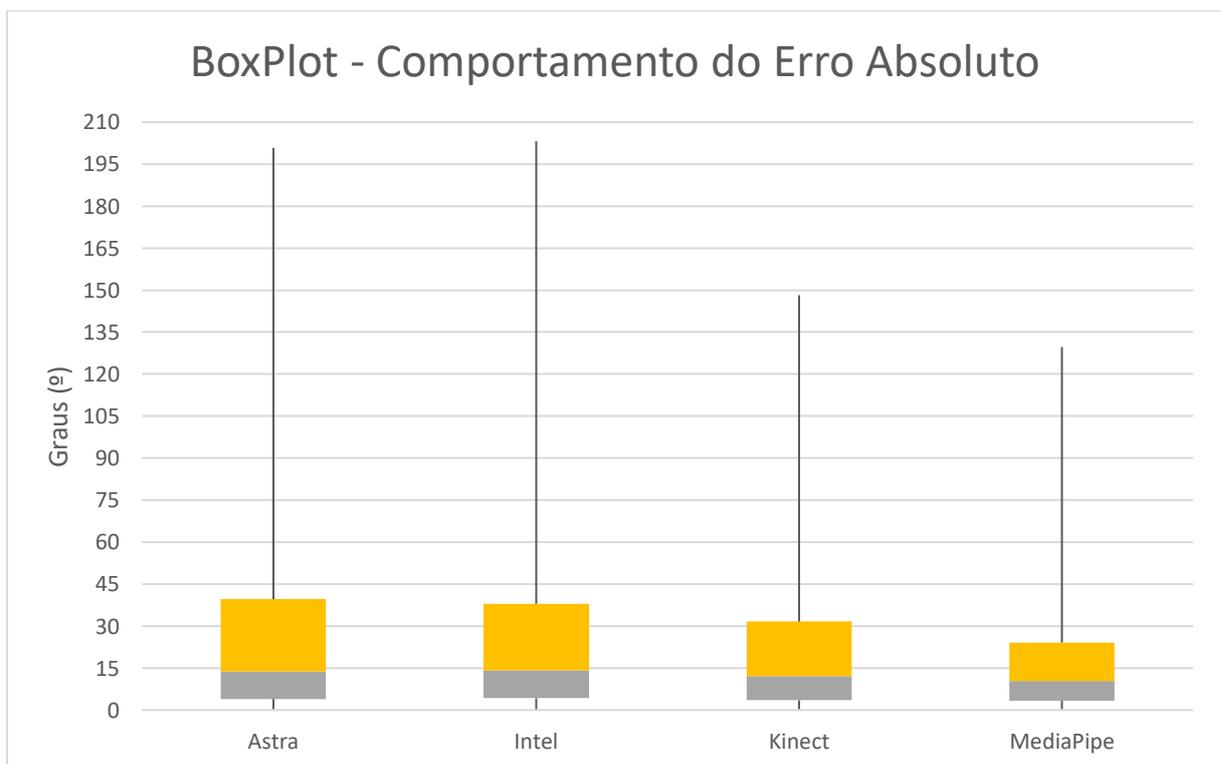
4.3 MEMBROS INFERIORES

Para a avaliação separada de membros inferiores foram realizadas duas formas de análise, primeiramente considerando todos os dados de MMII e posteriormente cada movimento e avaliando a média.

4.3.1 Análise Geral

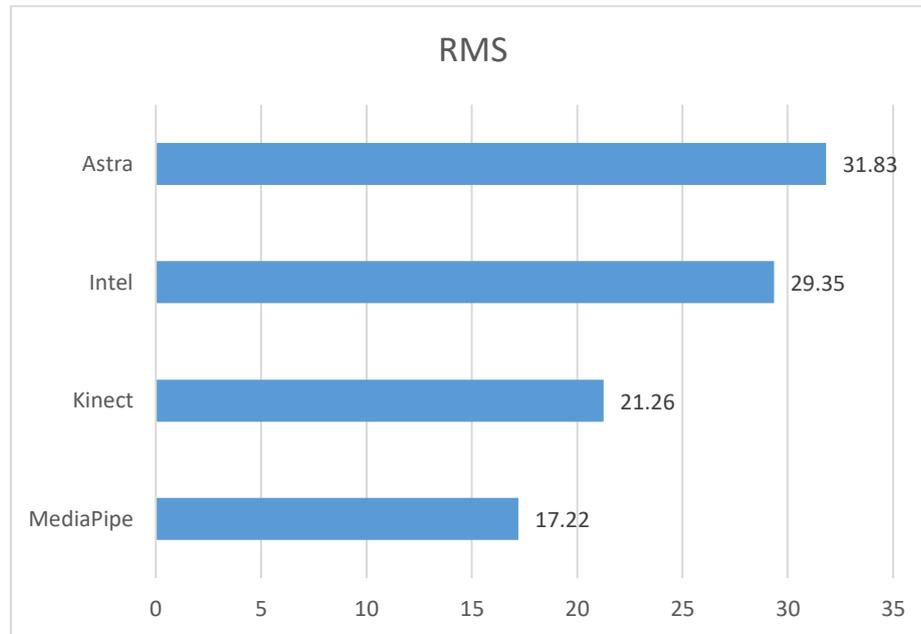
A Figura 33 expressa o BoxPlot para o erro absoluto das amostras em comparação pareada lado a lado dos sensores. Os resultados foram muito próximos, em comportamento, das comparações gerais do subtópico anterior. MediaPipe apresentou a menor mediana do Erro Absoluto, valor de 7,01°. Enquanto o Kinect obteve a segunda menor mediana, 8,56°. Por outro lado, Astra e Intel performaram com os maiores valores de mediana para o Erro, respectivamente, 9,86° e 9,9°. Esses valores indicam que a taxa de erro dentro da distribuição apresenta o meio dos valores nesses pontos, associando com o RMS é possível compreender juntamente com o BoxPlot o comportamento de dispersão do erro dentro da amostra. A Figura 34 exibe os valores de RMS, é possível visualizar que MediaPipe e Kinect apresentaram menor dispersão, embora seja comum em todos os sensores valores altos de erro máximo sendo possível resultado da sincronização ou ruído não filtrado.

Figura 33 - Box Plot do Erro Absoluto de cada sensor em relação ao Qualisys, para o movimento de membros inferiores.



Fonte: O Autor (2022).

Figura 34 - RMS do Erro Absoluto para cada sensor em frente a todos os dados de Membros inferiores em graus.



Fonte: O Autor (2022).

Após determinar as relações estatísticas e os resultados dos erros, calculou-se a correlação dos dados com o Qualisys para determinar o sensor mais próximo do padrão ouro. A Tabela 8 exibe a correlação de Pearson entre os equipamentos RGB-D em ordem crescente da esquerda para direita, intervalo de proximidade de 0 a 1. O Kinect e MediaPipe ambos obtiveram índice de Bom.

Tabela 8 - Correlação de Pearson dos sensores para todos os dados para dados de MMII. Azul (>0,75 e <0,9) e Amarelo (>0,5 e <0,75)

	Astra	Intel	Kinect	MediaPipe
Qualisys	0,62	0,66	0,80	0,87

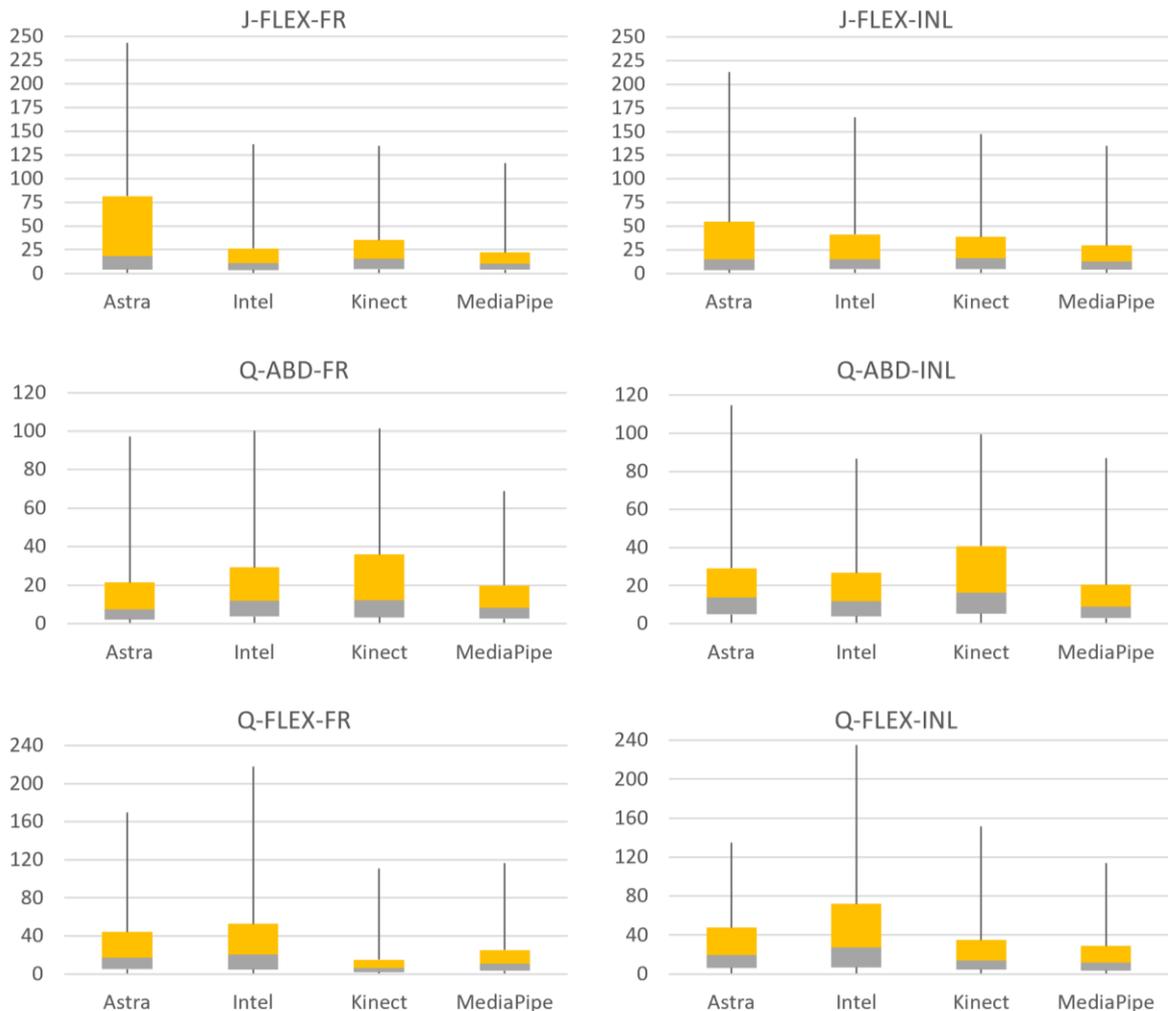
Fonte: O Autor (2022).

4.3.2 Análise por movimento

Foi realizada a mesma análise da seção anterior para cada tipo de movimento gravado com o objetivo de verificar o desempenho em cada articulação e inclinação de gravação. Pelos testes estatísticos, foi determinado que as amostras não se comportam como dados normalizados, portanto foi empregado o post-hoc não paramétrico de rank-sum de Friedman obtendo o valor de $p < 0,001$ para todos os sensores.

Utilizando os mesmos métodos de análise pra cada movimento, foram calculados o Erro Absoluto e o RMS para cada movimento. A Figura 35 condensa as informações da distribuição do E_A . Avaliando os resultados para cada movimento tem-se que: em Q-ADB-FR o sensor Astra teve o menor valor de mediana (5,39°); em Q-FLEX-FR o Kinect obteve o menor valor de mediana para o erro (4,28°); para os demais movimentos o MediaPipe apresentou os menores valores de Erro, sendo J-FLEX-FR (6,74°); J-FLEX-INL (8,56°); Q-ADB-INL (6,03°); Q-FLEX-INL (8,31°). Além disso, os movimentos de Q-ADB-FR e Q-FLEX-FR apresentaram em os segundos menores valores, respectivamente 5,63° e 7,67°.

Figura 35 - Box Plots agrupados do Erro Absoluto de cada movimento de MMII.



Fonte: O Autor (2022).

Para compreender todos os movimentos e valores medianos do erro a Tabela 9 apresenta os valores descritos para cada movimento. MediaPipe apresentou a

mediana média mais baixa assim como o desvio padrão. O sensor Kinect V2 possuiu a segunda menor média e desvio padrão da mediana do Erro Absoluto. Visto que os dados não são paramétricos a avaliação da mediana é um indicativo do comportamento dos dados, associando ao RMS é possível estabelecer e compreender o comportamento dos dados e avaliar qual apresentou melhor comportamento desejado.

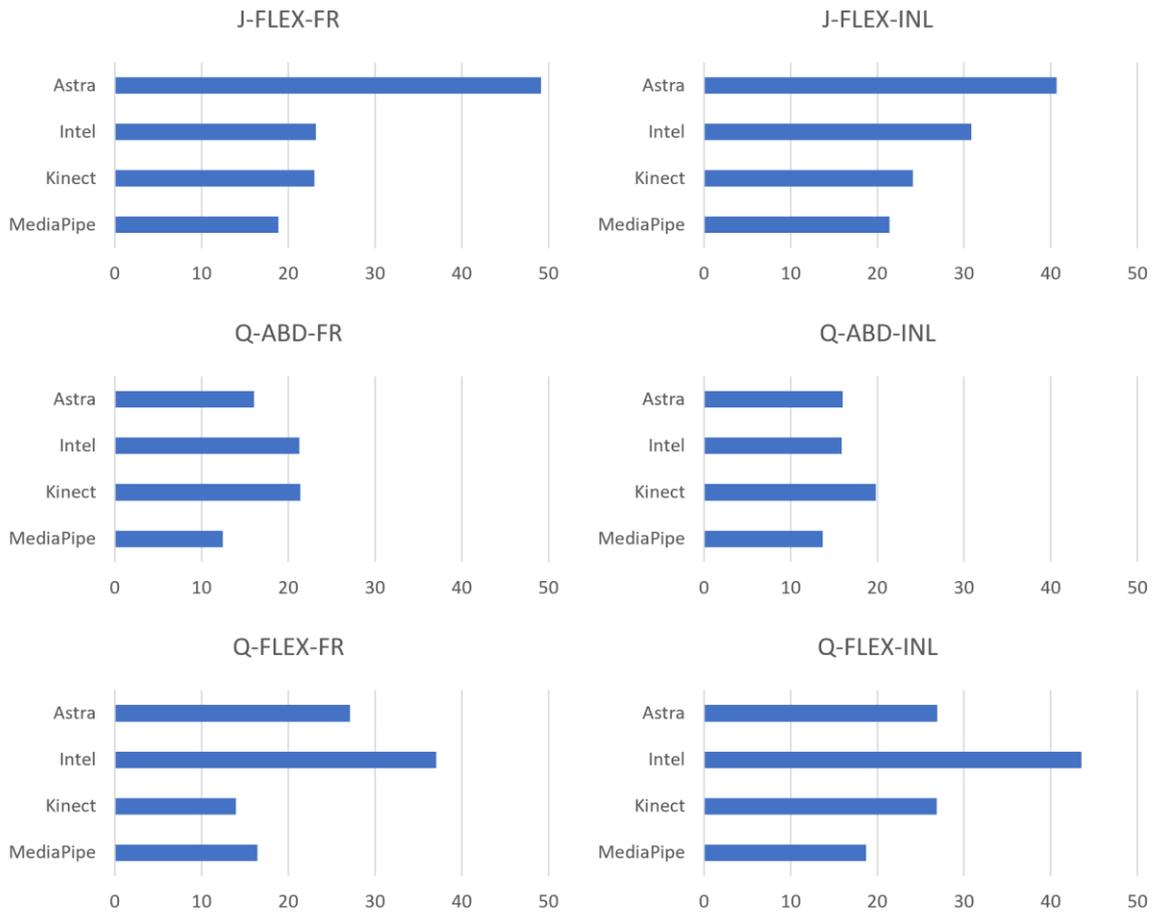
Tabela 9 - Tabela de Medianas do Erro Absoluto para membros inferiores em graus (°).

	Astra	Intel	Kinect	MediaPipe
J-FLEX-FR	14,34	7,3	10,74	6,74
J-FLEX-INL	11,3	10,26	11,38	8,56
Q-ABD-FR	5,39	8,08	8,86	5,63
Q-ABD-INL	8,79	7,79	10,89	6,03
Q-FLEX-FR	11,97	15,46	4,29	7,67
Q-FLEX-INL	13,22	20,5	9,64	8,31
Média	10,84 ± 3,26	11,57 ± 5,32	9,3 ± 2,62	7,16 ± 1,21

Fonte: O Autor (2022).

Avaliando os RMS do E_A dos movimentos juntamente com os Boxplots é possível compreender o comportamento dos dados em comparação com o padrão ouro e ter indicativos de qual apresentou um performance melhor em relação ao erro. Diante disso, a Figura 36 contém a comparação lado a lado de todos os movimentos em gráfico de barras. Percebe-se uma paridade entre o RMS e o Boxplot, 5 dos 6 menores valores de RMS coincidiram de ser o mesmo sensor com as menores medianas, embora sejam informações complementares elas não são correlatas. Apenas o movimento Q-ABD-FR que ocorreu discrepância, visto que o menor valor de mediana foi do Astra e do RMS do MediaPipe, isso ocorreu devido à dispersão dos dados do primeiro e terceiro quartil além de maior máximo, comparados ao Boxplot do MediaPipe.

Figura 36 - Gráficos de RMS do Erro Absoluto agrupado para todos os sensores para coleta de MMII.



Fonte: O Autor (2022).

Avaliando o comportamento médio do RMS ao longo dos movimentos o MediaPipe obteve o valor médio menor, assim como o desvio padrão. Desse modo pode-se interpretar que os dados do MediaPipe apresentaram menor dispersão dentre os sensores para membros inferiores.

Tabela 10 - Valores de RMS do Erro Absoluto e média para membros inferiores em graus (°).

	Astra	Intel	Kinect	MediaPipe
J-FLEX-FR	49,15	23,19	23,03	18,85
J-FLEX-INL	40,68	30,84	24,09	21,41
Q-ABD-FR	16,01	21,27	21,39	12,45
Q-ABD-INL	16,01	15,87	19,84	13,71
Q-FLEX-FR	27,13	37,09	13,96	16,41
Q-FLEX-INL	26,89	43,54	26,87	18,69
Média	29,31 ± 13,32	28,63 ± 10,44	21,53 ± 4,42	16,92 ± 3,39

Fonte: O Autor (2022).

Como na subseção anterior foi realizado o teste de correlação de Pearson para cada movimento, a Tabela 11 exibe a correlação de Pearson entre os equipamentos RGB-D, para todos os sensores e movimentos o valor de $p < 0,001$ nas matrizes de correlação. Kinect V2 obteve índice de correlação de Excelente em dois movimentos (Q-ABD-INL e Q-FLEX-FR) sendo o único a atingir essa métrica, embora na avaliação média tanto o Kinect e MediaPipe obtiveram índice de Bom.

Tabela 11 - Correlação de Pearson dos sensores para cada movimento de membros inferiores. Verde ($>0,9$), Azul ($>0,75$ e $<0,9$), Amarelo ($>0,5$ e $<0,75$) e Vermelho ($<0,5$). *P-valor* $< 0,001$ para todas as matrizes de correção de Pearson.

	Astra	Intel	Kinect	MediaPipe
J-FLEX-FR	0,31	0,83	0,83	0,89
J-FLEX-INL	0,51	0,69	0,81	0,86
Q-ABD-FR	0,54	0,26	0,82	0,69
Q-ABD-INL	0,7	0,37	0,91	0,66
Q-FLEX-FR	0,68	0,49	0,91	0,87
Q-FLEX-INL	0,71	0,41	0,69	0,84
Média	$0,58 \pm 0,15$	$0,51 \pm 0,22$	$0,83 \pm 0,08$	$0,8 \pm 0,10$

Fonte: O Autor (2022).

4.4 MEMBROS SUPERIORES

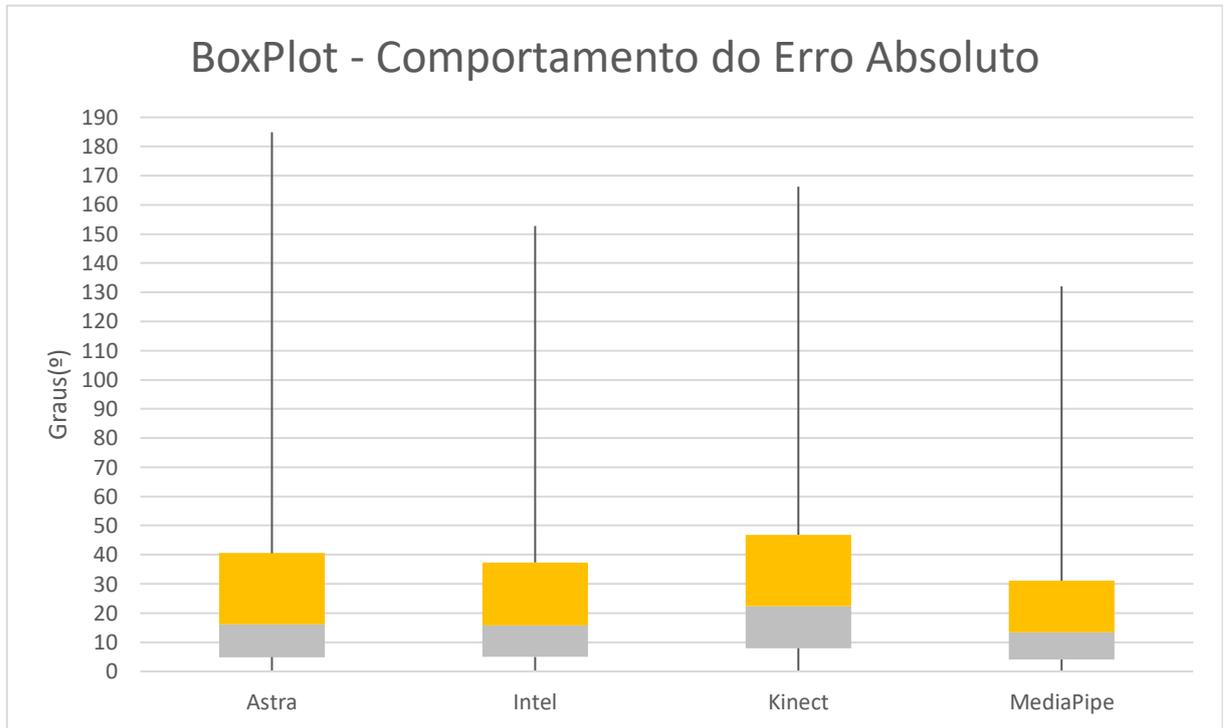
Foi avaliado o erro absoluto, RMS e correlação de Pearson entre o padrão ouro e os sensores para o conjunto total de dados, como também para cada movimento analisado. O processo foi o mesmo realizado para membros superiores, sendo possível a apresentação dos resultados de modo mais sucinto.

4.4.1 Análise Geral

Foi utilizado o mesmo passo a passo realizado para MMII, primeiramente foi analisado o conjunto de dados agrupados como apenas uma única amostra. Realizando o teste de normalidade foi verificado que as amostras não obedeciam a uma distribuição normal, com isso foi empregado o teste de Wilcoxon para avaliar a diferença entre os sensores, todas as amostras apresentam $p < 0,001$. O teste de rank-sum de Friedman obteve $p < 0,001$. Desse modo, foi avaliado o erro absoluto entre os equipamentos RGB-D e o padrão ouro. A Figura 37 exibe o erro absoluto das amostras em comparação pareada lado a lado dos sensores. Dos sensores, o

MediaPipe apresentou resultados de mediana de Erro Absoluto menor que os demais (9,33°), em contrapartida o Kinect V2 a pior mediana (14,49°).

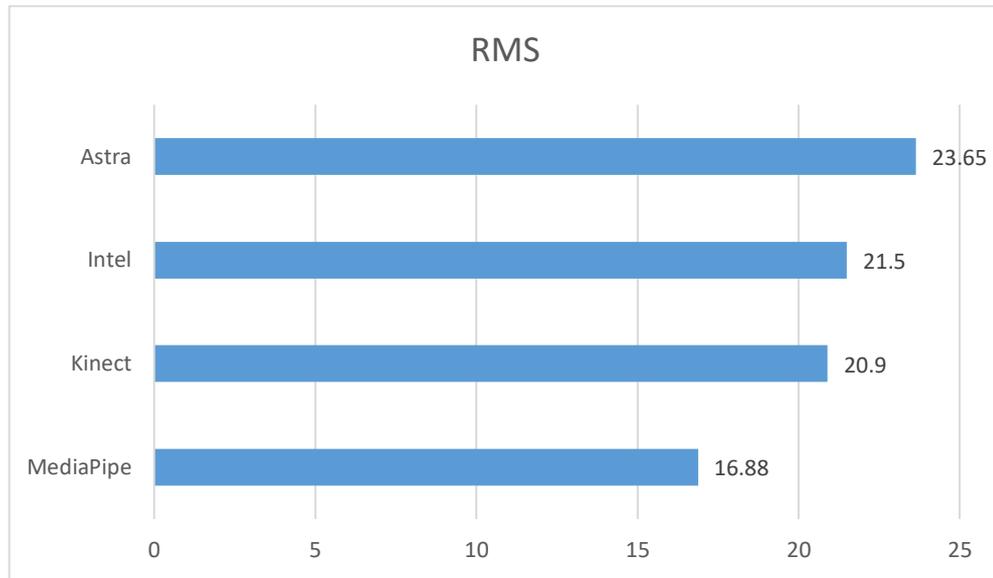
Figura 37 - Box Plot do Erro Absoluto de cada sensor em relação ao Qualisys, para o movimento de membros superiores.



Fonte: O Autor (2022).

Outro dado levantado foi o RMS para entender a variação do erro das amostras, a Figura 38 exibe os valores de RMS. MediaPipe apresentou o menor valor de RMS para o E_A , de 16,88°. Kinect e Intel tiveram resultados muito próximos de RMS, 20,9° e 21,5° respectivamente apenas uma diferença de 0,6°, um valor muito pequeno considerando o alcance do movimento. Astra apresentou o resultado mais elevado com 23,65°, uma diferença de 10% maior se comparada com o RMS do Intel (terceiro mais alto).

Figura 38 - Gráfico de RMS do Erro Absoluto para cada sensor em frente a todos os dados de Membros superiores.



Fonte: O Autor (2022).

Para finalizar os resultados agrupados foi calculada a correlação de Pearson, exibida na Tabela 12. Para os membros superiores, os sensores demonstraram melhor correlação do que para membros inferiores, pois todos alcançaram valores de Bom. Astra alcançou a classificação de Bom no valor limite da classe entre Moderado e Bom. A análise por movimento possibilitou uma melhor visualização do comportamento do sensor.

Tabela 12 - Correlação de Pearson dos sensores para todos os dados de MMSS. Azul (>0,75 e <0,9).

	Astra	Intel	Kinect	MediaPipe
Qualisys	0,75	0,77	0,82	0,85

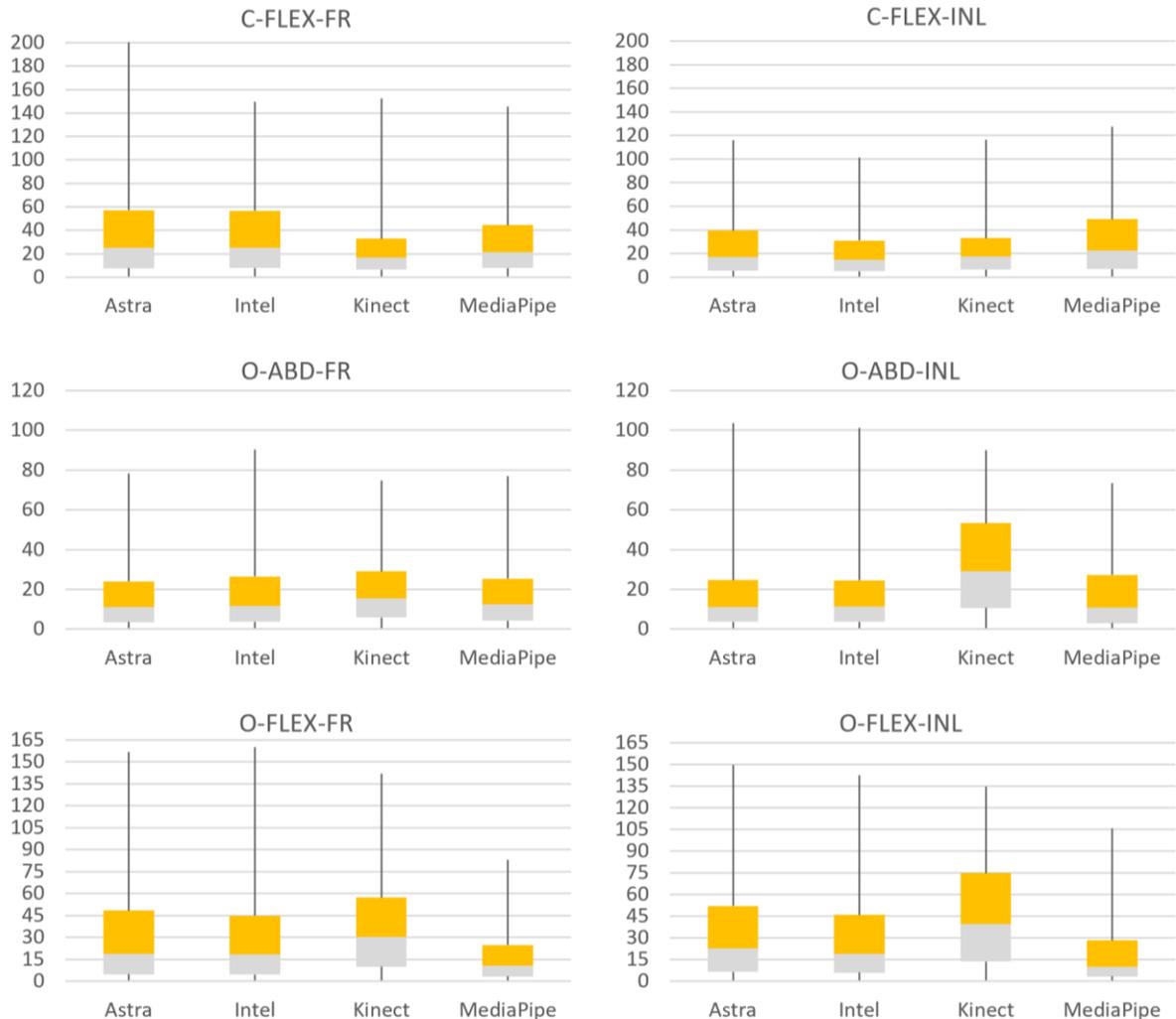
Fonte: O Autor (2022).

4.4.2 Análise por movimento

Assim como na seção anterior, os dados apresentam pelo teste de rank-sum de Friedman p -value < 0,001 considerando todos os sensores para a comparação. Utilizando os mesmos métodos de análise para cada movimento, foram calculados o Erro Absoluto e o RMS para cada movimento. A Figura 39 condensa as informações da distribuição do E_A . Em C-FLEX-FR Kinect obteve a menor mediana com 10,26°. Em C-FLEX-INL, Intel obteve a menor com 9,67°. Astra obteve os menores valores

de mediana em dois movimentos: O-ABD-FR, 7,5° e O-ABD-INL, 7,48°. Assim como ele, MediaPipe apresentou menores valores em dois movimentos: O-FLEX-FR 10,26° e O-FLEX-INL 6,81°.

Figura 39 - Box Plots agrupados do Erro Absoluto de cada movimento de MMSS.



Fonte: O Autor (2022).

Para compreender todos os movimentos e valores medianos do erro a *Tabela 13* apresenta os valores descritos para cada movimento. MediaPipe apresentou a mediana média mais baixa, seguido do Intel, ambos apresentaram um desvio padrão pequeno em comparação aos demais. A mediana mais baixa é um indicativo do comportamento, representando que em metade dos dados o erro foi ao menos menor que a mediana, sendo um indicador positivo seu valor mais baixo.

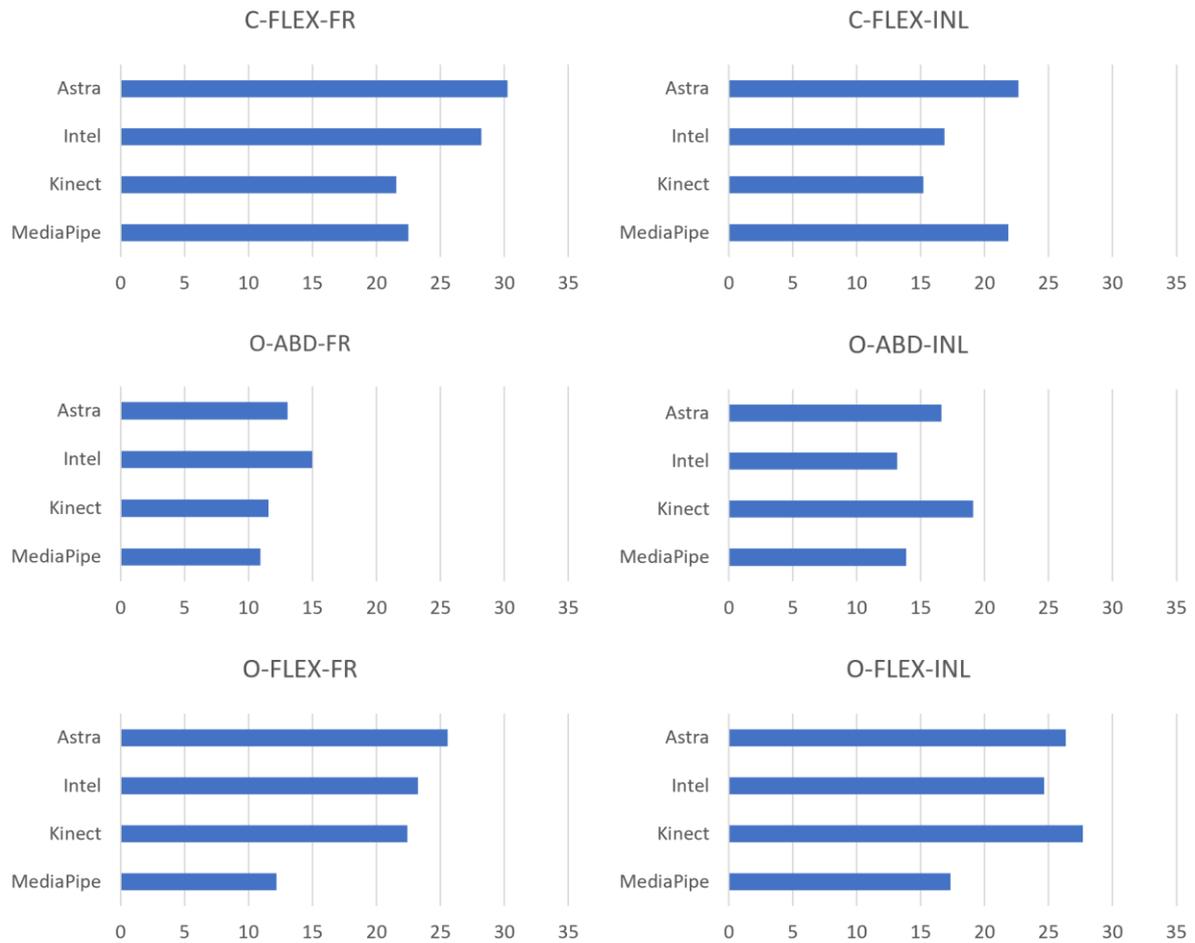
Tabela 13 - Tabela dos valores de mediana do Erro Absoluto para cada movimento avaliado.

	Astra	Intel	Kinect	MediaPipe
C-FLEX-FR	17,31	17,23	10,26	13,56
C-FLEX-INL	11,6	9,67	11,13	15,92
O-ABD-FR	7,5	8,03	9,6	8,17
O-ABD-INL	7,48	7,69	18,59	7,97
O-FLEX-FR	13,88	13,58	20,31	7,47
O-FLEX-INL	16,39	13,18	26,15	6,81
Média	12,36 ± 4,27	11,56 ± 3,74	16,01 ± 6,72	9,98 ± 3,79

Fonte: O Autor (2022).

Ademais, para avaliar o comportamento da dispersão do Erro é possível associar ao RMS. A Figura 40 apresenta os resultados do RMS do Erro Absoluto. Kinect apresentou menores valores de RMS em C-FLEX-FR e C-FLEX-INL, no segundo apresentou a segunda menor mediana (11,13°) mas com dispersão menor dos dados. Em O-ABD-INL, Intel apresentou o menor RMS com 13,17° seguido diretamente pelo MediaPipe com 13,86°. MediaPipe obteve os menores valores em ambos os movimentos que ele obteve mediana menor, O-FLEX-FR 12,19° e O-FLEX-INL 17,33°, além de O-ABD-FR com RMS de 10,94°.

Figura 40 - Gráficos de RMS do Erro Absoluto agrupado para todos os sensores para coleta de MMSS.



Fonte: O Autor (2022).

Em adicional, a Tabela 14 expressa dos RMS ao longo dos movimentos e a média comparando os resultados. O MediaPipe obteve a menor média de RMS juntamente do menor desvio padrão, seguido pelo Kinect. Portanto, percebe-se o comportamento menos disperso do MediaPipe.

Tabela 14 - Valores de RMS do Erro Absoluto e média para membros superiores.

	Astra	Intel	Kinect	MediaPipe
C-FLEX-FR	30,24	28,2	21,55	22,49
C-FLEX-INL	22,65	16,86	15,21	21,87
O-ABD-FR	13,05	14,98	11,56	10,94
O-ABD-INL	16,61	13,17	19,11	13,87
O-FLEX-FR	25,56	23,26	22,43	12,19
O-FLEX-INL	26,34	24,66	27,68	17,33
Média	22,41 ± 6,45	20,19 ± 6,01	19,59 ± 5,68	16,45 ± 4,93

Fonte: O Autor (2022).

Como na seção anterior foi realizado o teste de correlação de Pearson para cada movimento, a Tabela 15 exibe a correlação de Pearson entre os equipamentos RGB-D. Para membros superiores a correlação entre os equipamentos RGB-D e RGB foi alta, visto que a correlação mais baixa obteve a métrica de Bom. Kinect e MediaPipe obtiveram correlação Excelente.

Tabela 15 - Correlação de Pearson dos sensores para cada movimento me membros superiores. Verde (>0,9), Azul (>0,75 e <0,9) e Amarelo (>0,5 e <0,75).

	Astra	Intel	Kinect	MediaPipe
C-FLEX-FR	0,6	0,63	0,83	0,78
C-FLEX-INL	0,73	0,87	0,93	0,85
O-ABD-FR	0,91	0,87	0,96	0,97
O-ABD-INL	0,84	0,90	0,96	0,96
O-FLEX-FR	0,73	0,81	0,91	0,96
O-FLEX-INL	0,82	0,76	0,94	0,95
Média	0,77 ± 0,11	0,81 ± 0,10	0,92 ± 0,05	0,91 ± 0,08

Fonte: O Autor (2022).

5 DISCUSSÃO

O objetivo do trabalho foi avaliar a capacidade de análise de equipamentos RGB-D como fonte de estimativa de posição articular para avaliação biomecânica, sendo fundamental a avaliação da capacidade dos equipamentos em reconhecer e estimar a posição 3D das articulações possibilitando a medição das variações angulares, uma vez que amplitude de movimento é um parâmetro utilizado para acompanhar a evolução e estado do paciente (YANG e colab., 2014).

Os resultados demonstraram, que independente da análise, seja com todos os dados agrupados, seja por segmento corporal e movimento, alguns fatos sobre o comportamento dos equipamentos merecem destaque. A API MediaPipe apresentou resultados melhores nas métricas avaliadas. Astra e Intel apresentaram desempenhos intermediários na avaliação geral e MMSS. Por outro lado, o Kinect apresentou resultados piores na avaliação geral para o Erro Absoluto, embora para MMII o segundo melhor resultado.

O MediaPipe obteve os melhores resultados de Erro Absoluto e de RMS, de modo que a mediana do Erro foi a mais baixa em média nas análises de MMSS ($\bar{E}_A = 9,98^\circ \pm 3,79$ e $RMS = 16,45^\circ \pm 4,93$) e MMII ($\bar{E}_A = 7,16^\circ \pm 1,21$ e $RMS = 16,92^\circ \pm 3,39$). Considerando todos as gravações para os membros, MediaPipe performou com os resultados de $\bar{E}_A = 8,57^\circ \pm 3,06$ e $RMS = 16,68^\circ \pm 4,05$. Como a distribuição dos dados não se comportaram de modo paramétrico o dado de mediana traduz mais informações de comportamento do que a média. Juntamente a mediana, o RMS representa a dispersão dos dados de Erro Absoluto no estudo. Avaliando todos os senhores que apresentaram significância estatística ($p\text{-value} < 0,001$) é perceptível que o MediaPipe apresentou uma mediana de erro mais baixa para o erro e com o RMS menor também, sendo assim é possível afirmar o melhor resultado para o MediaPipe.

Nesse viés é necessário avaliar o erro em frente a outras formas de mensuração. Hancock et al.(HANCOCK e colab., 2018) em seu trabalho avaliou diferentes formas e equipamentos de goniometria comuns em ambientes clínicos para medição de ângulo de joelho em flexo extensão. No trabalho foram avaliados goniômetro digital da Halo, goniômetro de braços longos e curtos, estimativa visual e goniômetro de aplicativos de celular, obtendo a estimativa de erro para cada equipamento, respectivamente: $6^\circ; 10^\circ; 14^\circ, 14^\circ$ e 12° . Diante disso, percebe-se que o

erro do MediaPipe estaria abaixo da escala de medição de goniômetros de braços curtos e longos (erro entre 10° e 14°), enquanto os demais sensores estariam no dentro o intervalo. Considerando o erro médio do MediaPipe de $8,57^{\circ}$ ele está abaixo do erro de ferramentas utilizadas comumente na clínica fisioterapia para a avaliação de ângulo, sendo uma ferramenta que utiliza apenas uma câmera RGB e consegue estimar dados de profundidade e mediante a isso calcular variações angulares tridimensionais. Dessa forma, o desempenho do MediaPipe foi surpreendente conseguindo ter métricas de avaliação melhores que três equipamentos RGB-D (Astra, Intel e Kinect V2) para medição de variação angular.

Ademais, avaliando a literatura na utilização de equipamento RGB, Martin et al avaliou a aplicação de OpenPose juntamente com Kinect v2 (combinação de rastreamento do OpenPose com as imagens de profundidade do Kinect) para a medição de ângulo de tornozelo, obtendo valores de correlação ($r = 0,89$) com o padrão ouro, resultado classificado como Bom (MARTÍN e colab., 2021). Diferentemente do trabalho apresentado, Martin utilizou uma técnica mista e obteve a mesma classificação que o Kinect sozinho alcançou nesse trabalho para MMII, classificação de correlação de Bom.

Vilas-Boas, em seus trabalhos e pesquisas, avaliou o Kinect V2 quanto à análise espacial e angular de sua inferência. Em ambos os trabalhos verificados neste estudo, verificou-se a avaliação dos ângulos das principais articulações durante a marcha. Para os ângulos do quadril, Vilas-Boas obteve resultados de correlação Pobre ($r = 0,13$ e $r = 0,30$) enquanto para o joelho os resultados foram de Excelente e de Bom ($r = 0,94$ e $r = 0,76$) (VILAS-BOAS, Maria Do Carmo e CHOUPINA e colab., 2019; VILAS-BOAS, Maria Do Carmo e ROCHA e colab., 2019). Mesmo apresentando resultado Excelente em um tipo de movimento, a análise média do Kinect para variação angular ainda está na escala de correlação Bom nesse trabalho. Vale ressaltar que uma das piores métricas do Kinect em dispersão foi em abdução de quadril, mesmas articulações que apresentou os piores valores no trabalho de Vilas-Boas, corroborando com os resultados negativos para o Kinect na avaliação de quadril.

Para membros inferiores, sensores Astra e Intel apresentaram os piores resultados tanto de correlação quanto numérico. Ambos apresentaram resultado moderado de r , como também, os resultados mais altos de Erro Absoluto e RMS, valores muito distantes dos outros sensores. Kinect v2 e MediaPipe apresentaram

resultados bons de correlação. MediaPipe apresentou resultados mais próximos do padrão ($\bar{E}_A = 12,51 \pm 11,33$ e $RMS = 17,22$), enquanto o Kinect apresentou resultados intermediários entre os sensores de $\bar{E}_A = 9,3^\circ \pm 2,62$ e $RMS = 21,53^\circ \pm 4,42$. Já Astra e Intel apresentaram, respectivamente, os valores de $\bar{E}_A = 10,84^\circ \pm 3,27$ e $RMS = 29,31^\circ \pm 13,32$, e $\bar{E}_A = 11,57^\circ \pm 5,32$ e $RMS = 28,63^\circ \pm 10,44$.

Analisando os resultados para membros superiores é possível perceber uma semelhança nos resultados, visto que o MediaPipe desempenho melhor que os demais sensores, e Kinect apresentou o pior resultado de Erro e RMS ($\bar{E}_A = 16,01^\circ \pm 6,72$ e $RMS = 19,59^\circ \pm 5,68$). A correlação de Pearson dos sensores foi melhor para MMSS do que MMII, para todos. Conforme visto nas Tabela 8, Tabela 12 e Tabela 15 a correlação foi melhor para todos os dados os sensores obtendo ao mínimo a correlação de moderado. Astra e RealSense apresentaram resultado de moderado e bom na avaliação de todos os dados analisados como uma única amostra. Entretanto quando avaliado cada movimento ambos apresentaram uma correlação de bom. Para valores de \bar{E}_A e RMS ambos apresentaram resultados próximos aos de MMII, Astra obteve $\bar{E}_A = 12,36^\circ \pm 4,27$ e $RMS = 22,41^\circ \pm 6,45$, enquanto o Intel obteve $\bar{E}_A = 11,56^\circ \pm 3,74$ e $RMS = 20,19^\circ \pm 6,02$. Não foram encontrados na literatura artigos com avaliação do desempenho de métricas angulares desses equipamentos.

Ademais, o Kinect e o MediaPipe apresentaram melhor correlação ao Qualisys, quando avaliado cada tipo de movimento ambos apresentaram correlação de Excelente. Mangal e Tiwari (MANGAL e TIWARI, 2021) em seu trabalho em avaliação de disfunção muscular encontrou, também, valores de correlação excelente para movimentos de abdução e flexão de ombro, respectivamente $r = 0,98$ e $r = 0,96$.

Dessa maneira, o equipamento deve ser capaz de representar as características naturais das funções que descrevem o movimento. O padrão ouro utilizado Qualisys apresentaram uma taxa de captura de 180hz e um precisão métrica de 0,05mm na área de gravação, características muito superiores ao deslocamento humano, visto que a marcha humana tem aproximadamente 0,75 Hz (IQBAL e colab., 2020). Ante o exposto, torna-se crucial que as curvas geradas pelos dados coletados apresentem uma aproximação.

Contudo, a captação de perturbações e de ruídos pelos equipamentos RGB-D alteraram o formato da curva, razão pela qual se justificou a aplicação de filtros

como método fundamental para a reconstrução apropriada dos dados. Assim como variações de erro muito abruptas pode ser fruto de ruídos não filtrados. Rocha et al. utilizou um conjunto de filtros de média móvel e Butterworth para suavizar os dados de eliminando as interferências (ROCHA e colab., 2018), sendo utilizados os mesmos padrões de filtros para todos os sensores RGB-D no presente estudo. Ainda assim, as curvas apresentaram padrões de perturbações que não foram 100% removidas como pode ser visto no **Erro! Fonte de referência não encontrada.** com os dados sincronizados. Diante desse resultado, foi possível perceber a sensibilidade do equipamento a interferência.

De modo geral, nenhum dos sensores testados na pesquisa não apresentaram curvas com o mesmo padrão de suavidade do Qualisys, mesmo com a utilização de filtros capazes de reduzir significativamente as imperfeições, descontinuidades e picos das curvas captadas.

5.1 LIMITAÇÕES E DIFICULDADES

As limitações ocorreram na coleta dos dados, visto a baixa quantidade de marcadores reflexivos foi necessário que a coleta fosse feita separadamente dobrando o tempo de setup e prolongando a duração da coleta. Além do mais, a qualidade de reflexão não estava no máximo de desempenho o que acarretou perdas de identificação de marcadores e subsequentemente a não reconstrução da articulação no Visual 3D.

As dificuldades teóricas em encontrar publicações que tivessem comparado o MediaPipe a padrões devido a uma API nova impuseram adversidades na análise dos resultados e metodologia de pesquisa.

5.2 TRABALHOS FUTUROS

O resultado da pesquisa demonstrou a viabilidade da aplicação do MediaPipe para fins de medição de variação angular em fins de avaliação biomecânica, os próximos passos da pesquisa serão a busca ativa de outros métodos RGB que apresentem cálculo de profundidade, como também a realização de testes em outras atividades mais complexas e funcionais e testes com usuários em ambiente clínico.

6 CONCLUSÃO

A aplicação de técnicas de captura e análise de movimento representa ferramentas poderosas na área de reabilitação e pode trazer diversas vantagens para o tratamento de pacientes. A aplicação de equipamentos Mocaps ópticos de marcadores reflexivos de modo amplo em ambientes clínicos é limitada devido ao alto custo dos mesmos assim como a alta demanda de tempo e mão de obra especializada para o *setup*. Considerando o contexto vivenciado na pandemia, houve em muitos casos a impossibilidade e/ou a necessidade de afastamento do paciente de centros de tratamento pelo risco de contaminação. Desse modo, evidenciou-se a necessidade de acompanhamento remoto de modo fiel para a continuidade de tratamentos de reabilitação motora.

Nessa perspectiva de utilização em ambiente clínico e até homecare, aplicações RGB-D e RGB podem representar a mudança de paradigma na área da saúde. Entretanto, para isso é necessário compreender a proximidade desses equipamentos do padrão ouro. Visto isso, este trabalho buscou analisar a capacidade de avaliação e estimativa angular de articulações. Foram investigados um total de 3 equipamentos RGB-D (Kinect V2, Astra e Intel RealSense) e uma técnica baseada em imagens RGB (MediaPipe). Dentre os resultados, os equipamentos da Intel e Astra tiveram os piores valores, tanto de correlação com o padrão ouro, quanto numéricos de estimativa do ângulo do movimento, dando indícios da sua inviabilidade para aplicações clínicas.

Em contrapartida, Kinect V2 e MediaPipe apresentaram os melhores valores. O resultado do Kinect V2 corrobora com os dados da literatura com correlações de Bom e Excelente para avaliação angular. O MediaPipe apresentou um resultado surpreendente, sendo equiparado ao Kinect V2 com resultados numéricos melhores e dentro de escalas de erro de equipamentos clínicos. Desse modo, isso representa a possibilidade de desenvolvimento e aplicações de técnicas utilizando apenas componentes RGB para aplicações de reabilitação de forma equiparada a soluções que utilizam tanto imagens RGB quando de profundidade.

REFERÊNCIAS

- ABDULHAY, Enas e colab. **Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease.** Future Generation Computer Systems, v. 83, p. 366–373, 2018. Disponível em: <<https://doi.org/10.1016/j.future.2018.02.009>>.
- ALBERT, Justin Amadeus e colab. **Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study.** Sensors (Switzerland), v. 20, n. 18, p. 1–22, 2 Set 2020.
- BAKER, Richard. **Gait analysis methods in rehabilitation.** Journal of NeuroEngineering and Rehabilitation, v. 3, p. 1–10, 2006.
- BAZAREVSKY, Valentin e colab. **BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs.** 11 Jul 2019. Disponível em: <<https://arxiv.org/abs/1907.05047v2>>. Acesso em: 24 ago 2021.
- BAZAREVSKY, Valentin e colab. **BlazePose: On-device Real-time Body Pose tracking.** 17 Jun 2020. Disponível em: <<https://arxiv.org/abs/2006.10204v1>>. Acesso em: 24 ago 2021.
- BEAK, Seung Hui e colab. **Upper torso and pelvis linear velocity during the downswing of elite golfers.** BioMedical Engineering Online, v. 12, n. 1, p. 1–12, 2013.
- BERNARDINA, Gustavo R.D. e colab. **Moving system with action sport cameras: 3D kinematics of the walking and running in a large volume.** PLoS ONE, v. 14, n. 11, p. 1–13, 2019. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0224182>>.
- BOLINK, S. A.A.N. e colab. **Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit-stand transfers and step-up transfers: Comparison with an optoelectronic motion capture system.** Medical Engineering and Physics, v. 38, n. 3, p. 225–231, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.medengphy.2015.11.009>>.
- BOSSO, Luciana Rocco e GOLIAS, Andrey Rogério Campos. **A postura de atletas de ginástica rítmica: análise através da fotometria.** Revista Brasileira de Medicina do Esporte, v. 18, n. 5, p. 333–337, 2012.
- BURLE, Alexandre De Queiroz e colab. **Real-time approach for gait analysis using the Kinect v2 sensor for clinical assessment purpose.** Proceedings - 2020 22nd Symposium on Virtual and Augmented Reality, SVR 2020, p. 144–153, 2020.
- CAO, Zhe e colab. **OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields.** IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 43, n. 1, p. 172–186, 2021.
- CAPPOZZO, Aurelio e colab. **Human movement analysis using**

stereophotogrammetry. Part 1: Theoretical background. *Gait and Posture*, v. 21, n. 2, p. 186–196, 2005.

CHEN, Chengjun e colab. **Repetitive assembly action recognition based on object detection and pose estimation.** *Journal of Manufacturing Systems*, v. 55, n. May, p. 325–333, 2020. Disponível em: <<https://doi.org/10.1016/j.jmsy.2020.04.018>>.

CIPPITELLI, Enea e colab. **Validation of an Optimized Algorithm to Use Kinect in a Non-Structured Environment for Sit-to-Stand Analysis.** [S.l: s.n.], 2015.

D'ANTONIO, Erika e colab. **Validation of a 3D Markerless System for Gait Analysis Based on OpenPose and Two RGB Webcams.** *IEEE Sensors Journal*, v. 21, n. 15, p. 17064–17075, 1 Ago 2021.

DA GAMA, A.E.F. e colab. **Markerless gesture recognition according to biomechanical convention.** XXIV Brazilian Congress on Biomedical Engineering, p. 2033–2036, 2014.

DA GAMA, Alana Elza Fontes e colab. **MirrARbilitation: A clinically-related gesture recognition interactive tool for an AR rehabilitation system.** *Computer Methods and Programs in Biomedicine*, v. 135, p. 105–114, 2016a. Disponível em: <<http://dx.doi.org/10.1016/j.cmpb.2016.07.014>>.

DA GAMA, Alana Elza Fontes e colab. **MirrARbilitation: A clinically-related gesture recognition interactive tool for an AR rehabilitation system.** *Computer Methods and Programs in Biomedicine*, v. 135, p. 105–114, 1 Out 2016b.

DOLATABADI, Elham e colab. **The Toronto Rehab Stroke Pose Dataset to Detect Compensation during Stroke Rehabilitation Therapy.** 2017.

GIL, Elton e MOURA, Xavier. **Desenvolvimento de um sistema de captura de baixo custo destinado a determinação da cinemática humana .** **Desenvolvimento de um sistema de captura de baixo custo destinado a determinação da cinemática humana .** v. 4, p. 13–20, 2015.

GitHub - tensorflow/tflite-support: TFLite Support is a toolkit that helps users to develop ML and deploy TFLite models onto mobile / IoT devices. Disponível em: <<https://github.com/tensorflow/tflite-support>>. Acesso em: 27 ago 2021.

GOMIDE, João Victor Boechat e colab. **Captura de Movimento e Animação de Personagens em Jogos.** VIII Brazilian Symposium on Games and Digital Entertainment, p. 1–15, 2009.

HAFER, Jocelyn F. e colab. **Measuring markers of aging and knee osteoarthritis gait using inertial measurement units.** *Journal of Biomechanics*, v. 99, p. 109567, 2020. Disponível em: <<https://doi.org/10.1016/j.jbiomech.2019.109567>>.

HANCOCK, Graeme Ethan e HEPWORTH, Tracey e WEMBRIDGE, Kevin. **Accuracy and reliability of knee goniometry methods.** *Journal of Experimental Orthopaedics*. [S.l: s.n.], 2018

- HANSEN, Clint e colab. **Sequence-dependent rotation axis changes and interaction torque use in overarm throwing.** *Journal of Sports Sciences*, v. 34, n. 9, p. 878–885, 2016.
- HASLER, Nils e colab. **Markerless Motion Capture with Unsynchronized Moving Cameras.** p. 224–231, 2009.
- HUANG, Lin e colab. **Survey on depth and RGB image-based 3D hand shape and pose estimation.** *Virtual Reality and Intelligent Hardware*, v. 3, n. 3, p. 207–234, 2021. Disponível em: <<http://dx.doi.org/10.1016/j.vrih.2021.05.002>>.
- IQBAL, Muhammad e colab. **Comparison of dual task specific training and conventional physical therapy in ambulation of hemiplegic stroke patients: A randomized controlled trial.** *Journal of the Pakistan Medical Association*, v. 70, n. 1, p. 7–10, 2020.
- KUNST, Victo e colab. **Mini-MirrAR: Criando um dispositivo portatil para sistemas interativos de reabilitação motora.** [S.d.].
- KURIHARA, K. e colab. **Optical motion capture system with pan-tilt camera tracking and real time data processing.** *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, v. 2, n. May, p. 1241–1248, 2002. Disponível em: <<http://nobunaga.t.u-tokyo.ac.jp/publications/pdf2002/icra02/kurihara.pdf>>.
- LACHAT, Elise e colab. **Assessment and calibration of a RGB-D camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling.** *Remote Sensing*, v. 7, n. 10, p. 13070–13097, 2015.
- LAFAYETTE, Thiago e TEIXEIRA, João e DA GAMA, Alana. **Hybrid Solution for Motion Capture with Kinect v2 to Different Biotypes Recognition.** [S.l.]: Springer Singapore, 2019. Disponível em: <https://doi.org/10.1007/978-981-13-2119-1_39>.
- LI, Min e colab. **Three-Dimensional Pose Estimation of Infants Lying Supine Using Data from a Kinect Sensor with Low Training Cost.** *IEEE Sensors Journal*, v. 21, n. 5, p. 6904–6913, 1 Mar 2021.
- LIN, Wei e DING, Jiangtao. **Behavior detection method of OpenPose combined with Yolo network.** *Proceedings - 2020 International Conference on Communications, Information System and Computer Engineering, CISCE 2020*, p. 326–330, 2020.
- LIU, Yebin e GALL, Juergen e STOLL, Carsten. **Markerless Motion Capture of Multiple Characters Using Multiview Image Segmentation.** v. 35, n. 11, p. 2720–2735, 2013.
- LUGARESI, Camillo e colab. **MediaPipe: A Framework for Building Perception Pipelines.** 14 Jun 2019. Disponível em: <<https://arxiv.org/abs/1906.08172v1>>. Acesso em: 24 ago 2021.

LUN, Roanna e ZHAO, Wenbing. **A Survey of Applications and Human Motion Recognition with Microsoft Kinect**. [S.l.: s.n.], 2015. v. 29. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S0218001415550083>>.

MANGAL, Naveen Kumar e TIWARI, Anil Kumar. **A review of the evolution of scientific literature on technology-assisted approaches using RGB-D sensors for musculoskeletal health monitoring**. Computers in Biology and Medicine. [S.l.]: Elsevier Ltd. , 1 Maio 2021

MANUEL, Jose e PUEO, Basilio e JIMENEZ-OLMEDO, Jose Manuel. **El uso de la tecnología de captura de movimiento para el análisis del rendimiento deportivo**. 2017.

Marker Set Guidelines - Visual3D Wiki Documentation. Disponível em: <https://www.c-motion.com/v3dwiki/index.php/Marker_Set_Guidelines>. Acesso em: 13 dez 2019.

MARTÍN, Guillermo Díaz San e colab. **Automatic ankle angle detection by integrated rgb and depth camera system**. Sensors, v. 21, n. 5, p. 1–21, 1 Mar 2021.

Matplotlib — Visualization with Python. Disponível em: <<https://matplotlib.org/>>. Acesso em: 24 maio 2022.

MENACHE, Alberto. **Understanding motion capture for computer animation**. [S.l.]: Morgan kaufmann, 2011. Disponível em: <https://books.google.com.br/books?id=bgkzx4q88ScC&hl=pt-BR&source=gbs_slider_cls_metadata_7_mylibrary>. Acesso em: 17 mar 2018.

MOHAN, Jithin e ASHOK, A. **A Reliable Robot Workspace Monitoring System Using Kinect v2**. 2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering, ICRIEEECE 2018, p. 686–689, 1 Jul 2018.

MOTTA, Everton Simões Da. **Desenvolvimento de um método para a captura de movimentos humanos usando uma câmera RGB-D**. 2016. 2016.

NumPy. Disponível em: <<https://numpy.org/>>. Acesso em: 24 maio 2022.

O'BRIEN, Jonathon e BRACEWELL, Robert Martyn e CASTILLO, Juan Alberto. **The effects of kinesthetic and visual motor imagery on interjoint coordination in the hemiplegic index finger: An experimental study using the index of temporal coordination**. Brazilian Journal of Occupational Therapy, v. 29, p. 1–21, 2021.

OTTE, Karen e colab. **Instrumental assessment of stepping in place captures clinically relevant motor symptoms of parkinson's disease**. Sensors (Switzerland), v. 20, n. 19, p. 1–13, 1 Out 2020.

PHAM, Trinh Thi Doan e colab. **Moving object detection with kinect v2**. 2016 IEEE International Conference on Consumer Electronics-Asia, ICCE-Asia 2016, 3 Jan

2017.

PLASTIRAS, George e KYRKOU, Christos e THEOCHARIDES, Theocharis. **You Only Look Once: Unified, Real-Time Object Detection**. ACM International Conference Proceeding Series, 2018.

PORTNEY, L. G. e WATKINS, M. P. **Foundations of clinical research: applications to practice**. 3rd edn ed. Philadelphia: [s.n.], 2015.

ROCHA, Ana Patrícia e colab. **System for automatic gait analysis based on a single RGB-D camera**. PLoS ONE, v. 13, n. 8, 1 Ago 2018.

SAMIR, Mohammed e GOLKAR, Ehsan e RAHNI, Ashrani Aizzuddin Abd. **Comparison between the Kinect™ V1 and Kinect™ V2 for respiratory motion tracking**. IEEE 2015 International Conference on Signal and Image Processing Applications, ICSIPA 2015 - Proceedings, p. 150–155, 2016.

SciPy. Disponível em: <<https://scipy.org/>>. Acesso em: 24 maio 2022.

SHETTY, Sachin. **Identify Parkinson ' s Disease Using Gait Analysis**. [S.d.].

SHOTTON, Jamie e colab. **Real-time human pose recognition in parts from single depth images**. Studies in Computational Intelligence, v. 411, p. 119–135, 2013.

SONG, Young Seop e colab. **Validation of attitude and heading reference system and microsoft kinect for continuous measurement of cervical range of motion compared to the optical motion capture system**. Annals of Rehabilitation Medicine, v. 40, n. 4, p. 568–574, 2016.

TEKIN, Bugra e BOGO, Federica e POLLEFEYS, Marc. **H+O: Unified egocentric recognition of 3D hand-object poses and interactions**. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, v. 2019-June, p. 4506–4515, 2019.

TensorFlow. Disponível em: <<https://www.tensorflow.org/?hl=pt-br>>. Acesso em: 27 ago 2021.

TÖLGYESSY, Michal e colab. **Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2**. Sensors (Switzerland), v. 21, n. 2, p. 1–25, 2021.

VILAS-BOAS, Maria do Carmo e CHOUPINA, Hugo Miguel Pereira e colab. **Full-body motion assessment: Concurrent validation of two body tracking depth sensors versus a gold standard system during gait**. Journal of Biomechanics, v. 87, p. 189–196, 18 Abr 2019.

VILAS-BOAS, Maria Do Carmo e ROCHA, Ana Patrícia e colab. **Validation of a single RGB-D camera for gait assessment of polyneuropathy patients**. Sensors (Switzerland), v. 19, n. 22, 2 Nov 2019.

VISWAKUMAR, Aditya e colab. **Human Gait Analysis Using OpenPose**. Proceedings of the IEEE International Conference Image Information Processing, v. 2019- Novem, p. 310–314, 2019.

WASENMÜLLER, Oliver e STRICKER, Didier. **Comparison of kinect v1 and v2 depth images in terms of accuracy and precision**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 10117 LNCS, p. 34–45, 2017.

WU, Qingqiang e colab. **RGB-D Videos-Based Early Prediction of Infant Cerebral Palsy via General Movements Complexity**. IEEE Access, v. 9, p. 42314–42324, 2021.

YANG, Yang e colab. **Reliability and validity of kinect RGB-D sensor for assessing standing balance**. IEEE Sensors Journal, v. 14, n. 5, p. 1633–1638, 2014.

ZENNARO, S. e colab. **Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications**. Proceedings - IEEE International Conference on Multimedia and Expo, v. 2015- August, 2015.

ZHANG, Fan e colab. **MediaPipe Hands: On-device Real-time Hand Tracking**. 2020. Disponível em: <<http://arxiv.org/abs/2006.10214>>.