



**DIOGO MIGUEL AFONSO CARDOSO**

Licenciado em Ciências da Engenharia Eletrotécnica e de  
Computadores

# DETEÇÃO DE MOVIMENTO BASEADA EM SINAIS DE RÁDIO FREQUÊNCIA

MESTRADO EM ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES

Universidade NOVA de Lisboa  
Março, 2022



# DETEÇÃO DE MOVIMENTO BASEADA EM SINAIS DE RÁDIO FREQUÊNCIA

**DIOGO MIGUEL AFONSO CARDOSO**

Licenciado em Ciências da Engenharia Eletrotécnica e de Computadores

**Orientador:** Rodolfo Alexandre Duarte Olivera  
*Professor associado com agregação, Universidade NOVA de Lisboa*

**Coorientador:** António Eduardo Carreiro Furtado  
*Investigador, Instituto de Telecomunicações*

## **Deteção de Movimento Baseada em Sinais de Rádio Frequência**

Copyright © Diogo Miguel Afonso Cardoso, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

## AGRADECIMENTOS

Primeiramente, gostaria de agradecer ao professor Rodolfo Oliveira, que me acompanhou durante este ano. Agradeço a sua disponibilidade em ajudar, o seu compromisso para que o trabalho finalizasse todas as etapas e, sobretudo, a sua boa disposição constante e otimismo em qualquer situação.

Agradeço ao meu co-orientador António Furtado de igual forma pela disponibilidade, e por-me ter guiado sobretudo ao nível dos resultados práticos.

Quero agradecer a todos os meus amigos, aos que me acompanharam durante todo o meu percurso académico e a todos aqueles que entraram na minha vida, desde os meus tempos de infância. Obrigado a todos por me ajudarem a construir este caminho.

Um especial obrigado à minha namorada Marina Bandarra, por todo o carinho, apoio e motivação que me deu ao longo dos últimos anos.

Por último, mas o agradecimento mais importante aos meus pais, João e Paula. Quero agradecer por terem sido os grandes responsáveis pela minha educação, e por sempre insistirem e batalharem para que chegasse até aqui. Nunca deixaram que nada me faltasse e deram sempre a força necessária para que superasse qualquer obstáculo.

A todos um muito obrigado.

## RESUMO

Nos últimos anos temos assistido a um interesse crescendo no rastreamento de movimento humano e no uso de sistemas de localização, onde se destaca a detecção por radiofrequência. As antenas FM representam os modelos mais comuns nos sistemas de radiodifusão atuais, no que toca à presença de ruído e interferências. Ainda que a classificação de imagem seja cada vez mais apoiada por diversos algoritmos de análise, fazer o uso dos sinais RF oferece vantagens a nível de sistemas de vigilância ou de aplicações de detecção de alvos.

Neste trabalho, é feito um estudo a vários conjuntos de imagem representativos de vários movimentos, e é proposto um algoritmo de classificação dos movimentos. Através da ferramenta Mediapipe Pose identificam-se 33 pontos localizados ao longo de um corpo humano. Dessa forma, foi feito o estudo sobre as coordenadas das articulações ao longo do conjunto, e considerando os vetores de deslocamento e os ângulos de deslocamento ao longo das imagens, o modelo adotado permite suportar um classificador final de estimação do contexto em cada transição de imagem, baseando-se em cada articulação. A classificação mostrou-se eficaz nas previsões efetuadas, inclusive quando uma das classes está embebida noutra, e vice-versa.

Com a introdução dos sinais RF, foi utilizado um sistema de rádio definido por *software* com duas antenas omnidirecionais. Foram definidos sub-intervalos para análise e correlação de métricas aplicadas à amplitude e fase do sinal recebido em cada antena, e os contextos no vídeo de imagem foram caracterizados. Foi possível observar diferenças entre as antenas, assim como identificar o possível instante da transição, entre os cenários envolvidos.

Ambos os domínios estudados mostraram-se eficazes na detecção de movimento num determinado cenário. As potencialidades do domínio RF devem ser aproveitadas, devido às suas características e capacidades que conseguem ter em sistemas de detecção.

**Palavras-chave:** Classificação supervisionada, Conjunto de imagem, Mediapipe Pose, Sinais RF, Rádio definido por *software*

## ABSTRACT

In the last years we have witnessed a growing interest in tracking human motion and in the use of location systems, where radiofrequency detection stands out. FM antennas represent the most common models in today's broadcasting systems, regarding the presence of noise and interference. Even if image classification is increasingly supported by various analysis algorithms, making use of RF signals offers advantages in surveillance systems or target detection applications.

In this work, a study is made of several sets of images representing various movements, and a motion classification algorithm is proposed. Using the Mediapipe Pose tool, 33 points located along a human body are identified. That way, a study was carried out on the coordinates of the joints along the set, and considering the displacement vectors and displacement angles along the images, the model adopted allows to support a final classifier of context estimation in each image transition based on each articulation. The classification proved to be effective in the predictions made, even when one of the classes is included in another, and vice-versa.

With the introduction of RF signals, it was used a Software-Defined Radio Setup with two omnidirectional antennas. The signal was received by the antennas, collected by a nearby radio FM station. In parallel the capture of a certain action occurring and in the path of the signal to the antennas was made. Sub-intervals were defined to the analysis and metric's correlation applied to the amplitude and phase of the received signal in each antenna, and contexts of the image video were characterized. It was possible to observe differences between the two antennas, as well as identify the possible transition instant between the envolved scenarios.

Both domains studied proved to be effective in movement detection in a given scenario. The potential of the RF domain must be taken into account, due to it's characteristics and capabilities that they can have in detection systems.

**Keywords:** Supervised classification, Image set, Mediapipe Pose, RF signal, Software-Defined Radio

# ÍNDICE

<b>Índice de Figuras</b>	<b>viii</b>
<b>Índice de Tabelas</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Objetivos . . . . .	1
1.3 Contribuições . . . . .	2
1.4 Estrutura do documento . . . . .	2
<b>2 Trabalho relacionado</b>	<b>4</b>
2.1 Descrição de imagem . . . . .	4
2.1.1 Detecção de movimento em imagens . . . . .	4
2.1.2 Computação de vetores de movimento . . . . .	6
2.1.3 Estimação de pose e representação do movimento . . . . .	8
2.2 Radar FMCW . . . . .	10
2.2.1 Arquitetura de <i>sensing</i> . . . . .	11
2.2.2 Processamento de sinal e <i>background</i> matemático . . . . .	12
2.2.3 Implementação de um radar FMCW . . . . .	15
2.3 Séries temporais e análise de características de classificação . . . . .	18
2.3.1 Representação de séries temporais e características de classificação . . . . .	18
2.3.2 <i>Features</i> globais . . . . .	22
2.4 Sinal de rádio e presença em imagem . . . . .	24
2.4.1 Métodos para reconhecimento de atividade e de gestos ao nível do rádio FMCW . . . . .	25
2.5 Trabalhos na área . . . . .	26
2.5.1 Reconhecimento de atividade humana geral . . . . .	27
2.5.2 Reconhecimento de gestos . . . . .	27
2.5.3 Monitorização de sinais vitais . . . . .	30

<b>3</b>	<b>Datasets de imagem</b>	<b>33</b>
3.1	Classificação de imagem . . . . .	33
3.1.1	Machine Learning . . . . .	33
3.2	Mediapipe <i>framework</i> . . . . .	34
3.2.1	Mediapipe Pose . . . . .	34
3.3	Consciência de contexto . . . . .	36
3.3.1	<i>Entradas e saídas do modelo</i> . . . . .	36
3.3.2	<i>Dataset</i> . . . . .	36
3.3.3	Modelo estatístico . . . . .	37
3.4	Análise do movimento . . . . .	38
3.4.1	Vetores deslocamento e angular . . . . .	39
3.4.2	Estudo de métricas . . . . .	39
<b>4</b>	<b>Resultados experimentais</b>	<b>42</b>
4.1	Ferramentas utilizadas . . . . .	42
4.1.1	Python . . . . .	42
4.2	Introdução do algoritmo usado . . . . .	43
4.3	Etiquetagem e preparação do estudo . . . . .	44
4.3.1	Preparação dos <i>datasets</i> . . . . .	45
4.4	Visualização gráfica das saídas do modelo . . . . .	47
4.4.1	Gráficos diferenciais . . . . .	49
4.4.2	Influência geral de cada articulação . . . . .	52
4.5	Vetores deslocamento e angular . . . . .	54
4.5.1	Componente $x$ e $y$ : análise por sub-classes . . . . .	55
4.5.2	Vetores deslocamento e angular: análise por sub-classes . . . . .	57
4.5.3	Vetores deslocamento e angular: análise pelas classes completas . . . . .	58
4.6	Métricas e classificador final . . . . .	61
<b>5</b>	<b>Sinais RF</b>	<b>67</b>
5.1	Recolha de dados . . . . .	67
5.1.1	Rádio definido por <i>software</i> . . . . .	67
5.1.2	Bandas de tecnologia . . . . .	68
5.2	Caracterização dos <i>datasets</i> . . . . .	68
5.3	Resultados experimentais . . . . .	70
5.3.1	Métricas aplicadas . . . . .	70
5.3.2	Correlação do valor das amplitudes entre as duas antenas . . . . .	72
<b>6</b>	<b>Conclusões</b>	<b>76</b>
	<b>Bibliografia</b>	<b>77</b>

## ÍNDICE DE FIGURAS

2.1	Ilustração do uso de pontos de correspondência (Fonte Própria). . . . .	7
2.2	Técnica de Imagens da História do Movimento para detecção do movimento humano. (Adaptado de [6]). . . . .	7
2.3	Diagrama de blocos de um radar FMCW simples (Fonte Própria). . . . .	11
2.4	Transmissão do sinal por um radar FMCW (Fonte Própria). . . . .	12
2.5	(a) Sinais recebidos por um radar FMCW e espectro das respectivas frequências <i>beat</i> (Adaptado de [23]). (b) Frequência <i>beat</i> de referência detetada no processo de calibração da deteção de um alvo (Adaptado de [24]). . . . .	13
2.6	Em cima o sinal transmitido no domínio frequência-tempo, em baixo o sinal <i>beat</i> recebido de um alvo em movimento (Adaptado de [25]). . . . .	14
2.7	(a) Espectro do sinal modulado LO, representando a frequência instantânea de uma rampa do sinal. (b) Espectro do sinal de receção, que sofreu reflexão noutros objetos. (c) Sinal IF produzido à saída do <i>mixer</i> (Adaptado de [23]). . . . .	16
2.8	Diagrama de blocos de um radar FMCW em banda base complexa (Adaptado de [27]). . . . .	16
2.9	(a) Espectro do sinal modulado LO, representando a frequência instantânea de uma rampa do sinal. (b) Espectro do sinal de receção, que sofreu reflexão noutros objetos. (c) Sinal que entra no <i>mixer</i> em quadratura. (d) Sinal IF que sai do <i>mixer</i> em quadratura (Adaptado de [23]). . . . .	17
2.10	Medidas de semelhança de séries temporais, no domínio do tempo, com base em características de classificação (Adaptado de [28]). . . . .	19
2.11	Diagrama de Voronoy. Distâncias de poder com $p = r = [2$ (mais à esquerda, 1, 0.75 (mais à direita)] (adaptado de [31]). . . . .	20
2.12	Espectrograma da assinatura de Doppler de uma determinada pessoa a caminhar (Adaptado de [40]). . . . .	28
2.13	Perfis de Doppler no domínio tempo-frequência de gestos avaliados, onde é visível que cada segmento pode englobar mudanças de Doppler no lado positivo e/ou negativo do gráfico (Adaptado de [42]). . . . .	29

2.14	Variação de fase provocada pelos sinais vitais. Nas ondas de frequência mais baixa temos a respiração, onde os picos e os vales correspondem à expiração e inspiração, respetivamente, às quais se sobrepõem frequências mais altas correspondentes ao batimento cardíaco (Adaptado de [47]). . . . .	31
2.15	Periodicidade do sinal afetada pelo movimento de um membro (Adaptado de [47]). . . . .	31
3.1	Mediapipe Pose. . . . .	35
3.2	Modelo inicial. . . . .	37
3.3	Modelo estatístico. . . . .	38
3.4	Levantar dos dois braços por parte de uma pessoa, e os histogramas respetivos para o indicador da mão esquerda, e para o indicador do pé esquerdo. . . .	40
4.1	Análise ao movimento usando o algoritmo Mediapipe Pose. . . . .	44
4.2	<i>Dataset</i> de imagem. . . . .	45
4.3	Indivíduo sentado. . . . .	46
4.4	Indivíduo a levantar e baixar a mão. . . . .	46
4.5	Exemplo da construção de resultados. . . . .	46
4.6	Distribuições retiradas por articulação. . . . .	47
4.7	Em (a) temos o gráfico de pontos referente à orelha esquerda do indivíduo, em (b) referente ao cotovelo direito, em (c) ao indicador da mão direita e, por fim, o gráfico (d) correspondente ao indicador do pé esquerdo. . . . .	48
4.8	Em (a) temos o gráfico de pontos referente à orelha esquerda do indivíduo, em (b) referente ao cotovelo direito, em (c) ao indicador da mão direita e, por fim, o gráfico (d) correspondente ao indicador do pé esquerdo. . . . .	50
4.9	Gráfico de distâncias do indicador da mão esquerda, em (a); em (b) para o indicador da mão direita. . . . .	51
4.10	Gráfico de distâncias para os ombros e para os joelhos do indivíduo nas classes envolvidas. . . . .	52
4.11	Gráfico de distâncias para o nariz e para o indicador do pé direito nas classes envolvidas. . . . .	53
4.12	Deslocamentos vetoriais acima de um <i>threshold</i> definido de 5, para cada uma das articulações do Mediapipe Pose. . . . .	53
4.13	Deslocamentos vetoriais acima de um <i>threshold</i> definido de 9, para cada uma das articulações do Mediapipe Pose. . . . .	54
4.14	Em (a) e em (b), o deslocamento em $x$ e em $y$ , respetivamente, do indicador direito para uma altura em que o indivíduo está sentado. Em (c) e em (d), o deslocamento em $x$ e em $y$ , respetivamente, também do indicador direito e para o cenário em que o indivíduo está a levantar o braço direito. . . . .	56

4.15	Análise por sub-classe: em (a) e em (b), os histogramas para os vetores deslocamento no caso em que o indivíduo está sentado e no caso em que está a mexer o braço, respetivamente. Em (c) e em (d) exatamente a mesma ordem de cenários que em (a) e em (b) mas agora para os deslocamentos angulares efetuados. . . . .	58
4.16	Análise na classe completa, para o indicador esquerdo da pessoa: em (a) e em (b), os histogramas para os vetores deslocamento no caso em que o indivíduo está sentado e no caso em que está a mexer o braço, respetivamente. Em (c) e em (d) exatamente a mesma ordem de cenários que em (a) e em (b) mas agora para os deslocamentos angulares efetuados. . . . .	59
4.17	Análise na classe completa: em (a), (b) e (c), para os deslocamentos angulares efetuados do mindinho direito, do pulso direito e do cotovelo direito, respetivamente, na classe em que se está a mover o braço. . . . .	60
4.18	Histogramas dos vetores deslocamento nos dois cenários, para o cotovelo esquerdo. . . . .	61
4.19	Classificação de deslocamentos na classe de estar sentado. . . . .	64
4.20	Classificação de deslocamentos na classe de mover o braço. . . . .	65
4.21	Matriz de confusão da classificação do movimento em cada deslocamento. . . . .	66
5.1	Câmara e <i>setup</i> do rádio definido por <i>software</i> . . . . .	68
5.2	Potência de recepção de múltiplas estações de rádio FM. . . . .	69
5.3	Contextos em estudo. . . . .	69
5.4	Média das amplitudes do sinal, nas duas antenas. . . . .	72
5.5	Variância dos valores da amplitude do sinal, para as duas antenas. . . . .	73
5.6	Média dos valores de fase dos sinais das duas antenas. . . . .	74
5.7	Variância dos valores de fase, para as duas antenas. . . . .	74
5.8	Correlação cruzada entre as amplitudes do sinal da antena de referência com a antena de <i>sensing</i> , para quatro intervalos de tempo em estudo. . . . .	75

## ÍNDICE DE TABELAS

2.1	Comparação das tecnologias de deteção da atividade humana referidas. . .	32
4.1	<i>Datasets</i> trabalhados. . . . .	44
4.2	Movimentos possíveis. . . . .	45
4.3	Cor atribuída a cada ponto dos <i>plots</i> consoante a visibilidade. . . . .	47
4.4	Números associados a cada articulação. . . . .	54
4.5	Média dos histogramas de cada articulação, na classe de estar sentado. . . .	62
4.6	Média dos histogramas de cada articulação, na classe de mexer o braço. . .	62
5.1	Sub-intervalos de tempo em estudo, e identificação com um número de cada um. . . . .	71



# INTRODUÇÃO

## 1.1 Motivação

Com o constante avanço das tecnologias de telecomunicação, vários foram os dispositivos e máquinas que depressa surgiram, capazes de interagir entre si e com o ambiente envolvente. Nos últimos anos temos assistido a um maior interesse no rastreamento de movimento humano e no uso de sistemas de localização, onde se destaca a detecção por radiofrequência. Com o aparecimento daquilo a que chamamos *Internet of Things* (IoT), os sinais de Radio Frequência (RF) têm sido usados para capturar determinados eventos num ambiente IoT. Fazer o uso dos sinais RF, oferece vantagens a nível de sistemas de vigilância ou de aplicações de detecção de alvos.

As soluções baseadas em RF tornaram-se bastante atrativas devido ao baixo custo e às propriedades de fácil implementação. Os sinais RF têm atrasos de propagação bastante curtos devido à rapidez dos mesmos e ao serem transmitidos, refletidos ou espalhados por incidirem em objetos, é possível extrair informações dos alvos, tais como a posição, velocidade ou direção do movimento. Para além disso, os sensores RF destacam-se pela capacidade de penetrarem materiais e de não serem afetados por condições ambientais, exteriores ao sistema, e são extremamente precisos no que toca às medidas de detecção efetuadas. Os sinais RF são, assim, bastante usados como características da detecção, impulsionando e servindo como entrada à classificação e aprendizagem de sistemas, através de algoritmos de *Machine Learning*.

## 1.2 Objetivos

O principal objetivo deste trabalho é realizar um estudo da modelação do movimento humano, através de sinais de rádio, construindo um modelo que ao ser aplicado ao sinal de rádio, nos permita verificar padrões e associá-los aos contextos de mobilidade. Para isso, é necessário atingir os seguintes pontos:

- Definir inputs no sistema de imagem e no sistema RF, para que haja processamento de características para aprendizagem nas imagens e no sinal RF, respetivamente;

- Análise da correlação entre informação visual e RF, de forma a estudar a sua semelhança, a identificar temporizações para “tempos mortos” e a estabelecer zonas de picos/transições de atividade;
- Desenvolvimento do modelo que represente todo o conhecimento associado a cada um dos contextos a classificar e que ofereça informação útil a ser usada num ou mais métodos de estimação, necessários na previsão do tipo de contexto.

### 1.3 Contribuições

Esta dissertação vem fomentar os algoritmos de classificação de imagem, auxiliados com técnicas de *Machine Learning*, e introduzir a tecnologia das ondas de rádio, dando início a um estudo da correlação entre as duas partes.

Foi realizado um estudo à aplicação do algoritmo Mediapipe Pose a conjuntos de imagem. Foi analisada em cada contexto em questão a influência das articulações presentes à medida que paralelamente se avaliou progressivamente os resultados obtidos de acordo com aquilo que seria de esperar. Foi construído um classificador de forma a prever movimentos futuros, indicando para cada articulação do algoritmo, se determinada transição de movimento foi corretamente classificada.

Feito o estudo da imagem, introduziu-se o estudo sobre *datasets* RF, ao serem aplicadas métricas sobre os sinais obtidos por um sistema de rádio definido por *software* à medida que determina ação ia ocorrendo, o que provocou determinado comportamento nos sinais recebidos pelas antenas.

O estudo nesta dissertação abre asas a trabalhos futuros, de forma a dar continuidade ao trabalho aqui desenvolvido, quer no contexto de imagem, quer no contexto RF bem como nos dois em coexistência.

### 1.4 Estrutura do documento

O presente relatório encontra-se dividido em 6 capítulos. No Capítulo 2 é iniciado o estudo da deteção de movimento no contexto de imagem e introduzido o contexto de rádio representado pelo radar FMCW, para que no fim sejam apresentados métodos de correlação entre cada um dos contextos. O capítulo é finalizado com um conjunto de trabalhos na área, isto é, com a descrição breve de sistemas já implementados.

No Capítulo 3 é dado início à descrição do *dataset* de imagem. É descrito o que se pretendia como resultado final, como as amostras foram recolhidas, com recurso ao vídeo, assim como quais os passos que levaram à identificação de alvos em movimento e à construção do classificador final adotado.

No capítulo 4 estão evidenciados os resultados experimentais, onde uma vez aplicado o algoritmo selecionado, foi conseguido pegar em imagens e as classificar nas classes em

estudo. O classificador foi construído, e o modelo foi avaliado em termos da eficácia e precisão dos resultados obtidos.

O estudo do contexto RF deu-se no capítulo 5. Ficou descrito como as amostras foram recolhidas através do sistema implementado, e tendo em conta *datasets* semelhantes em relação ao que foi visto no contexto de imagem, a presença de movimento em imagens foi avaliada e comparada entre os diversos sinais de Rádio-Frequência.

Por fim, no capítulo 6 são apresentadas as lições retiradas desta dissertação assim como as conclusões aos resultados obtidos.

## TRABALHO RELACIONADO

### 2.1 Descrição de imagem

Atualmente, a descrição de imagem assume-se como um processo fundamental na representação e análise de vídeo digital. Cada quadro assume-se como uma forma retangular de elementos de imagem, ou pixels, e cada um desses elementos é constituído por 1 ou mais bits, de forma a termos escalas de cor e detalhe das mais simples às mais variadas. A alteração de pixels na imagem que nos está a ser transmitida é o fator crucial para a deteção e reconhecimento de formas. Um cenário em mudança pode ser entendido como uma sequência de imagens, e esse fenómeno de ação observado ao longo do tempo é devido ao movimento dos objetos da imagem, do seu observador, ou de ambos. Alterações num conjunto de imagens adjacentes no tempo, proporcionam um conjunto de características referentes ao movimento de objetos que descrevem uma determinada trajetória e é a partir desta ideia que surge a necessidade da análise e deteção do movimento humano.

Para aprender um determinado cenário presente numa imagem, podem ser aplicados modelos de mistura Gaussianos no espaço RGB, mas outros espaços de cor como HSV ou YUV funcionam bem como alternativas [2] [3]. Modelos de mistura gaussianos são modelos probabilísticos que fazem com que numa determinada imagem possam ser formados conjuntos de dados com base nos atributos de semelhança que apresentam. Desta forma constituem um fator importante na caracterização de ambientes, assim como na criação de modelos de cor e estudo do movimento humano, abordados nos capítulos seguintes.

#### 2.1.1 Deteção de movimento em imagens

A deteção de humanos em sequências de imagens é parte fundamental para monitorizarmos a atividade e comportamento normal de uma pessoa ou outra forma em movimento no dia a dia, algo que é imprescindível em sistemas de vigilância, sistemas de segurança em carros ou até mesmo em sistemas de interação humano-robô.

Em [4] e a fim de partir para o estudo do movimento de uma pessoa, uma análise completa passa por alguns passos:

- Inicialização: Assegura uma correta interpretação do cenário em consideração;
- Rastreamento: Estabelece relações coerentes na pessoa, assim como dos respectivos membros que a compõem ao longo de várias imagens (*frames*);
- Estimação de pose: Responsável por identificar como o corpo humano e/ou os membros individuais estão configurados;
- Reconhecimento: Identifica o movimento capturado e procede à sua classificação.

Detetar movimento em imagens, ou *tracking*, pode ser visto como uma etapa preliminar da estimação de pose e também do reconhecimento do próprio movimento. Fazer o rastreamento a partir da análise visual do movimento humano origina a junção de três processos:

- Segmentação figura-fundo: É o ponto de partida da maioria dos algoritmos de seguimento e consiste em separar os objetos de interesse (humanos) do resto da imagem (plano de fundo). Para deteção de movimento em imagens, podemos ter que considerar vários cenários dependendo da estaticidade da câmara, de termos um ou mais objetos de interesse ou ainda do plano de fundo ser ou não constante;
- Representação: As entidades segmentadas têm de ser descritas por uma representação conveniente. Uma representação baseada em objetos pode ser realizada por: uma representação por pontos; uma representação através de caixas em que o sujeito é representado por caixas fechadas que contêm os píxeis ou regiões identificadas com o processo da segmentação figura-fundo; uma representação de forma obtida através de métodos de subtração de uma *frame* atual com uma *frame* anterior, utilizados de igual forma na segmentação figura-fundo; uma representação do tipo bolha que segue a segmentação figura-fundo. Um indivíduo é representado por uma bolha ou por um conjunto de bolhas que partilham algumas características. Estas semelhanças podem ser a nível de fluxo, de cores ou ambos.
- Correspondências temporais: Dado o estado de várias pessoas em frames anteriores, o objetivo é o de determinar qual o estado das mesmas na frame atual. Esta análise de correspondência está muitas vezes interligada com a previsão, o que envolve um possível conhecimento avançado sobre o estado do objeto na *frame* seguinte, o qual é comparado com o estado na *frame* atual.

Entre outras formas, é permitido fazer a segmentação figura-fundo com base no movimento. A solução ideal passa por encontrar padrões de movimento semelhantes aos de um humano, em vez de padrões de aparência, rejeitados pela hipótese de haver características correspondentes a posições específicas do corpo.

Com o que foi referido no parágrafo anterior, a técnica de *Support Vector Machines* pode ser utilizada para constituir classificadores binários treinados, para determinar se

uma dada região numa imagem corresponde a um movimento humano. A função de classificação de padrões pode ser aprendida através de uma função de base radial, definindo um hiperplano ao longo do espaço de características. De acordo com a parametrização desta última e recorrendo-se a misturas de modelos Gaussianos, temos a formação de clusters de imagem que permite o agrupamento de características comuns presentes. Os pesos da função de classificação são selecionados, para que então possamos ter menos classificações incorretas no que toca a ver se determinada zona da imagem pertence ou não a uma pessoa.

### 2.1.2 Computação de vetores de movimento

É entendido como campo de movimento uma matriz de vetores 2D, representando o movimento de pontos numa cena 3D. Um vetor de movimento é formado desde o ponto em 3D com imagem no tempo  $t$  até ao ponto em 3D com imagem no tempo  $t + \delta$ . A cada vetor de movimento pode estar associada uma velocidade instantânea no tempo  $t$ .

O fluxo de imagem corresponde ao campo de movimento calculado, considerando que a intensidade da imagem perto dos pontos correspondentes é relativamente constante. O processamento de fluxo de imagem está presente em alguns métodos referenciados de seguida, com a finalidade de se fazer a deteção daquilo que são as mudanças a acontecer num espaço inicial captado por uma determinada câmara.

#### 2.1.2.1 Uso de pontos de correspondência

Segundo [5], um campo de movimento disperso pode ser calculado identificando e distinguindo-se os pares de pontos que correspondem em duas imagens tirados nos instantes  $t$  e  $t + \delta$ . Pontos pertencentes aos cantos do objeto serão bem sucedidos para o seguimento de imagens tanto monocromáticas como de cor, e são chamados de pontos de alto interesse. Um operador de interesse centrado num píxel  $P$ , vai calcular as variâncias de intensidade para a vizinhança, mais concretamente nas direções vertical, horizontal ou diagonal. Se o mínimo das variações ultrapassar um determinado limiar de decisão, então  $P$  é tornado num ponto de interesse. Uma vez detetado o conjunto destes pontos na imagem 1 no tempo  $t$ , cabe por fim ver quais os pontos a que correspondem na imagem 2 no tempo  $t + \delta$ . Este método está exemplificado na Figura 2.1, em que a partir de um ponto de interesse  $(T_x, T_y)$  na imagem 1 dentro de uma determinada região, irá ser procurado na mesma zona da imagem 2 a melhor correspondência para uma pequena vizinhança de  $(T_x, T_y)$ , onde irá estar a cabeça  $(H_x, H_y)$  do nosso vetor de movimento.

#### 2.1.2.2 Imagem da história do movimento

Em [6], é proposto o uso de *Motion History Images*, que se traduzem em Imagens da História do Movimento (MHI). Trata-se de um método para análise de movimento em imagens, em que à medida que novas frames vão chegando, estas vão-se sobrepôr às

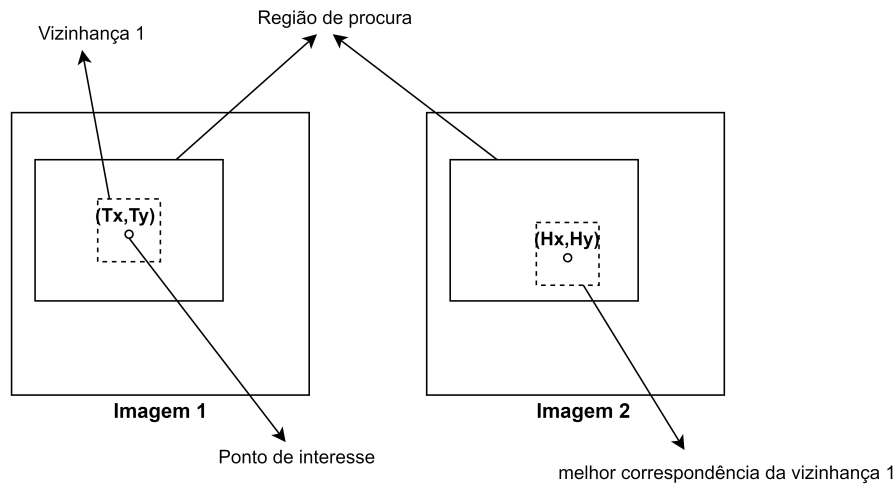


Figura 2.1: Ilustração do uso de pontos de correspondência (Fonte Própria).

formas de imagem já existentes cujos valores de brilho vão diminuindo. As representações por MHI conseguem codificar intervalos entre dois quadros espaçados no tempo, numa única imagem. Assim qualquer gesto que uma pessoa tenha na imagem terá escalas de tempo abrangidas por MHI.

Com este processo, é possível codificar de forma direta um ponto no espaço de imagem ao que chamamos Imagem da História do Movimento cronometrada (tMHI). Um gradiente da tMHI é usado para determinar o fluxo óptico normal da imagem, o que leva à obtenção de vetores de direção que apontam no sentido do movimento. Estes irão apontar numa direção ortogonal aos contornos da parte do corpo que está a mexer em cada instante da tMHI, o que faz com que haja um fluxo de movimento presente.

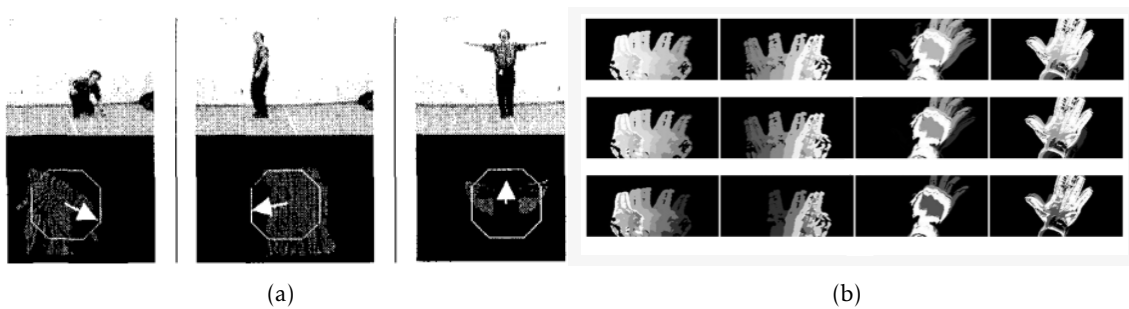


Figura 2.2: Técnica de Imagens da História do Movimento para detecção do movimento humano. (Adaptado de [6]).

### 2.1.2.3 Aplicação de filtros de movimento em pares de imagens no tempo

Em [7] é realizado o estudo sobre a deteção de pedestres recorrendo-se a padrões de movimento. Neste método, em pares de imagens no tempo, é extraída informação sobre um determinado deslocamento através da diferença entre as mesmas. Regiões onde o valor absoluto da diferença é elevado correspondem a movimento.

Dados sobre a direção do movimento podem ser retirados a partir da diferença entre a primeira imagem e versões deslocadas da segunda imagem.

$$\Delta = abs(I_t - I_{t+1}) \quad (2.1)$$

$$U = abs(I_t - I_{t+1} \uparrow) \quad (2.2)$$

$$L = abs(I_t - I_{t+1} \leftarrow) \quad (2.3)$$

$$R = abs(I_t - I_{t+1} \rightarrow) \quad (2.4)$$

$$D = abs(I_t - I_{t+1} \downarrow) \quad (2.5)$$

As equações 2.1 à 2.5 originam cada uma a sua imagem resultante, e aquela cuja energia for menor, ou seja, onde for menos perceptível a presença de pixels verdadeiros na imagem (imagem  $\Delta$ , U, L, R ou D), dá-nos a direção tomada pelo movimento.

#### 2.1.2.4 Imagem de dupla diferença

Segundo [8], pode-se obter movimento em regiões captadas de *video frames* através de imagens de dupla diferença, assim como estimar a pose de forma parcial nessa região, descrito de seguida na Secção 2.1.3 do presente documento.

Uma imagem de dupla diferença é originada a partir de 3 *frames* consecutivas numa *stream* de vídeo. Considerando os pares de imagens ' $t - 1$ ' e ' $t$ ', ' $t$ ' e ' $t + 1$ ', é realizada a diferença entre os mesmos pelo que se obtém duas imagens de diferença. De seguida, estas vão ser binarizadas. Este método surge como uma solução ao *ghosting effect*, que se traduz num fenómeno de erro provocado por uma resposta lenta no tempo, isto é, ao executar-se a diferença entre duas *frames* consecutivas, a *frame* inicial ainda não desvaneceu completamente, efeito que vai provocar uma diminuição na eficácia com que se deteta o movimento. Contudo, uma imagem de dupla diferença ao ser uma multiplicação de duas imagens, vai incluir pixels isolados com ruído o que afecta a deteção de movimento. Desta forma, cada 4 por 4 pixels na imagem de dupla diferença vão ser agrupados em blocos. Se mais de metade dos pixels de um bloco forem verdadeiros (marcados com o valor 1), então o bloco é verdadeiro. Este procedimento permite evitar ruído e torna o processamento de imagem mais leve.

#### 2.1.3 Estimação de pose e representação do movimento

A estimação de pose é fundamental para se gerar a forma do objeto, neste caso de uma pessoa, para que o estudo sobre o movimento seja totalmente eficaz. Estimar a configuração da estrutura de articulações subjacentes a um determinado indivíduo, permite depois estudar a respetiva cinemática, presente nos métodos anteriormente referidos. Uma vez retirados e preparados os dados através do *tracking*, a estimação de pose encarrega-se por extrair informação específica de imagem a partir dos mesmos, desde os contornos de uma pessoa às mãos ou cabeça.

O uso de um modelo humano ou de características que se possam assemelhar ao mesmo é, aqui, o fator crucial para a construção da pose uma vez que o sistema é direcionado para análise sobre uma pessoa, o que requer a incorporação de conhecimento sobre humanos no seu processo. No estudo realizado em [4], a estimação de pose é dividida em 3 classes principais, distinguidas pela existência ou não de modelos *a priori*, isto é, a criação de modelos deduzidos por outros anteriores.

#### 2.1.3.1 *Model Free*

Neste modelo não são usados modelos *a priori*. Um ponto de partida ao trabalho aqui envolvido, relaciona-se com a procura do formato corporal através de *blobs*, contrastivas com o uso de características como pontos, linhas ou contornos. O mesmo termo, segundo [9], consiste em considerar vetores de píxels com as respetivas coordenadas espaciais, o que conduz à formação de clusters com base no nível espectral e espacial. Um humano é modelado como um conjunto conectado de bolhas, com distribuições gaussianas individualmente atribuídas assim como um mapa de suporte que indica os píxels pertencentes a cada bolha.

Construções probabilísticas de partes do corpo foram introduzidas para uma estimação de pose baixo-cima a duas dimensões. Começando primeiro por detetar as suas localizações prováveis, as suas interligações permitiam, posteriormente, obter os formatos que melhor correspondiam às observações. Aprender o subespaço de configurações de um corpo possibilita fazer o mapeamento das características visuais em configurações do corpo inteiro [10]. Por aprendizagem [11], são feitas aproximações estatísticas através da reconstrução de movimentos 3D de um indivíduo, a partir de sequências de imagem monoculares. Através de uma única câmara que capta vídeo a duas dimensões, rastreando o movimento de posições do corpo na imagem ao longo do tempo, as reconstruções 3D são inferidas dessas mesmas posições, e comparadas com um conjunto de treino de movimentos humanos 3D para determinar aquelas que são plausíveis.

Partindo do que foi mencionado no parágrafo anterior, um modelo oculto de *Markov* modela a variedade de características que pode ter o sistema dinâmico referente ao estudo do movimento de partes do corpo, isto é, tendo em conta trajetórias observadas e as configurações e reconstruções anteriormente faladas, o modelo oculto de Markov oferece-nos a probabilidade de termos determinada pose numa *frame* futura [12].

#### 2.1.3.2 **Uso indireto de modelo**

Os métodos nesta secção utilizam um modelo *a priori* para estimação da pose de um sujeito. O mesmo é tido como referência ou tabela de procura, com informação relevante que vai ajudar na interpretação dos dados medidos.

Foi visto anteriormente, que a segmentação de determinado alvo de estudo pode ser realizada com base na recuperação do plano de fundo inicial. São então geradas formas de pessoas que podem servir como *input* à localização de partes do corpo [13]. Dito por

outras palavras, a pose geral ou total de um indivíduo é calculada, lançando um ponto de partida à descoberta de poses mais específicas. Com o avançar do dinamismo presente nas restrições do seguimento em imagens, uma filtragem recursiva de valores de píxel *raw*, foi aproveitada para dar origem a um modelo de pose dinâmico formado por bolhas [14].

Trabalhos mais recentes, têm em conta as dependências espaciais entre partes do corpo. Estes modelos são baseados noutros modelos gráficos baseados em árvore [15], que codificam estas relações no espaço entre partes adjacentes de uma determinada pessoa, e que seguem uma certa cadeia cinemática. Existindo uma ampla variação no que toca à aparência humana devido à pose, ou à sua não linearidade resultante da existência de roupa ou de condições de imagem, foram introduzidos os chamados modelos de estrutura pictórica [16]. Os mesmos representam a aparência de cada parte, separadamente, e capturam explicitamente configurações espaciais de partes que são independentes da aparência.

### 2.1.3.3 Uso direto de modelo

No uso direto de um modelo a priori para a estimação de pose, este representa o sujeito observado, ao estar constantemente em atualização devido a novas observações. Os modelos são bastante detalhados [4], facilitando a capacidade de se lidar com uma possível oclusão de imagem ou ainda a introdução de restrições a nível cinemático no sistema. A filosofia seguida consiste em prever a pose do modelo que vai estar presente na imagem seguinte (modelo sintetizado). Diversas tentativas de estimação de pose são realizadas e, através de medidas de semelhança, a pose correta é encontrada e usada para atualizar o modelo anterior.

Enquanto que a estimação de pose a 2 dimensões já possui traços bem definidos, o mesmo não se pode dizer da estimação de pose a três dimensões. Em [17], é proposto um modelo hierárquico flexível para estimação de pose 3D, onde o tronco de uma pessoa corresponde ao nível superior, os membros superiores pertencem ao nível seguinte que depende do anterior, seguindo-se os membros inferiores. A pose vai depender de diferentes parâmetros, como as juntas (locais de ligação entre os dois ossos) ou o tamanho dos ossos.

## 2.2 Radar FMCW

Os dispositivos de radar tornaram-se comuns nos países desenvolvidos. Estes baseiam-se num sinal que é emitido por uma antena transmissora, por normal ao encontro de um alvo pretendido, que reflete parte da radiação eletromagnética de volta para uma antena recetora. O tempo de propagação destes sinais, entre o emissor e o recetor, é bastante pequeno - os mesmos transmitem-se à velocidade da luz. As diferenças que se verificam na onda recebida face à transmitida estabelecem-se na ordem dos picossegundos, e a

técnica do FMCW procura mapear essas mudanças na frequência portadora do sinal transmitido através do espectro do sinal recebido. Através do sinal de radiofrequência transmitido, é possível calcular a distância a que o alvo se encontra numa determinada área, assim como a direção ou velocidade a que se move.

Os radares FMCW (*Frequency Modulated Continuous Wave*) surgiram, em grande parte, por se tratarem de uma implementação de baixo custo e por terem um reduzido consumo a nível energético [18] [19]. Devido ao elevado comprimento de onda que as ondas emitidas por um radar FMCW podem ter, este opera bem sob um tempo ou condições de luminosidade variáveis, fator que torna o espectro RF bastante atrativo para o *sensing* em aplicações e sistemas interativos.

### 2.2.1 Arquitetura de *sensing*

Um radar FMCW transmite uma portadora contínua, que vai ser modulada por uma função periódica, sinusoidal ou dente de serra, de forma a fornecer dados sobre um alvo a detetar.

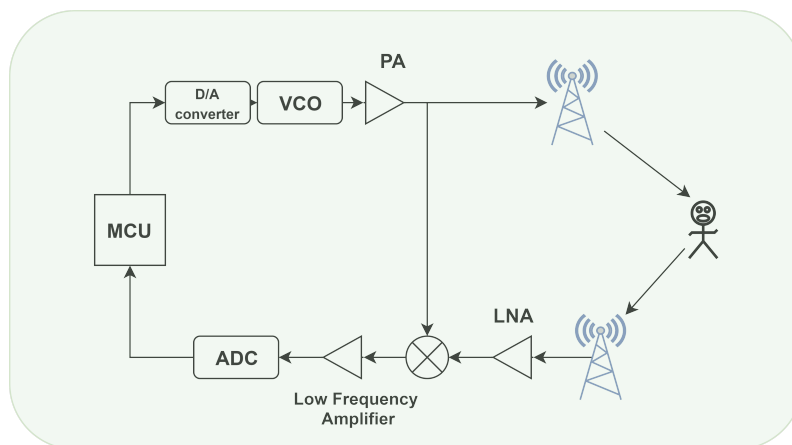


Figura 2.3: Diagrama de blocos de um radar FMCW simples (Fonte Própria).

A Figura 2.3 ilustra uma representação do radar estudado. O oscilador de tensão controlada (VCO), em conjunto com o conversor digital-analógico, que gerou a entrada em VCO, originam o sinal que vai ser transmitido pela antena de transmissão. Antes de ser transmitido, o sinal passa ainda por um amplificador de potência (PA) com um ganho associado pré-estabelecido. Na chegada do sinal refletido pelo alvo, à antena recetora, o mesmo é amplificado pelo LNA (*Low Noise Amplifier*) e vai entrar num *mixer*, numa multiplicação no tempo com o sinal de referência (sinal antes de ser emitido pela antena de transmissão). O sinal de alta frequência resultante é submetido a um filtro passa-baixo, antes de ser convertido em sinal digital e interpretado por uma unidade de controlo.

O hardware do sistema apresentado, pode ser visto de forma mais detalhada. No trabalho desenvolvido por [20] para deteção de alvos humanos, foi utilizado um radar FMCW com uma banda de rádio ISM de 24 GHz. Os blocos *D/A converter*, VCO e PA, correspondem a um circuito integrado de microondas. O oscilador de tensão controlada

está ligado a dois *prescalers*, com frequências de output de 23-kHz e 1.5 GHz, que fazem reduzir a frequência do sinal produzido pelo conversor digital-analógico. Ainda nesta fase, um *Phase-Locked Loop* é usado para estabilizar o VCO na frequência de 24 GHz do circuito e gerar as rampas de frequência para detecção do alvo. Por fim, temos uma secção de banda base analógica, constituída por um amplificador de baixo ruído, pois o alvo pode provocar um sinal muito baixo em amplitude, e ainda o processamento do sinal, que no nosso diagrama de blocos corresponde ao ADC e à unidade de controlo.

### 2.2.2 Processamento de sinal e *background* matemático

Para introduzirmos a análise aos sinais envolvidos num radar FMCW, a figura 2.4 mostra-nos as frequências dos mesmos temporalmente, considerando o tratamento de sinal a nível de hardware referido na secção anterior tal como um efeito de Doppler nulo.

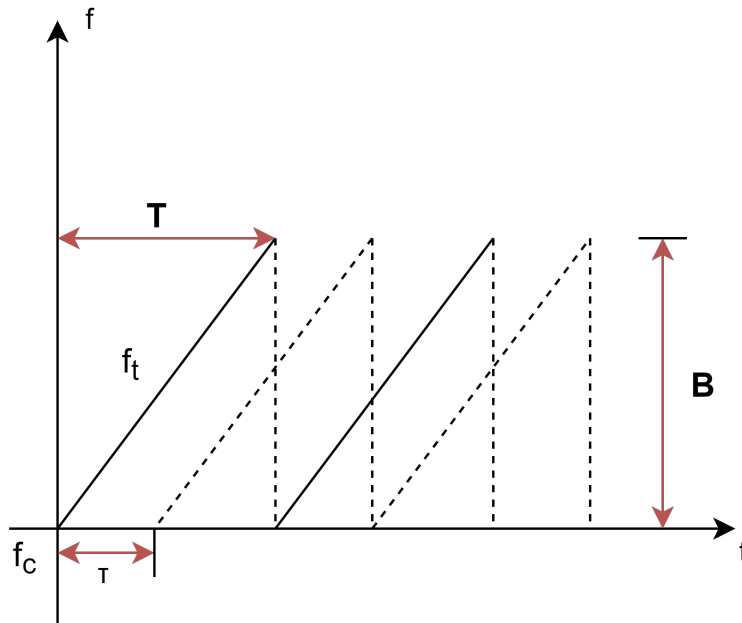


Figura 2.4: Transmissão do sinal por um radar FMCW (Fonte Própria).

As linhas sólidas correspondem aos gráficos dos sinais transmitidos e as linhas a tracejado aos sinais recebidos. A frequência do *chirp* (sinal transmitido) é dada, a qualquer momento, por:

$$f_t = f_c + B * \frac{t}{T}, \quad (2.6)$$

onde  $f_c$  representa a frequência da portadora, B a largura de banda do sinal e T o período do *chirp* [21]. O sinal emitido vai embater num alvo, é refletido, e parte do que foi refletido vai chegar à antena recetora com um atraso do tempo de propagação equivalente a  $\tau$  dado por:

$$\tau = \frac{2R}{c}. \quad (2.7)$$

Na equação anterior,  $R$  consiste no alcance do objeto, sendo que  $c$  corresponde à velocidade da luz. A resolução de alcance do radar FMCW é descrito através de:

$$\Delta R = \frac{c}{2B} \quad (2.8)$$

Assumindo que  $\Delta f$  representa a diferença entre as frequências dos sinais recebido e transmitido, o alcance do objeto é calculado a partir de:

$$R = \frac{c\Delta f T}{2B}. \quad (2.9)$$

A frequência *beat* do sinal,  $\Delta f$ , calcula-se a partir da diferença da frequência da onda emitida e da onda refletida.

### 2.2.2.1 Transformada rápida de Fourier e efeito de Doppler

De forma a ser realizada a desmodulação do sinal recebido, é realizado, sobre o mesmo, o processamento de uma transformada rápida de Fourier complexa (FFT), de forma a que o sinal no domínio do tempo passe para o domínio da frequência, ou seja, corresponde a um algoritmo rápido de aplicação da transformada de Fourier discreta em amostras no domínio do tempo. Ela é aplicada em cada chirp recebido, e é tipicamente usada para estimar diversos parâmetros como o alcance ou velocidade do alvo, ou ainda o ângulo de incidência do sinal [22]. A dimensão da matriz FFT depende, de forma proporcional, do número de parâmetros a serem determinados, e a complexidade aumenta consoante uma maior dimensionalidade da FFT.

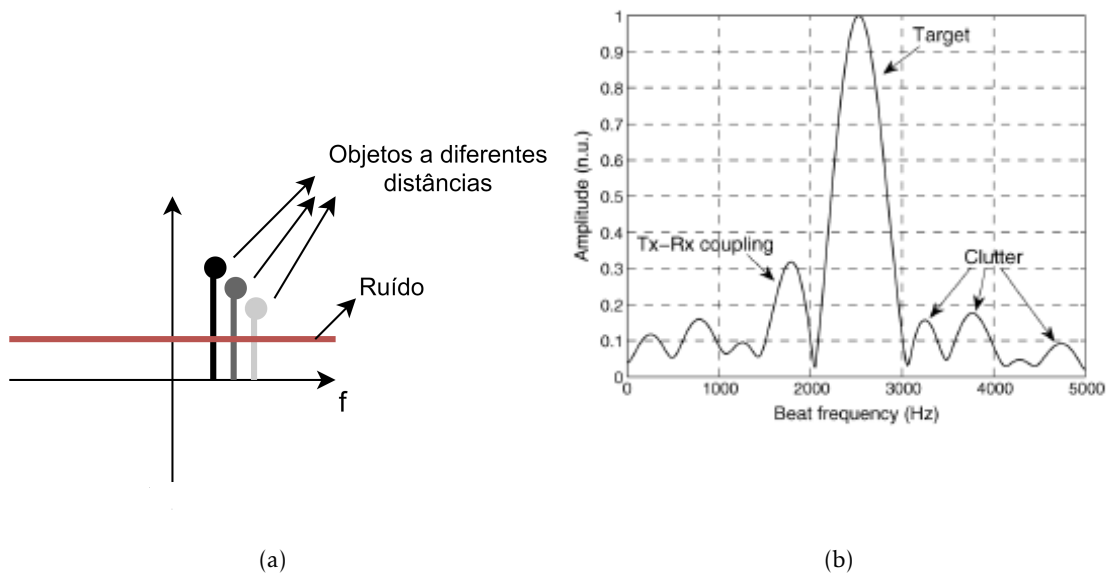


Figura 2.5: (a) Sinais recebidos por um radar FMCW e espectro das respectivas frequências *beat* (Adaptado de [23]). (b) Frequência *beat* de referência detetada no processo de calibração da deteção de um alvo (Adaptado de [24]).

Existe sempre uma fase associada à frequência *beat* do sinal recebido, que muda linearmente com o número de amostras do sinal transmitido. Uma mudança na fase indica como a frequência do sinal muda em consequência do número de períodos. No sistema de um radar FMCW um "atraso" equivale a uma "mudança de frequência".

Para pessoas que estejam em movimento, a frequência *beat* do sinal recebido abrange uma componente de Doppler. Esta depende da velocidade relativa entre o radar e o alvo. Àquilo a que podemos denominar por um radar Doppler, estes podem estar sujeitos a problemas, como a existência de *offsets* DC ou de desequilíbrios numa implementação em fase/quadratura. No entanto, os mesmos radares funcionam no princípio de que a obtenção do alcance do alvo é baseada em medições de fase, o que confere bastante precisão ao processo [24]. Nos radares Doppler, monitorizar o movimento de um indivíduo assenta na aplicação de uma FFT para retirar informações de fase.

### 2.2.2.2 Extração do sinal baseada em fase

Na figura 2.6 está representada, no gráfico superior, uma cadeia de transmissão modulada por uma onda dente de serra no domínio frequência-tempo. No gráfico inferior são apresentados os sinais *beat* recebidos de um alvo encontrado.

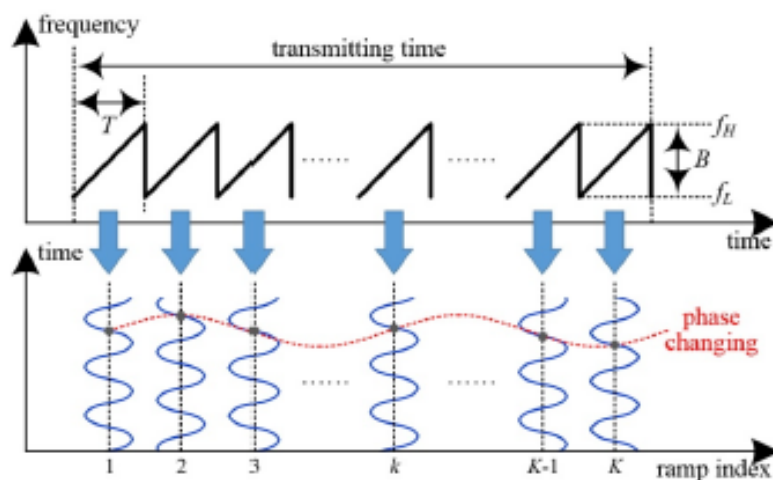


Figura 2.6: Em cima o sinal transmitido no domínio frequência-tempo, em baixo o sinal *beat* recebido de um alvo em movimento (Adaptado de [25]).

A figura acima representa uma técnica do radar FMCW Doppler-alcance 2D baseada em rampa rápida.

É visível pelo gráfico do sinal recebido, que a velocidade radial e o alcance permanecem idênticos para todas as rampas, devido ao tempo de transmissão ser muito inferior ao tempo de movimento do alvo. As informações de velocidade são retiradas das mudanças de frequência, devido ao efeito de Doppler. Pela linha vermelha a tracejado no gráfico das ondas recebidas, é observável que se tem uma mudança de fase constante, provocada pela frequência de Doppler do sinal.

A transformada rápida de Fourier que é aplicada sobre o sinal recebido, vai ter  $M$  amostras se o sinal recebido for repartido em sub-janelas temporais, e cada uma dessas amostras vai ter  $N$  pontos. A fase em cada conjunto é extraída e "desembrulhada" por:

$$\varphi_m(n) = \begin{cases} \varphi_m(n) - 2\pi & \varphi_m(n) - \varphi_m(n+1) > \pi, \\ \varphi_m(n) + 2\pi & \varphi_m(n) - \varphi_m(n+1) < -\pi \end{cases} \quad (2.10)$$

onde  $n \in [1, N-1]$  corresponde ao ponto de amostragem.

Podem ocorrer alguns problemas de interferência ao nível da antena de transmissão. É muitas vezes uma auto-interferência, causada por falha de isolamento do sinal nos caminhos de transmissão ou de receção (obstáculos à passagem do sinal). A mesma interferência pode ter origem na eletrónica analógica associada, quando nas rampas de transmissão do sinal temos a passagem repentina da frequência mais alta para a mais baixa [25], tal como é visível na figura 2.5. Em [26] é proposto um modelo de erro que procura solucionar imprecisões nas medidas de alcance assim como falsas deteções de alvo. É criada uma nova cadeia de processamento do sinal, que permite detetar frequências de vibração muito mais altas que a taxa de *chirp* normal e então separá-las, uma vez que estão a causar interferência.

### 2.2.3 Implementação de um radar FMCW

#### 2.2.3.1 Implementação em banda base real

Muitas das implementações destes radares usam o diagrama de blocos que foi visto na Figura 2.4 deste documento. São compostas por um único *mixer* real e por uma única cadeia ADC. O módulo do oscilador local gera um sinal linear de onda contínua modulado na frequência,  $\cos(\phi(t))$ , que passa pelo amplificador de potência antes de ser transmitido. O sinal recebido, é submetido ao *mixer* com o sinal que havia sido gerado pelo oscilador antes da transmissão, de forma a produzirem a frequência *beat* de output que passa pelo ADC antes de ser processada.

A Figura 2.7 ilustra o espectro de amplitude dos sinais de transmissão (LO), receção (RX) e da frequência *beat* (IF), numa implementação em banda base real [23].

No espectro RX, o sinal de interesse está contido na parte "*In band*" do mesmo, enquanto que a porção "*Image Band*" é desprovida de qualquer sinal de interesse. Esta divisão relaciona-se com o facto de o sinal recebido estar sempre atrasado em relação ao sinal transmitido e, este ruído provocado pela imagem de banda do espectro, vai estar presente no espectro do sinal IF.

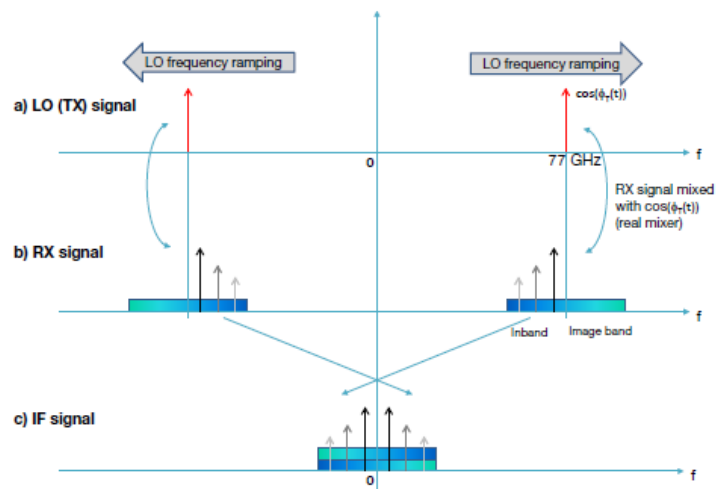


Figura 2.7: (a) Espectro do sinal modulado LO, representando a frequência instantânea de uma rampa do sinal. (b) Espectro do sinal de recepção, que sofreu reflexão noutros objetos. (c) Sinal IF produzido à saída do *mixer* (Adaptado de [23]).

### 2.2.3.2 Implementação em banda base complexa

Apesar de muitos dos sistemas de radares FMCW existentes usarem recetores em banda base real, verificou-se que apenas uma parte do espectro do sinal recebido continha informação útil, e nada mais podia ser aproveitado do resto (banda de imagem). Em consequência, a arquitetura de um radar FMCW também pode ser feita recorrendo a 2 *mixers*, um em fase e outro em quadratura, para fazer a desmodulação do sinal recebido. Um esquema tal como é mostrado na figura 2.8, permite extrair a fase e a amplitude instantânea do sinal no domínio do tempo.

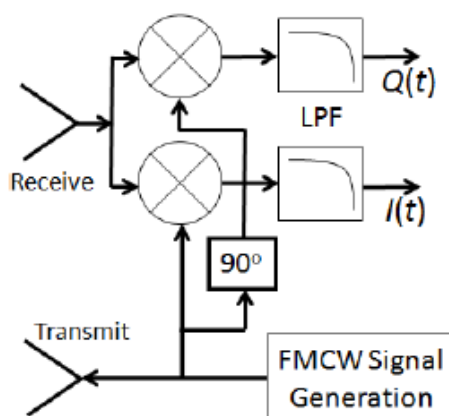


Figura 2.8: Diagrama de blocos de um radar FMCW em banda base complexa (Adaptado de [27]).

Na Figura 2.9, considera-se que o sinal em quadratura representa uma versão sin do sinal gerado pelo oscilador do FMCW, enquanto que o sinal em fase corresponde a uma versão cos e, portanto, teremos 2 cadeias ADC no tratamento do sinal recebido.

Foi visto que numa implementação em banda base real não havia qualquer sinal útil de reflexão do alvo, num dos lados do espectro. Ao ser utilizada uma arquitetura de receção em banda base complexa, as frequências IF de output, à saída do *mixer*, conseguem ser separáveis, pelo que a parte da banda de imagem do espectro sem qualquer sinal de interesse consegue facilmente ser detetada. Desta forma, o benefício mais claro que uma arquitetura em banda base complexa oferece é o melhoramento da figura-ruído do sinal, isto é, eliminando o ruído que existia numa implementação em banda base real.

A Figura 2.9 mostra o espectro das diferentes fases pelas quais um sinal passa, numa implementação em banda base complexa.

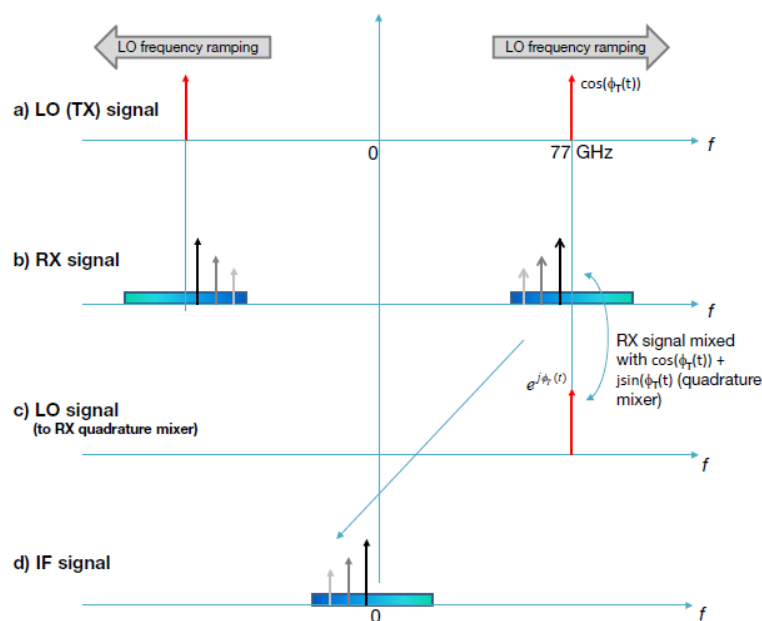


Figura 2.9: (a) Espectro do sinal modulado LO, representando a frequência instantânea de uma rampa do sinal. (b) Espectro do sinal de receção, que sofreu reflexão noutros objetos. (c) Sinal que entra no *mixer* em quadratura. (d) Sinal IF que sai do *mixer* em quadratura (Adaptado de [23]).

Com uma banda base complexa, é possível fazer o monitoramento da banda de imagem do espectro, com o intuito de se detetar interferência e/ou medir o nível de ruído térmico de forma eficaz [23]. Por exemplo, picos de energia na banda de imagem analisada, podem provir de um outro sistema de radar interferente.

Por outro lado, uma duplicação da cadeia de ADC para suportar uma implementação em banda base complexa não provoca um aumento da carga na memória do sistema nem da complexidade a nível do processamento do sinal. Visualizando os gráficos das figuras 2.9c e 2.9d, os mesmos correspondem aos *outputs* dos ADCs, numa implementação em banda base real e em banda base complexa, respetivamente. Na implementação real, é constatável que as amostras vão para processamento a uma taxa de duas vezes a frequência de *beat* máxima. Numa arquitetura complexa não é preciso dobrar o ritmo e ainda é possível diminuí-lo. Como a banda de imagem do sinal pode ser detetada e eliminada,

a largura de banda do espectro é reduzida a metade, e as amostras dos canais em quadratura e em fase podem ser enviadas para serem processadas, a uma taxa equivalente à frequência de *beat* máxima.

## 2.3 Séries temporais e análise de características de classificação

No mundo que nos rodeia, o passar do tempo é o fator que limita a nossa experiência, e estudar as suas dinâmicas e implicações nos diferentes processos do mundo real, está presente no nosso quotidiano. As mesmas dinâmicas capturadas através de medições sucessivas ao longo do tempo originam séries temporais, e o estudo destas permite levar a uma vasta gama de características resultantes e que se pretendem obter.

Cada série temporal vai equivaler a um grande número de pontos medidos que podem ser bastante diversos, demonstrando toda a dinâmica envolvida, o que leva à tentativa de perceber quais os padrões que se podem retirar de forma à compreensão dos processos dinâmicos subjacentes. Em conjuntos de dados não trabalhados e ordenados, é fácil a obtenção de perguntas face a uma amostra retirada. Contudo, são mais frequentes e interessantes as dúvidas relativamente à série temporal inteira e quais as dependências temporais existentes, como verificar como um primeiro valor se correlaciona com um segundo valor no futuro. A partir de dados de uma série temporal, com determinada forma gráfica, diferentes classes de métodos são introduzidas de seguida, como a estacionariedade dos valores ou a entropia, constituindo diversas métricas de caracterização que obtêm as propriedades da série temporal, através de características de classificação.

### 2.3.1 Representação de séries temporais e características de classificação

Representar uma série temporal através de características de classificação, parte da ideia da definição de uma medida de semelhança entre pares de séries. Quando uma nova série surge, ela é classificada pela sua correspondência com a série temporal de treino mais semelhante, que infere qual a classe do novo conjunto de pontos. No entanto, não existe uma "melhor" igualdade entre séries. Uma sucessão de valores medidos no tempo pode ser semelhante com um conteúdo em frequência semelhante, mas se considerarmos o estudo de *burst rates*, duas séries análogas são aquelas com taxas idênticas. Se avaliarmos duas séries apenas pelos valores parecidos nos mesmos instantes de tempo, a semelhança entre as mesmas é quantificável com recurso a uma simples métrica como a distância Euclidiana entre os vetores das duas séries temporais.

Partindo do que foi mencionado no parágrafo anterior, são usadas métricas sobre aquilo que se retira fisicamente, que por sua vez geram *features* sob o quadro temporal retirado. Se forem retiradas duas variáveis aleatórias no tempo a distância entre os mesmos é denominada como distância estatística. Esta pode ser inclusivamente aplicada entre um ponto individual de uma amostra e um conjunto de outros pontos, e é um termo válido

e usado como métrica se e só se a distância considerada for positiva em módulo, ou nula caso os pontos coincidam.

Para séries temporais de comprimento igual, uma medida de semelhança entre séries, captura o quão semelhantes são os dois conjuntos, durante o determinado período tempo. O mesmo é demonstrado na figura 2.10-A. Uma aproximação à classificação de séries temporais tendo em conta as características de classificação, está representada na figura 2.10-B.

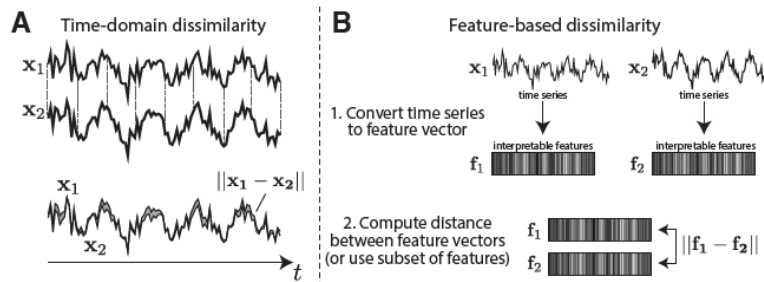


Figura 2.10: Medidas de semelhança de séries temporais, no domínio do tempo, com base em características de classificação (Adaptado de [28]).

A representação baseada em características de classificação de uma sequência de valores que partem de *inputs* no domínio do tempo, permite ter um bom conhecimento acerca das propriedades dos dados medidos que são úteis para conjuntos futuros, tal como tinha sido referido. É essencial referir como a representação é construída, assim como a metodologia usada na seleção das características [29].

É importante ter alguns *datasets* previamente construídos através de experiências anteriores, seja de imagem ou de sinal de rádio, de forma a aplicarem-se as distâncias estatísticas já referidas. Os conjuntos resultantes da computação de alguns parâmetros dos *datasets*, incluem dados com as características de classificação mais relevantes de uma série no tempo, e representam assim um grupo finito de *inputs*, num algoritmo de *clustering*, que vai agrupar a nova série temporal em estudo na classe das *features* mais importantes a retirar da mesma. Os conjuntos de características previamente processados a partir dos *datasets*, são aplicados para que haja uma classificação e são menos sensíveis ao ruído, daí ser eficiente o uso dos mesmos [30].

### 2.3.1.1 Tipos de distâncias usadas

Tal como foi introduzido, são utilizadas distâncias e medidas de semelhança entre 2 séries temporais, de forma a extrair informação e propriedades da imagem. Primeiramente, estas distâncias partem dos requisitos que são sempre superiores ou iguais a zero, sendo que é nula se os dois pontos medidos corresponderem ao mesmo, e que a distância num dos sentidos é obviamente simétrica à distância no sentido inverso. Os diferentes métodos que não cumpram com as condições anteriores, são apenas vistos como medidas de semelhança, e não como uma distância. O conjunto de distâncias usadas entre imagens,

são divididas em distâncias de poder, distâncias nas leis de distribuição e distâncias de correlação [31].

### 1. Distâncias de expoente

Medem a distância entre 2 vetores  $x$  e  $y$  de comprimento  $n$ , pela fórmula:

$$dp = \sum_{n=1}^n |x_i - y_i|^p)^{\frac{1}{r}}, \quad (2.11)$$

em que o valor dos parâmetros  $p$  e  $r$ , nos dão a métrica que está a ser utilizada para extração de características. A título exemplificativo, quando os dois parâmetros tomam o valor 2, a métrica utilizada corresponde à distância Euclidiana.

A Figura 2.11 mostra um diagrama de Voronoy, que é construído com base na distância de expoente, e consoante se varia os parâmetros mencionados. O tom de sombreamento de uma dada região, representa o conjunto de pontos para os quais o ponto de origem inserido naquela região está mais perto que qualquer um dos outros, localizados nas restantes regiões.

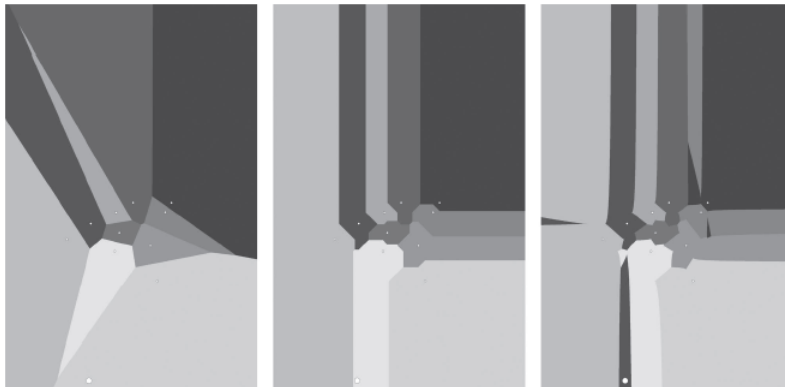


Figura 2.11: Diagrama de Voronoy. Distâncias de poder com  $p = r = [2$  (mais à esquerda),  $1, 0.75$  (mais à direita)] (adaptado de [31]).

Como outros dois exemplos de uma distância de poder, temos a distância de Mahalanobis, ou a função de distância heterogénea. Em (2.12) está identificada a distância de Mahalanobis numa versão simplificada, onde  $x$  é um vetor que contém todas as dimensões de uma única série temporal observada,  $\mu$  é um vetor que constitui o centro de massa de toda a informação retirada,  $n$  é o número de elementos de  $x$  e  $\mu$ , e  $d(x_i, \mu_i)$  corresponde à diferença entre os elementos  $i$  de  $x$  e de  $\mu$ .

$$m_{\mu,\sigma}(x) = \sum_{i=1}^n \frac{d(x_i, \mu_i)}{\sigma_i} \quad (2.12)$$

### 2. Distâncias nas leis de distribuição

Neste tipo de distância, as medidas são descritas com base na distribuição de probabilidade do *dataset* de variáveis, num mesmo intervalo. Nesta categoria, temos o exemplo da distância de Bhattacharya, mostrada em (2.13), que classifica as diferentes *features* pela sua capacidade em distinguir uma classe das restantes.

$$\rho(P_1, P_2) = \sum_{x \in X} \sqrt{p_1(x)p_2(x)}, \quad (2.13)$$

onde  $P_1$  e  $P_2$  correspondem às distribuições de probabilidade de variáveis no domínio  $X$ , enquanto que  $p_1(x)$  é a probabilidade de  $x$  ocorrer em  $P_1$ , e  $p_2(x)$  é a probabilidade de  $x$  ocorrer em  $P_2$ .

Outro exemplo de uma distância nas leis de distribuição é a distância  $\chi^2$ . A mesma está representada na equação (2.14), onde  $x$  e  $y$  são vetores extraídos de comprimento  $n$ ,  $p(x_i)$  é a probabilidade da ocorrência de um elemento  $i$  em  $x$ , assim como  $p(y_i)$  diz respeito à probabilidade de ocorrer um elemento  $i$  em  $y$ .

$$d(x, y) = \sum_{i=1}^n \frac{(p(x_i) - p(y_i))^2}{p(y_i)} \quad (2.14)$$

### 3. Distâncias de correlação

Distâncias e semelhanças de correlação visam caracterizar a correlação existente entre dois *datasets*. Esta comparação é entendida como uma distância ou medida de semelhança, e não como uma distribuição de probabilidades ou magnitude dos vetores. A correlação de *rankings* de Spearman,  $\rho$ , é um exemplo de distância de correlação, e está definida em (2.15).  $X_c$  e  $Y_c$  representam os postos das variáveis discretas  $X$  e  $Y$ ,  $x_i$  e  $y_i$  contém o posto  $i$  em  $X_c$  e  $Y_c$  respetivamente, e  $n$  é o número de elementos em  $X_r$ .

$$\rho(X_c, Y_c) = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)} \quad (2.15)$$

Outro exemplo deste tipo de distâncias é a correlação de postos de Kendall,  $\tau$ , que é nos dada pela equação (2.16). Nesta operação, é usada a função sinal,  $sgn$ , que calcula a subtração entre o número de pares discordantes dos postos e o número de pares concordantes,

$$\tau(X_r, Y_r) = \frac{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n sgn(x_i - x_j) * sgn(y_i - y_j)}{n(n-1)} \quad (2.16)$$

onde a função sinal é definida por

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \quad (2.17)$$

### 2.3.2 Features globais

As *features* que são extraídas, não são representativas de uma solução universal e podem não ser aplicadas a cada domínio de problema. Diferentes mecanismos foram desenvolvidos com o objetivo de otimizar as *features* e medidas de cada série temporal, associando as mesmas a um certo tipo de padrão.

As *features* globais são vistas como algoritmos que quantificam padrões numa série, ao longo do intervalo de tempo de análise [28], e conseguem ser produzidas através de um vasto número de mecanismos complexos. Para se obter uma calibração precisa das medidas retiradas, algumas delas são calculadas em ambas as séries no tempo de dados *raw*, e apenas neste tipo de dados [30] em detrimento de outros, como os dados TSA (com tendência e ajuste sazonal). Segundo o mesmo estudo existem nove *features* estatísticas, clássicas e desenvolvidas, para uma caracterização global de séries temporais: a tendência, a periodicidade, a correlação em série, a distorção, a curtose, a não-linearidade, a auto-similaridade e a teoria do caos.

#### 2.3.2.1 Representação vetorial das *features*

Existem alguns métodos de análise que podemos aplicar sobre séries no tempo definindo as suas propriedades, tais como a estacionaridade de valores, correlações lineares, métodos da literatura da análise física não-linear, ajuste do modelo linearmente ou não-linearmente ou até mesmo simples processos gaussianos, e todos eles são codificados algoritmicamente como operações. Cada operação,  $\beta$ , trata-se de um algoritmo que toma a série temporal,  $X = (x_1, x_2, \dots, x_n)$ , como *input*, dando origem a um simples número real como *output*,  $\beta : \mathbb{R}^N \rightarrow \mathbb{R}$ .

A estacionaridade de uma determinada janela temporal, é conseguida a partir de um sistema com parâmetros fixos e constantes ao longo da captação de imagem (ou distribuições de probabilidade dos parâmetros invariáveis), isto é, as medidas de estacionaridade mostram como as dependências temporais vão variando. Uma métrica plausível para o valor da estacionaridade média define-se por:

$$EM(\tau) = \frac{\text{std}(\overline{x_{1:w}}, \overline{x_{w+1:2w}}, \dots, \overline{x_{(m-1)w+1:mw}})}{\text{std}(x)}, \quad (2.18)$$

na qual em numerador temos o desvio padrão calculado a partir do conjunto de médias de sub-janelas não sobrepostas da série temporal a ser analisada, cada uma de comprimento  $w$ , enquanto no denominador temos o desvio padrão da série no tempo [28].

Por sua vez, a autocorrelação mede a correlação entre séries temporais que estão separadas por um certo atraso:

$$AC(\tau) = \langle x_t x_{t+\tau} \rangle = \frac{1}{s_x^2(N-\tau)} \sum_{t=1}^{N-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}), \quad (2.19)$$

onde  $\tau$  representa o atraso de uma série temporal  $x$ , com variância  $s_x^2$  e média  $\bar{x}$ .

Por fim, outra das formas de fazer uma representação vetorial de amostras temporais, é aplicando uma transformada de Fourier em tempo discreto, tal como já havia sido referida na Sub-seção 2.2 deste documento. Esta representa a projeção do sinal do domínio do tempo, para o domínio da frequência através de:

$$\bar{x}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^N x_n \exp\left(\frac{2\pi kn}{N}\right) \quad (2.20)$$

Na extração de *features* de imagens, ao usar-se uma técnica DFT (*Discrete Fourier Transform*), uma das técnicas comuns é o uso apenas dos primeiros  $k$  coeficientes, que representam as baixas frequências do sinal. Esta seleção de *features*, deve-se em grande parte ao facto das séries temporais em análise terem bastante ruído, que precisa de ser removido. Contudo, numa transformada de Fourier discreta, cada característica retirada é influenciada por toda a informação temporal retirada, uma vez que os coeficientes descrevem frequências globais [32].

Uma técnica melhor e mais eficiente passa por usar os maiores  $k$  coeficientes de cada série temporal, que se traduz numa maior conservação da quantidade ótima de energia, necessária para que, na reconstrução do modelo por extração de *features*, ele se assemelhe o melhor possível ao original.

### 2.3.2.2 Seleção e classificação das *features*

Em [33], dá-se o estudo de monitorizar a atividade de idosos para prever quedas. São usados um número mínimo de sensores e os dados são analisados de forma a determinar se os mesmos, correspondendo a diferentes atividades, podem ser agrupados por classes. Um conjunto de *features* foram calculadas através de acelerações *raw* gravadas, em janelas temporais de seis segundos, e deram a origem a outras que partiram as amostras em quadros mais pequenas, de forma a analisar transições de movimento. A tendência média e a diferença média em janela, onde os intervalos de seis segundos são divididos em 12 sub-intervalos, calculados por

$$\mu T = \sum_{i=2}^{12} (|\mu_i - \mu_{i-1}|), \quad (2.21)$$

onde a média de cada período de 0.5 segundos é obtida e subtraída à média da janela seguinte. A tendência média é então encontrada através do somatório de todas as subtrações possíveis, dentro da janela de seis segundos inicial. De seguida, a tendência de variância e a diferença de variância em janela podem ser calculadas de forma semelhante, sendo que é a variância que é obtida e não a média, para cada uma das janelas da seguinte forma:

$$\sigma T^2 = \sum_{i=2}^{12} (|\sigma_i^2 - \sigma_{i-1}^2|). \quad (2.22)$$

Uma vez que uma dada série temporal  $y(k)$  é analisada de acordo com as equações (2.21) e (2.22), ela é dividida em  $N$  intervalos de comprimento  $n$ , sem sobreposição, cada um com uma tendência  $y_n$  associada. A flutuação média  $F(n)$  correspondente ao intervalo total, é dada por

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N [y(i) - y_n(i)]^2}. \quad (2.23)$$

## 2.4 Sinal de rádio e presença em imagem

Existem algumas técnicas de deteção através de WiFi, baseadas em diferentes propriedades de camada física. Elas podem ser aplicadas diretamente no *hardware* de raiz, mas também em *hardware* customizado como é o caso da implementação de um radar FMCW. Estes operam até larguras de banda de 1.79 GHz, em oposição aos 20 MHz dos dispositivos WiFi de diversas normas. Por isso, oferecem um tempo de resolução muito melhor nos seus resultados [34]. Por esta razão, as técnicas por FMCW podem não fazer grande diferença se for feita apenas um levantamento da atividade humana geral, mas na identificação de gestos ou até mesmo de sinais vitais, são preferíveis as suas utilizações.

Com tudo o que foi introduzido na Sub-secção 2.2 deste documento, a expressão matemática para o sinal analítico, na sua forma complexa, transmitido por uma antena de um radar FMCW, é formulada por:

$$s_T = A_T \exp \left[ j2\pi \left( f_c t + \frac{1}{2} K t^2 + \varphi_0 \right) \right], \quad (2.24)$$

na qual  $A_T$  corresponde à amplitude do sinal transmitido,  $f_c$  à frequência portadora do sinal,  $K$  à chamada "inclinação da modulação de frequência",  $t$  ao "tempo rápido" num intervalo  $[-\frac{T}{2}, \frac{T}{2}]$  onde  $T$  é o período de uma forma de onda e, por fim,  $\varphi_0$  constitui a fase inicial.

### 2.4.1 Métodos para reconhecimento de atividade e de gestos ao nível do rádio FMCW

Os radares FMCW, ao contrário de outras técnicas de deteção sem fios, conseguem separar as reflexões do sinal de rádio provocadas por diferentes objetos. O espectro RF tem propriedades bastante atrativas para sistemas de deteção, pela sua robustez face a condições ambientais adversas. Este tipo de sensores emitem sinais que são bastante rápidos e precisos, para além de conseguirem atravessar diferentes tipos de material.

Movimentos humanos, como mover um braço em qualquer direção, provoca sinais de radar que não são estacionários, e tipicamente conseguem ser analisados no domínio frequência-tempo. Na prática, podemos ter o mesmo domínio representado num espectrograma, obtido por uma transformada rápida de Fourier sobre os dados recolhidos e observados, que é por sua vez utilizado para extração de características.

#### 2.4.1.1 Power Burst Curve

Este método é aplicável ao domínio frequência-tempo. Para um determinado sinal discreto  $s(n)$  de comprimento  $L$ , o espectrograma é obtido por aplicação da FFT ao sinal recebido, e é aplicado um mecanismo de janela deslizante a todos os pontos da FFT de modo a filtrar o sinal dado por:

$$S(n, k) = \left| \sum_{m=0}^{L-1} s(n+m)h(m)e^{-j2\pi\frac{mk}{L}} \right|^2, \quad (2.25)$$

onde  $n = 1, \dots, L$ , em que  $L$  corresponde ao tamanho da janela deslizante, e  $k = 1, \dots, K$  que simboliza a janela em si, isto é, o deslocamento ao longo do período do sinal discreto.

O início e o final de cada movimento podem ser retirados monitorizando a chamada *Power Burst Curve* [35], PBC, que mede a energia do sinal no espectrograma dentro de bandas de frequência específicas, ao realizar-se a soma do somatório do módulo das partes negativas do espectro, com o somatório das partes positivas. Por norma, existe uma certa banda à volta da frequência de 0 Hz que não é utilizada, uma vez que provoca erros de precisão:

$$S(n) = \sum_{k_1=K_{N1}}^{K_{N2}} |S(n, k_1)|^2 + \sum_{k_1=K_{P1}}^{K_{P2}} |S(n, k_1)|^2. \quad (2.26)$$

#### 2.4.1.2 Correlação canónica

Em [36], é feita a distinção de diversos tipos de movimento, exemplificando com o movimento de cair, de sentar, de torcer o corpo e de andar. Ao definir-se uma matriz  $V$  que contém imagens rastreadas no tempo  $x_i$ ,  $i = 1, \dots, M$ , de um movimento específico, temos:

$$V = (x_1, x_2, \dots, x_M). \quad (2.27)$$

Tomando a matriz  $V$  como normalizada, isto é, aplicando uma variância unitária a todos os valores da matriz, cada imagem vetorizada vai ter de tamanho  $1 \times N$ . O subespaço  $d$ -dimensional de um simples movimento corporal pode ser obtido com recurso a uma decomposição em valores singulares (SVD), sob a matriz  $V$ . Para calcular a distância entre dois subespaços  $\phi_1$  e  $\phi_2$ , de forma a proceder-se à deteção do movimento, é usada a análise de correlação canónica entre os mesmos, que é definida por:

$$\cos\theta_i = \max_{u_i \in \phi_1} \max_{v_i \in \phi_2} u_i^T v_i, \quad (2.28)$$

obedecendo às condições  $\|u\| = \|v\| = 1$ ,  $u_i^T u_j = v_i^T v_j = 0$ ,  $i \neq j$ , onde  $u$  e  $v$  correspondem aos vetores das imagens. É realizada uma decomposição em valores singular das matrizes  $U$  e  $V$ , representativas das bases ortogonais dos dois subespaços [36].

#### 2.4.1.3 Extração da frequência de Doppler

Este método diz respeito ao domínio fase-tempo. Na desmodulação do sinal recebido e de forma a identificarmos ao detalhe a presença de movimento no mesmo, é necessária a remoção da componente de Doppler. No sistema de equações (2.10), tínhamos representada a fase de cada amostra do sinal. Uma função de *fitting* é conseguida por [37]:

$$F(n) = \sum_{k=0}^K m_k \varphi_m^k, \quad (2.29)$$

em que  $\varphi_m$  corresponde ao histórico de fase da  $n$ -ésima amostra presente no sinal recebido e  $m_k$  é o coeficiente do *fitting* linear de ordem  $K$  do número total de pontos da amostra. Este, procura minimizar a perda quadrática média  $S$  entre o sinal refletido e recebido, e a função de *fitting*.

$$S = \frac{1}{N_f} \sum_{n=1}^{N_f} (P(m) - \varphi_m)^2. \quad (2.30)$$

Depois de ser realizado todo o processo, correspondente à subtração da função de *fitting* ao sinal recebido, e ao longo de todas as amostras, a componente de Doppler que estava presente graficamente no sinal, é removida.

## 2.5 Trabalhos na área

Muitos foram os estudos que realçaram o uso de sinais *wireless*, sem fio, para a deteção das atividades humanas. O movimento de um corpo num ambiente onde existe a

propagação de sinais, tem impacto direto nos mesmos, o que se analisa através das mudanças significativas na amplitude e fase de sinais interferidos por certo indivíduo, face ao que foi transmitido. Com isto, concluiu-se que é possível achar a localização de um utilizador, sem que ele tenha na sua posse um dispositivo *wireless*, abrindo o caminho ao desenvolvimento de outros sistemas inteligentes [38].

O desenvolvimento de novas tecnologias de entrada, isto é, dispositivos que funcionam como uma plataforma de acesso e de interação com a informação, tem mudado a forma como o processamento a nível computacional se manifesta. Ao nível da modelação do movimento com sensores de radar, permitiu que a interação humano-computador se aprimorasse. No caso do radar FMCW, a sua utilização na deteção de movimento humano, foca-se em determinar gestos que englobam movimentos controlados e precisos.

### 2.5.1 Reconhecimento de atividade humana geral

Trabalhos recentes em imagens baseadas em RF, demonstraram a capacidade da obtenção de imagens com alta resolução em tempo-real, com o uso de um radar de onda milimétrica. Estes radares, estão frequentemente ligados a técnicas convencionais de *Machine Learning*. Os *Support Vector Machines*, a sincronização temporal dinâmica e os classificadores de floresta aleatória, são exemplos dos algoritmos de *Machine Learning* mais usados no que toca ao reconhecimento da atividade humana baseado em radares.

Diferentes tipos de movimento humano geram diferentes reflexões, que são também chamadas de assinaturas micro-Doppler. Em [39], foram captados sinais micro-Doppler de 4 movimentos diferentes de um indivíduo: a rastejar, a gatinhar, a correr e a andar. A classificação do movimento foi feita ao treinar-se um modelo oculto de Markov. Na Figura 2.12, vemos um exemplo de um espectrograma micro-Doppler retirado de uma pessoa a andar. As mudanças de frequência na região negativa correspondem a momentos em que a pessoa está-se a afastar do radar, enquanto que mudanças de Doppler positivas traduzem-se num aproximar do indivíduo ao radar.

Em [41], é feito o reconhecimento 3D de um indivíduo, o qual se divide na capacidade de achar o movimento de um braço e na capacidade de se detetar uma queda. De forma a perceber para onde alguém está a apontar, é necessário distinguir entre mover o corpo todo e mover apenas um braço, sabendo que a superfície de reflexão de um braço é muito mais pequena que a superfície de reflexão do corpo inteiro. Desta forma, para esta distinção é utilizada uma variância espacial (a variância do sinal ao longo do eixo vertical é significativamente maior quando o refletor é o corpo todo).

### 2.5.2 Reconhecimento de gestos

Com a computação cada vez mais presente no ambiente, o reconhecimento de gestos por parte de sistemas de deteção inteligentes, permite que se tenha um vasto conjunto de técnicas de interação com os humanos.

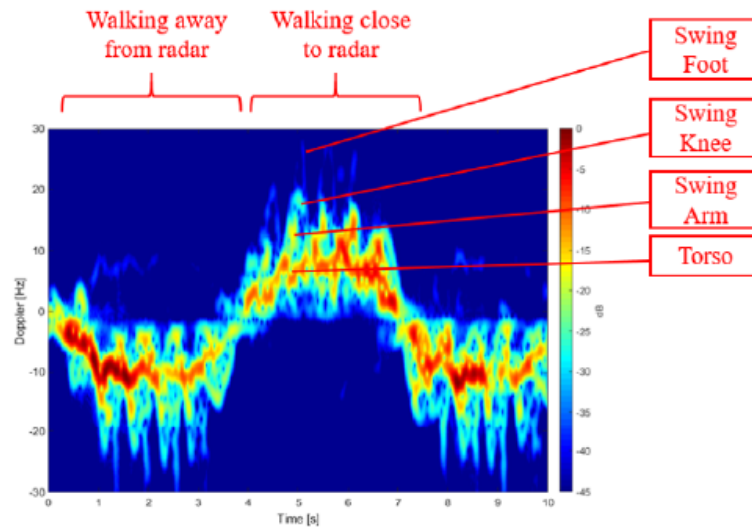


Figura 2.12: Espectrograma da assinatura de Doppler de uma determinada pessoa a caminhar (Adaptado de [40]).

De forma introdutória a outros estudos a referir, em [42] é apresentado o *WiSee*. Trata-se de um sistema *wireless* que captura a informação de gestos a partir das propriedades de mudança de frequência do sinal, isto é, permite a deteção de padrões de gestos, através das frequências de Doppler positivas e negativas extraídas do sinal recebido. O desafio impulsionador do sistema referido, é o facto do movimento humano provocar mudanças de frequência na ordem das dezenas de Hertz, pelo que são difíceis de detetar com outros sistemas sem fios, nomeadamente gestos pequenos, como por exemplo ao nível da mão. *WiSee* apresenta uma solução baseada no sinal recebido, transformando-o num pulso de banda estreita com baixa largura de banda.

De modo a extrair a informação de Doppler e a mapeá-la em gestos, o trabalho introduzido no parágrafo anterior é descrito em 3 passos:

- Extração de Doppler - Processa as mudanças de frequência dos sinais de banda estreita recebidos e, então, são aplicadas FFTs sucessivas às amostras.
- Segmentação - Identifica um conjunto de segmentos associados aos gestos, através de um processo semelhante à deteção de pacotes em sistemas de comunicação. Como num dado segmento é processada a energia média nas frequências de Doppler positivas e negativas, quando a diferença da mesma com o nível de ruído ultrapassa um limiar de decisão definido, então traduz-se no início de um segmento.
- Classificação - Determina o gesto mais provável dentro de um conjunto de possibilidades. Considerando que apenas existem 3 tipos de segmentos, sendo eles segmentos com apenas mudanças de Doppler na parte positiva, com apenas mudanças de Doppler na parte negativa e com mudanças nos dois lados, aos quais se

atribui os números "1", "-1" e "2", respetivamente. Desta forma, há a criação de padrões de gestos, que são utilizados para classificar novos sinais recebidos.

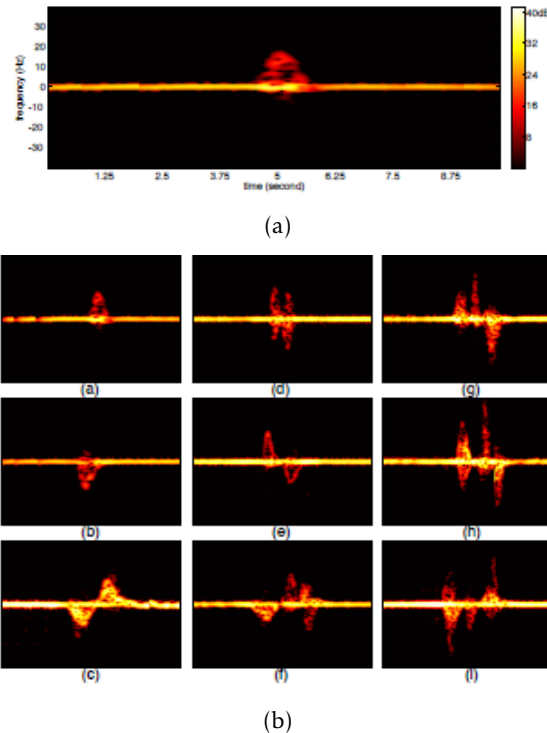


Figura 2.13: Perfis de Doppler no domínio tempo-frequência de gestos avaliados, onde é visível que cada segmento pode englobar mudanças de Doppler no lado positivo e/ou negativo do gráfico (Adaptado de [42]).

No seguimento do que foi referido anteriormente, técnicas como o modelo oculto de Markov ou uma sincronização temporal dinâmica podem aumentar o espaço de gestos possíveis. No trabalho realizado em [43], temos o exemplo de um sistema de reconhecimento de gestos com o radar FMCW, com base num modelo oculto de Markov. Os dados de radar recebidos são recolhidos e processados de forma a extrair-se os vetores com as características de classificação desejadas, os quais são usados para treino a um sistema de reconhecimento de gestos, baseado num modelo oculto de Markov previamente treinado. Uma vez que são usados vetores velocidade-energia para cada ponto de cada *frame* de um determinado movimento, é explicada a importância dos mesmos para evitar problemas de interferência.

Em [44] foi proposto o Soli, o primeiro sistema de onda milimétrica designado *end-to-end* para reconhecimento de gestos ao nível da mão. Fazer a associação de gestos a partir deste sistema passa, de forma inicial, por se fazer a extração de características de classificação, conectada a uma perspetiva de *Machine Learning* onde as mais importantes são seleccionadas durante a fase de aprendizagem. É construído um vetor de características que serve de *input* a outro algoritmo de *Machine Learning*, para classificação de gestos. De

seguida, é realizada a inferência de gestos possíveis, onde o sistema referido usa um classificador *Random Forest*, em detrimento de outros mais exigentes a nível computacional. Por fim, com o intuito de se melhorar a precisão das previsões feitas pelo classificador, é aplicado um filtro Bayesiano, o qual reduz significativamente erros esporádicos provocados por falsos-positivos.

Em muitos sistemas, quando um determinado gesto é servido como *input*, uma matriz de Doppler-Alcance, que representa uma característica de classificação resultante do movimento de entrada, é aprendida através de uma rede neuronal profunda e o resultado é inferido. No entanto, a classificação de gestos através de *Deep Learning* tem um problema comum, uma vez que movimentos iguais feitos por pessoas diferentes têm diferenças e podem afetar o resultado. Em [45] é introduzido um algoritmo de adaptação de domínio ao sistema de reconhecimento de gestos, que adapta a informação de um domínio existente quando novos dados chegam como *input*, provenientes de outro domínio. Ao nível do reconhecimento de gestos, o classificador é treinado com as características de classificação do domínio de origem com *labels* associadas. Na extração de *features*, são eliminadas certa parte para uma "camada de eliminação", o que vai reduzir erros de precisão provocados por problemas de *overfitting*. Assim que a última função de reconhecimento de gestos do domínio fornecido é completada, a função de perda do classificador é expressada na forma de uma entropia cruzada.

### 2.5.3 Monitorização de sinais vitais

Os sistemas de radar são úteis na monitorização de condições de saúde, nomeadamente ao nível respiratório e do batimento cardíaco. Os radares de onda contínua são os que mais se destacam entre as tecnologias existentes, oferecendo elevada precisão em detetar deslocamentos relativos em imagens.

Em grande parte, no estudo de sinais vitais são utilizados radares híbridos, os quais contêm o modo FMCW em conjunto com um modo Doppler. O lado FMCW é usado para captação de imagem 2D de uma dada cena, enquanto que a parte de Doppler atua na extração de fase e destaca o alvo em estudo. No entanto, estudos mais recentes trouxeram soluções de deteção de sinais vitais usando um sistema FMCW puro [46], através de um método baseado em desmodulação de fase com uma plataforma móvel [24].

Em movimentos maiores, como mexer um membro superior ou inferior, são vários os sensores que facilmente os detetam. Contudo, a deteção da respiração ou do batimento cardíaco reduzem significativamente esse lote de sistemas, devido às baixas amplitudes de sinal na ordem de um centímetro a uma fração de milímetro. A tecnologia avançada em [47], representa um algoritmo de medição dos sinais vitais de indivíduos presentes em cena, sendo capaz de separar os diversos sinais que provêm de diferentes utilizadores. Através do sinal recebido, pretende identificar quando o humano está praticamente estático, com apenas os sinais vitais a dominar, ou quando ele está a andar e a fazer mover algum membro. É visto que o movimento proveniente de um sinal vital é periódico,

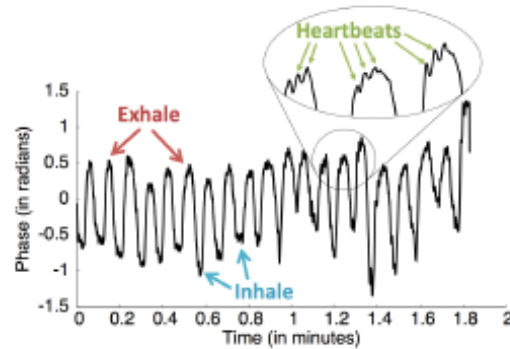


Figura 2.14: Variação de fase provocada pelos sinais vitais. Nas ondas de frequência mais baixa temos a respiração, onde os picos e os vales correspondem à expiração e inspiração, respetivamente, às quais se sobrepõem frequências mais altas correspondentes ao batimento cardíaco (Adaptado de [47]).

enquanto que de um braço, ou de uma perna, é aperiódico, e é a partir desta ideia que é feita a distinção entre movimentos. O mesmo é visível na Figura 2.15. Uma aproximação a medir a periodicidade do sinal recebido, consiste em aplicar a transformada de Fourier em cada janela temporal do mesmo, e ver quando o valor de pico da frequência é suficientemente mais alto que a potência média nas restantes frequências.

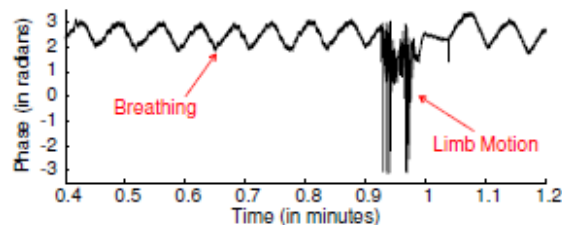


Figura 2.15: Periodicidade do sinal afetada pelo movimento de um membro (Adaptado de [47]).

Em [48] com o intuito de selecionar a frequência ideal para deteção de sinais vitais, foi realizada uma série de medições de radar em seis voluntários, deitados numa cama, com 4 ângulos de incidência do sinal distintos. Foram utilizados 5 parâmetros de *output* para descrever os sinais vitais de pessoas a dormir: a frequência de respiração estimada, a frequência do batimento cardíaco estimada, duas medidas de confiabilidade associadas a cada uma das frequências já enumeradas e um valor escalar que serviu como indicador de sinal vital. Como métodos de estimação da frequência de respiração e do batimento cardíaco foram introduzidas e detalhadas a aplicação de uma FFT e de uma auto-correlação.

Na tabela 2.1 estão resumidas todas as tecnologias apresentadas.

Tecnologia	Precisão	2D/3D
Padar et al. [39]	90%	2D
WiTrack [41]	96,9%	3D
WiSee [42]	94%	3D
Malysa et al. [43]	82,3%	3D
Soli [44]	92,10%	3D
Lee et al. [45]	93,58% 2D/99,06% 3D	Ambos
Yan et al. [46]	–	3D
Vital-Radio [47]	99%	3D
Anitori et al. [48]	55%	2D

Tabela 2.1: Comparação das tecnologias de detecção da atividade humana referidas.

## DATASETS DE IMAGEM

Este capítulo descreve como foram criados os datasets de imagem e RF, e quais as ferramentas e algoritmos usados no tratamento dos dados de forma a levar ao resultado do classificador final, e às características destacadas e processadas em cada um dos lados.

### 3.1 Classificação de imagem

Quando se fala em classificação de imagem, o nosso objetivo inicial é definir quais as métricas a usar sobre a imagem de forma à pre-definição de classes, ou *labels*, usadas em classificações futuras. Definindo qual a classe do *dataset* em estudo, permite depois o estudo do espaço latente assim como das distribuições Gaussianas associadas.

#### 3.1.1 Machine Learning

As técnicas de *Machine Learning* dizem respeito a um subcampo das ciências de computação, responsável por desenvolver modelos capazes de identificar padrões a partir de um conjunto de dados em estudo.

Os algoritmos de *Machine learning* contêm partes de código que auxiliam o utilizador na análise, exploração e obtenção de resultados em conjuntos de dados complexos. Os algoritmos podem ser divididos em três categorias definidas pelo tipo de *datasets* que estão a ser usados:

- Classificação supervisionada: O modelo é aprendido através dos dados de *input* e pelos dados de *output* que são esperados. É o tipo de classificação que mais necessita da participação humana, uma vez que são necessários inúmeros dados de treino e, por conseguinte, um modelo minimamente formulado para saber responder e classificar novos dados. Os algoritmos preveem com base num conjunto de exemplos etiquetados, que foram introduzidos como *input* no modelo pelo utilizador de forma a treiná-lo. Este tipo de classificação representa a técnica mais usada para a análise quantitativa de dados em imagens.

- Classificação não-supervisionada: Neste tipo de classificação, o modelo é aprendido apenas a partir dos dados de *input*, isto é, os mesmos não são etiquetados. Desta forma, usamos classificação não supervisionada quando não é expectável o tipo de resultado que se irá obter, e quando o número de dados etiquetados é bastante inferior ao número de dados que não estão etiquetados.
- Classificação por reforço: Algoritmos que aprendem através dos resultados de forma a decidir o que fazer depois. O "agente de reforço" decide o que deve ser feito para encontrar a melhor resposta possível do sistema, e mesmo na ausência de um conjunto de dados de treino, o modelo irá conseguir aprender pela sua experiência.

No âmbito desta dissertação, o interesse recai no uso da classificação supervisionada, aproveitando-se o conhecimento prévio obtido através dos dados de entrada do modelo, do conhecimento do utilizador, que fazem gerar *datasets* distintos com base nas suas características. Assim, será possível definir as métricas aplicadas para extração de características, de modo a serem posteriormente usadas em modelos de classificação futuros.

## 3.2 *Mediapipe framework*

Uma das tarefas mais complicadas dentro da visão computacional, é determinar a configuração de alto grau de liberdade no que toca ao corpo humano, com todos os seus membros, partes que se assemelham entre si e com variações na sua forma devido ao tipo de corpo, às roupas, à iluminação, entre muitos outros fatores.

O *Mediapipe* representa uma estrutura de desenvolvimento disponível, que contém código aberto, de forma a fornecer soluções de Machine Learning personalizáveis para mídia ao vivo e streaming. Este *framework*, pode suportar uma vasta coleção de algoritmos de *Machine Learning* para funções de processamento de mídia ou ainda para instanciar um conjunto de modelos de inferência.

### 3.2.1 *Mediapipe Pose*

O *Mediapipe Pose* [49], aplicado neste trabalho, representa uma solução de ML bastante confiável que oferece a estimação da pose de uma pessoa individual. A mesma é capaz de extrair 33 pontos-chave distribuídos ao longo de todo o corpo, a partir de frames de vídeo RGB, podendo ser também aplicada a um dataset de imagem previamente preparado. Os 33 pontos estão representados na figura 3.1.

Esta solução apresenta como output, em cada um dos pontos:

- As coordenadas x e y: normalizadas entre 0 e 1 pela largura e altura de imagem, respetivamente;
- Uma coordenada z: representa a profundidade do ponto de referência com a profundidade do ponto médio dos quadris (origem);

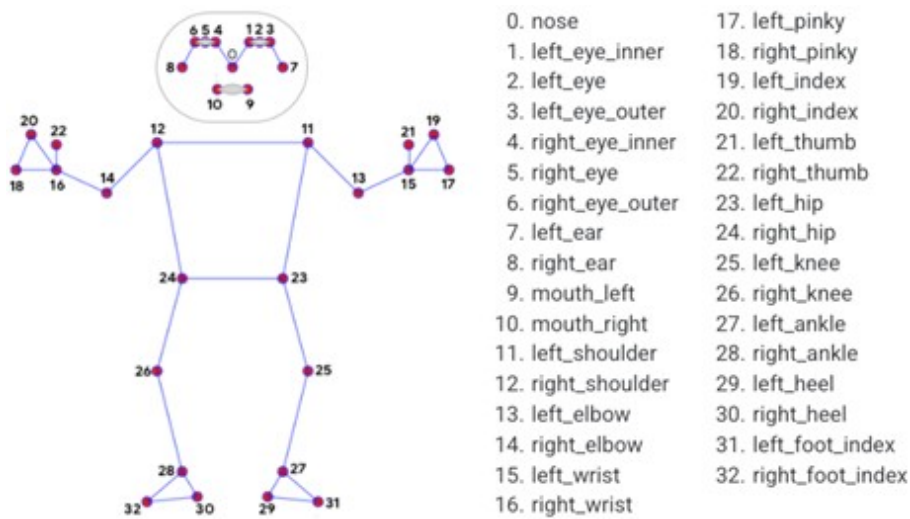


Figura 3.1: Mediapipe Pose.

- A visibilidade: um valor entre 0 e 1 indicando a probabilidade do ponto de referência ser visível, isto é, presente e não escondido, na imagem.

Outras características que pesaram na escolha da *framework* Mediapipe, incluem:

- Permitir o uso de imagens estáticas: Permite a detecção da forma humana mais distinta nas primeiras imagens procedendo à identificação mais clara dos *inputs* (33 pontos de referência detetados pelo algoritmo) ao longo do corpo e, nas imagens seguintes, aquilo que a *framework* necessita de fazer é rastrear os pontos já identificados, sem que mais nenhum método de detecção seja chamado. Isto permite com que modelos mais carregados para detecção possam ser introduzidos, enquanto o *pipeline* mantém um desempenho leve e rápido em tempo real;
- O uso de um valor mínimo de confiança na detecção: Faz com que seja estabelecido um limite na detecção de pessoas, o que leva a uma melhor qualidade dos dados, isto é, impõe um valor mínimo a partir do qual a classificação pode ser considerada bem sucedida.

A visibilidade como *output* ao nosso algoritmo, foi um fator inicial determinante na continuação do uso da ferramenta, avaliando-se em termos quantitativos o número de pontos com visibilidade 1 ou muito próxima. Um grande número de pontos com visibilidade 1 indica que podemos ter uma boa confiança no algoritmo, já que valores deste indicador vão do 0 ao 1, conforme são nada ou totalmente visíveis, respetivamente. A prévia construção de gráficos 3D, com as coordenadas  $x$  e  $y$  ao longo de todas as frames, dando-se cores diferentes aos pontos consoante o seu grau de visibilidade, possibilitou observar qual o desempenho que o algoritmo poderia ter em todo o trabalho.

### 3.3 Consciência de contexto

Neste trabalho, o objetivo inicial passa pela construção de um modelo de inferência de forma a identificar o tipo de movimento realizado entre duas imagens consecutivas.

#### 3.3.1 Entradas e saídas do modelo

Nesta análise de contexto, o tipo de movimento entre duas imagens consecutivas vai ser implementado/reconhecido de alguma forma. Assim, teremos como *inputs* ao nosso modelo:  $\gamma$ , que representa o tipo de articulação. É um identificador desde 1 até 33, classificação relativa à parte do corpo humano referente (algoritmo Mediapipe Pose tal como introduzido anteriormente); as variáveis  $x$  e  $y$ , indicadoras do deslocamento que foi realizado, por cada tipo de articulação.

Estes dados levarão à criação de um vetor de observações, do tipo:

$$(\gamma_{k-1}, x_{k-1}, y_{k-1}), (\gamma_k, x_k, y_k), (\gamma_{k+1}, x_{k+1}, y_{k+1}). \quad (3.1)$$

Como saídas do estudo feito através dos *inputs* referidos, devemos fazer algumas questões tais como quais os contextos ou sequências queremos identificar. E é a partir daqui que vai ser criado um modelo, que vai englobar este tipo de informação e ser treinado para avaliar novos movimentos, assim como novas sequências dos mesmos no futuro.

#### 3.3.2 Dataset

Foi visto que uma das características fundamentais dos métodos de Machine Learning supervisionados é o facto da representação de atividades estar dependente de um conjunto de atividades pré-definidas, que foram previamente estudadas e etiquetadas. Isto significa que a adição de uma nova classe requer um novo dataset de imagem ou que todo o sistema seja de novo treinado.

Para treinar o nosso modelo, o objetivo passou por 4 cenários diferentes, todos eles recolhidos previamente neste trabalho, para os quais houve a criação de 4 *datasets* principais: cenário de levantar e sentar numa cadeira, de estar sentado e levantar um ou os dois braços, de andar de um lado para o outro numa sala e, por fim, o cenário de andar à volta de uma cadeira. Dentro de cada cenário houve criação de *sub-datasets*, exemplificando que na ação de levantar e sentar na cadeira, a mesma é feita para ângulos diferentes. No movimento de levantar um dos braços, enquanto que num dataset é levantado apenas um dos braços, nos outros é levantado o outro ou os dois braços.

Com isto, ao existirem 4 tipos de movimento diferentes, queremos logo à partida ter 4 *outputs*. Pode também haver sequências com número de imagens variável para os diferentes movimentos, ou por exemplo existir movimentos no interior de outros.

Na figura 3.2, está apresentado um modelo ilustrativo daquilo que será a ideia inicial de implementação. Sendo  $\gamma$  o tipo de articulação,  $x$  e  $y$  as coordenadas *output* de cada

ponto no algoritmo Mediapipe Pose, o objetivo será conseguir, por exemplo, chegar à probabilidade de ter  $\gamma$ , dado que estou na sequência 1, à probabilidade de ter  $x$  dado que estou na sequência 2, e por aí adiante.

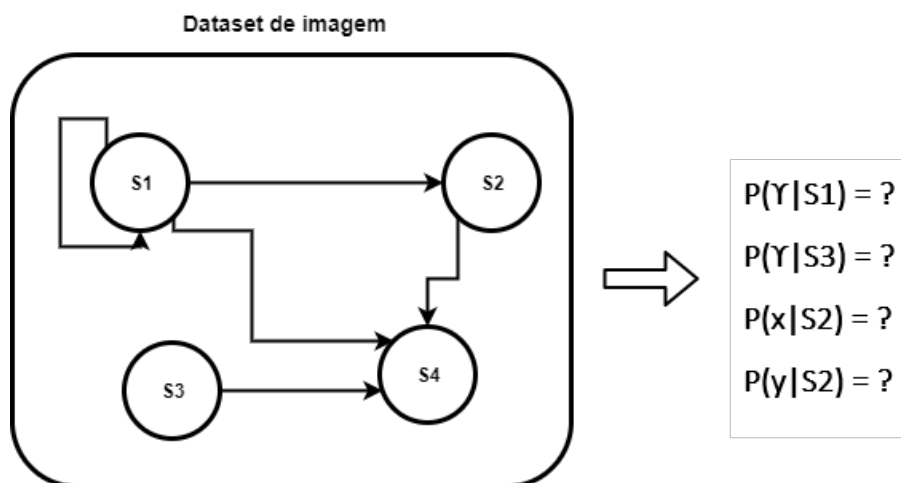


Figura 3.2: Modelo inicial.

Como referido anteriormente, num *dataset* de imagem que envolve os diferentes movimentos, começamos pelo princípio básico de se obter a probabilidade de termos determinado valor de *output* ao algoritmo mencionado na secção anterior, sabendo que estaremos num dado tipo de movimento identificados na imagem por  $S1$ ,  $S2$ ,  $S3$  e  $S4$ .

### 3.3.3 Modelo estatístico

Num determinado *dataset*, uma primeira abordagem ao estudo de todo o nosso algoritmo e que permite ter uma primeira perceção à eficácia do mesmo, é a construção de *plots* em 3 dimensões para cada uma das 33 partes do corpo humano identificadas. A variação do deslocamento definida a cada 2 pares de coordenadas para cada articulação, ao longo de todas as imagens do conjunto, dispostas no eixo restante, permitirá ver a distribuição de todos os pontos diferenciais, e observar se faz sentido a forma como estão dispostos. A título exemplificativo, num movimento em que estamos sentados e movemos apenas um dos membros, é normal que haja variações significativas para o gráfico desse membro, em determinados conjuntos de imagens consecutivas.

É então objetivo a aquisição de conhecimento através de um modelo estatístico. Para inferirmos estatisticamente o modelo, é necessário estudar as distribuições dos dados. Nesse sentido, definiu-se amostrar o movimento em  $x$ , em  $y$  e  $\gamma$  para cada articulação, pelo que temos a formação de domínios, tal como está explícito na figura 3.3.

Tal como já foi introduzido anteriormente, temos 4 tipos de movimento. Esse agrupar de ações representa o conjunto de observações possíveis, o espaço amostral, e iremos ter distribuições de probabilidade de cada um dos *outputs* do algoritmo do Mediapipe Pose implementado, isto é:

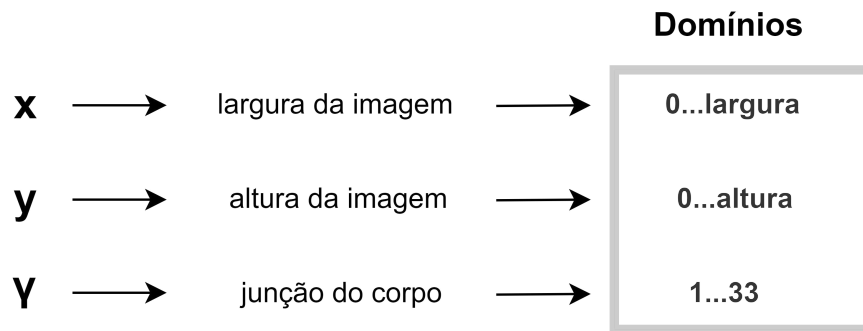


Figura 3.3: Modelo estatístico.

- Distribuições para as variáveis de saída  $x$  e  $y$ , indicadoras do posicionamento de cada uma das articulações trabalhadas, ao longo de todo um *dataset*. Isto permite fazer estudos sobre a posição dos membros num conjunto de *frames* de determinada classe dentro do *dataset* no seu todo, assim como correlacionar os histogramas, pertencentes à mesma classe, mas situados em espaços temporais diferentes.
- Distribuições para a variável  $\gamma$ , fazendo a análise de quantas articulações se realmente movimentam em cada conjunto de classe ou a distribuição desse número de movimentos em todo o *dataset*. Irá ser possível observar os diferentes histogramas, para os diferentes tipos de movimento.

Este tipo de modelos de inferência exploram muitas vezes as dependências de auto-correlação numa rede de dados, de forma a melhorar futuras previsões. No nosso caso, e pegando nas 33 articulações estudadas em todo o modelo, são geradas dependências entre os "atributos" das instâncias relacionáveis. Estamos a explorar a auto-correlação ao estimar distribuições de probabilidade das variáveis, partindo dos dados de entrada do modelo de forma a inferir certas *labels* (com o tipo de movimento) sobre dados futuros.

### 3.4 Análise do movimento

Muito mais que as distribuições de probabilidade que se podem retirar do estudo prévio das entradas e saídas ao nosso modelo, e prosseguindo com o estudo de prever movimentos futuros, é importante a caracterização dos *datasets* através de histogramas sobre o deslocamento entre pares de frames.

Do modelo estatístico introduzido anteriormente, o estudo recai sobre experiências sobre *datasets* de imagem, comparando os mesmos com modelos generativos através do cálculo da probabilidade de dados de teste. Quanto maior o número de amostras de teste, melhor o nosso modelo generativo. O trabalho sobre os diversos conjuntos de dados, permite depois ajustar um classificador, a fim de se obter funções de densidade de probabilidade sobre dados que correspondem ao nosso modelo.

### 3.4.1 Vetores deslocamento e angular

Um vetor de movimento é definido como um segmento de reta que vai desde um ponto de origem a um ponto de destino no espaço Euclidiano, que pode ser definido por  $n$ -dimensões. Um determinado movimento humano, que pretendemos estudar neste trabalho, é composto por  $n$ -vetores de movimento, onde existe a componente escalar e a componente angular. Estes vetores de movimento podem ser vistos como uma conceitualização do movimento como propriedade do espaço em vez de uma propriedade da trajetória do movimento. Vão ser por isso usados de forma a prever movimentos seguintes, e não tanto a trajetória no seu todo.

Uma vez que o nosso estudo não recai apenas sobre movimentos no plano 2D, onde teríamos um movimento em linha reta, é importante o estudo dos vetores angulares de forma a termos informação sobre a orientação do movimento. O deslocamento angular é útil de forma a quantificar o alcance do movimento em determinada articulação.

Pegando no algoritmo de imagem introduzido na subsecção 3.2.1, irão ser retirados os valores instantâneos do deslocamento linear e angular em cada articulação, aos quais serão aplicadas métricas fundamentais à construção de classificadores para o movimento. Tal como referido no nosso modelo estatístico, a análise de distâncias leva à construção de histogramas, que vão caracterizar os dados e permitir fazer distinções entre os diferentes casos dos *datasets*.

Na figura 3.4 está representado um indivíduo que a certa altura levanta os dois braços. No histograma exemplo relativo a um ponto na mão esquerda da pessoa, ainda que tenhamos uma maior concentração dos pontos em distâncias menores tal como no histograma relativo a um ponto no seu pé esquerdo, no histograma representativo da mão teremos a presença também de pontos com maior distância a ocorrer, registados quando o movimento propriamente dito foi feito.

### 3.4.2 Estudo de métricas

Com o desenvolvimento de todo o algoritmo, onde teremos histogramas da distância para cada uma das articulações e dentro de cada um dos cenários, o objetivo seguinte passa por criar métricas, baseadas na métrica da variância, que servirão como ponto de partida à classificação do modelo. O primeiro passo é então fazer o cálculo da média dos histogramas, contendo todos os deslocamentos, para cada uma das articulações identificadas no algoritmo:

$$\mu_j = \sum_{k=1}^n x_k p_k, \quad (3.2)$$

em que  $n$  representa o número total de deslocamentos  $x$ , e  $p$  representa a probabilidade do  $x$  respetivo que é fornecida pelo histograma.

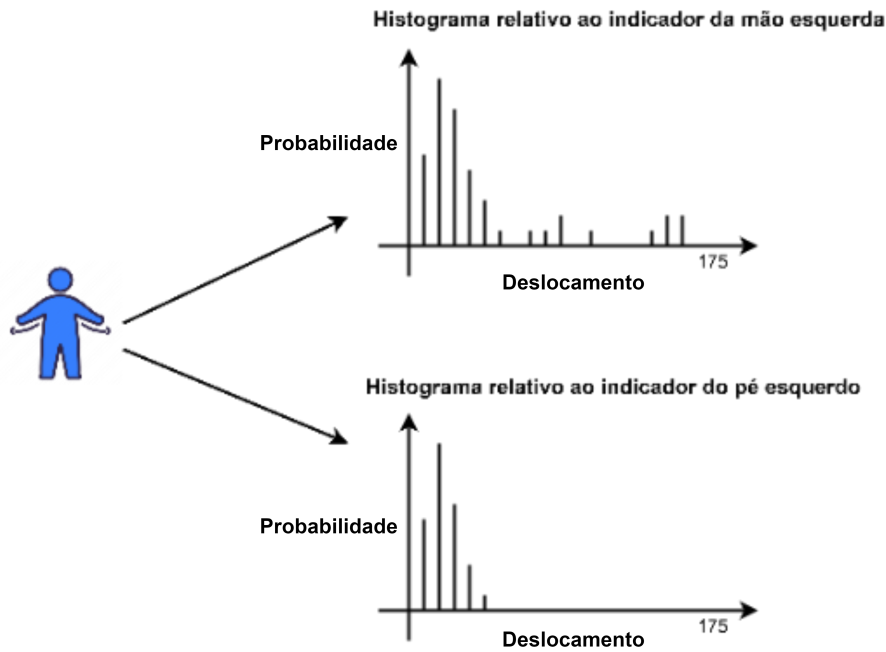


Figura 3.4: Levantar dos dois braços por parte de uma pessoa, e os histogramas respectivos para o indicador da mão esquerda, e para o indicador do pé esquerdo.

De seguida, foi calculada a distância para cada articulação dada pelo desvio à média, isto é, a diferença entre o valor observado e os valores que estão no histograma daquela classe e da articulação respetiva, que irão estar a ser pesados pelas probabilidades. Tendo o valor de  $x$  para cada amostra de deslocamento, a distância é dada pela equação 3.3.

$$d_j = \sqrt{(x_{amostra_j} - \mu_j)^2}. \quad (3.3)$$

Ao fazer o passo acima, imaginemos que se tem um deslocamento de valor 3. O que se está a fazer, é perceber qual a média dessa distância a todas as distâncias do histograma relativas à mesma articulação, o que no fundo irá dar o desvio do ponto a todo o histórico. O cálculo foi realizado para todas as distâncias e em cada classe, num determinado *dataset*.

Pegando nos cálculos já realizados até aqui, resta então calcular a distância total do conjunto de imagem, através da equação 3.4, onde são contabilizadas apenas as articulações que possuem suficiente número de pontos com boa visibilidade, referida em 3.2.1 como saída ao algoritmo que estamos a usar:

$$\frac{1}{J} \sum_{j=1}^J d_j, \quad (3.4)$$

onde  $J$  representa o número total de articulações contabilizadas.

### 3.4.2.1 Classificador

Devido à ausência de novos dados que permitissem testar o modelo sobre os mesmos, neste trabalho procedeu-se à etiquetagem de deslocamentos nos conjuntos usados para preparar o modelo. Para cada uma das distâncias em (3.3), volta-se a aplicar o mesmo cálculo. Contudo, desta vez será usado o valor de média da mesma articulação mas das outras classes que compõem o *dataset*, obtendo-se outros  $n$  valores de distância para cada par de *frames*, num conjunto com  $n$  classes. Daqui parte-se do princípio que a menor distância, no final, irá corresponder à classe prevista pelo modelo, e é dessa forma que cada deslocamento é classificado. Dentro de cada classe iremos ter a classificação de cada distância, que será determinada como classificação correta ou incorreta, uma vez que estamos a aplicar um método de classificação supervisionada.

## RESULTADOS EXPERIMENTAIS

Este capítulo descreve os resultados experimentais, apresentando-se os testes feitos e seguindo a explicação feita no capítulo anterior. Os mesmos irão validar toda a análise de imagem feita no âmbito desta dissertação e, por isso, os resultados vão ser comentados em todas as fases, comparando-os àquilo que seria esperado obter.

### 4.1 Ferramentas utilizadas

#### 4.1.1 Python

As soluções descritas no capítulo anterior e todo o modelo foi implementado em Python, uma linguagem de programação de alto-nível muito usada em desenvolvimento *web*, em criação de protótipos de *software*, entre muitas outras áreas. No âmbito deste trabalho, achou-se a ferramenta e solução mais viável a utilizar, devido ao grande enfoque em *Data Science* para fazer tratamento de dados, visualizações sobre eles e construir modelos.

Esta linguagem é também frequentemente conciliada com algoritmos de *Machine Learning*, tal como já foi mencionado no capítulo 3, levando à aprendizagem automática de modelos.

##### 4.1.1.1 Matplotlib

A biblioteca *Matplotlib* foi bastante utilizada ao longo da implementação, para criar as sucessivas visualizações sobre os dados trabalhados. É bastante fácil de utilizar e as suas inúmeras propriedades, permitindo controlar diversos detalhes visuais para melhor interpretação dos dados, levaram à sua preferência.

Esta biblioteca foi usada para traçar histogramas para caracterização dos cenários e testes que foram analisados, com a função *pyplot.hist* onde foi comum o uso dos parâmetros:

- *x*: valores de entrada. Pode ser introduzido um único vetor, ou uma sequência de vetores que não necessitam de ter o mesmo comprimento;

- *bins*: se é um inteiro, define o número de intervalos igualmente espaçados para classificar os dados do histograma. Pode receber também uma sequência, definindo os limites dos intervalos, em que os mesmos podem ser espaçados de forma desigual;
- *density*: se estiver a *True*, retorna uma densidade probabilística no histograma: cada classificação exibirá uma barra definida pela contagem da classificação dividido pelo número total de contagens mais a largura da barra;
- *rwidth*: representa a largura relativa das barras do histograma como uma fração da largura do intervalo.;

## 4.2 Introdução do algoritmo usado

Depois de alguns testes introdutórios feitos no início do trabalho, foi decidido usar o algoritmo Mediapipe Pose [49], onde ficou realçado, de imediato, o número de articulações que o mesmo é capaz de identificar, seja em imagens estáticas ou em conjuntos de *frames* em tempo real.

Começou-se por aplicar o mesmo algoritmo a qualquer um dos conjuntos de imagem já referidos. Usou-se um valor mínimo de confiança na deteção de 0.5, e foi feito o processamento em cada imagem, das partes do corpo referenciadas pelo Mediapipe Pose.

Em 4.1, temos os resultados destes primeiros testes, onde usando-se as diferenças dos valores de articulações entre frames, viu-se que era possível determinar a direção de um determinado movimento de um indivíduo, comprovando a utilidade e o número de estudos que o algoritmo poderia fornecer.

Na figura, as três imagens usam as coordenadas  $x$  e  $y$ , saídas ao Mediapipe Pose, do nariz, de forma a detetar a direção do movimento efetuado. Na imagem da direita, destaca-se que o algoritmo consegue detetar em forma de previsão pontos do corpo humano, ainda que estes estejam a ser ocultados pela cadeira presente.

Esta pré-análise à ferramenta que vamos utilizar no nosso estudo dá claramente a entender que a mesma pode ser usada não só sobre imagens estáticas, como funcionará de igual forma aplicada a quadros em tempo real.

A partir desta secção descrevem-se os resultados obtidos da análise de imagem feita a partir dos *datasets* iniciais, bem como os resultados que levaram à construção do classificador final. Será reportado o desempenho do mesmo, introduzido num algoritmo de classificação supervisionada, como referido no capítulo 3.

Este treino para um modelo final a adotar, foi feito para ser possível aplicar a cada *dataset*, ou *sub-dataset* dentro de cada um, existindo a divisão por classes de movimento, tal como é implícito pelo tema desta dissertação. Desta forma, uma vez que o modelo foi construído com base no uso de imagens estáticas, o mesmo fica apto na sua totalidade para ser aplicado a futuras análises de imagem em tempo real.

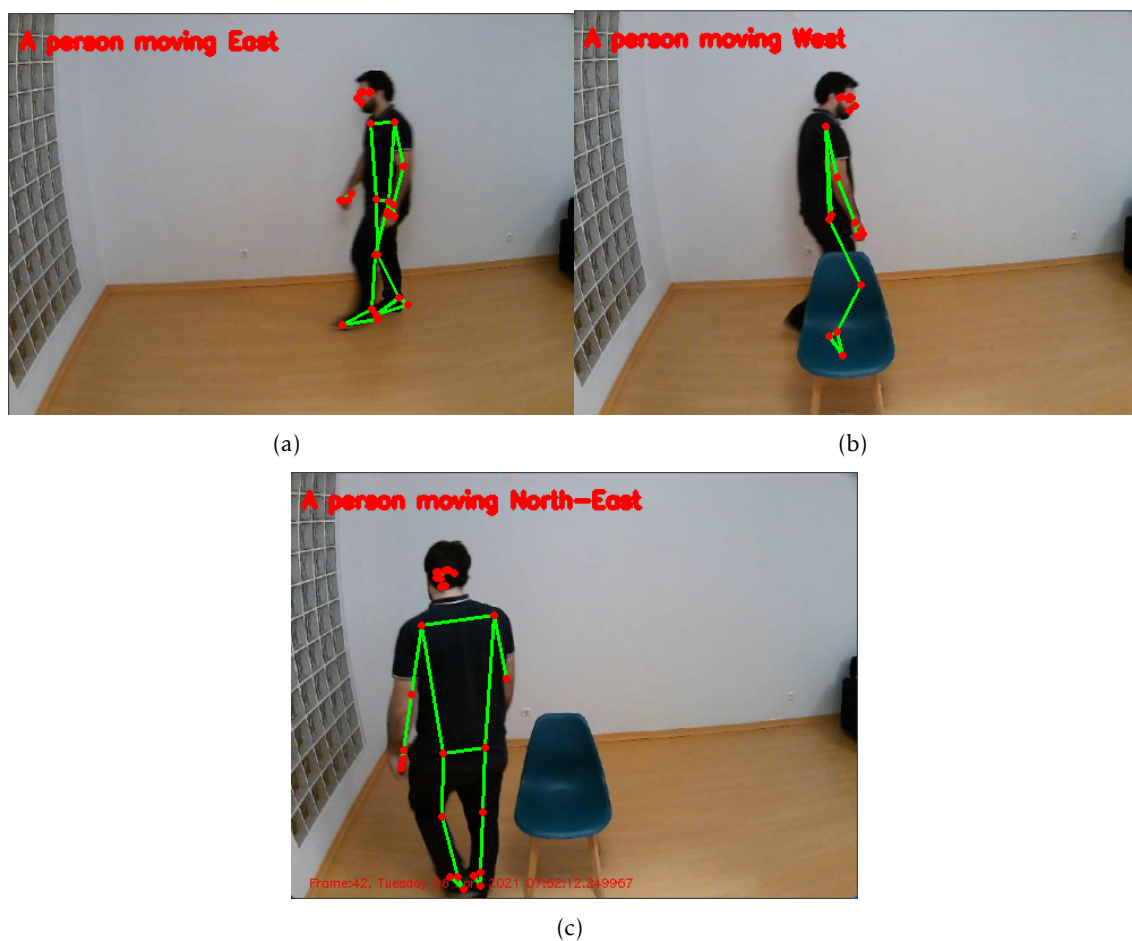


Figura 4.1: Análise ao movimento usando o algoritmo Mediapipe Pose.

### 4.3 Etiquetagem e preparação do estudo

Já foi introduzido que este trabalho inclui quatro tipos de movimento, identificados na tabela 4.1, onde dentro de cada um deles temos a formação de *sub-datasets*.

Tabela 4.1: *Datasets* trabalhados.

<i>Dataset case</i>	Movimento
12	Levantar e sentar numa cadeira
33	Sentado, levantar e baixar um ou os dois braços
44	Andar de um lado para o outro numa sala
55	Andar em torno de uma cadeira

Em cada um dos conjuntos, cada *frame* foi previamente etiquetada com o número da ação correspondente, ilustradas na tabela 4.2. Desta forma, foi possível trabalhar os dados e obter cada resultado demonstrado mais à frente neste capítulo.

Por fim, foram retiradas e guardadas, para cada um dos diferentes tipos de *dataset* as imagens com a identificação dos pontos efetuada pelo Mediapipe Pose. Este passo foi

Tabela 4.2: Movimentos possíveis.

Número da ação	Ação
1	Em pé
2	Sentar na cadeira
3	Levantar da cadeira
4	Estar sentado na cadeira
5	Estar sentado e levantar uma ou as duas mãos
6	Sentado e com a(s) mão(s) levantada(s)
7	Estar sentado e baixar uma ou as duas mãos
8	Caminhar

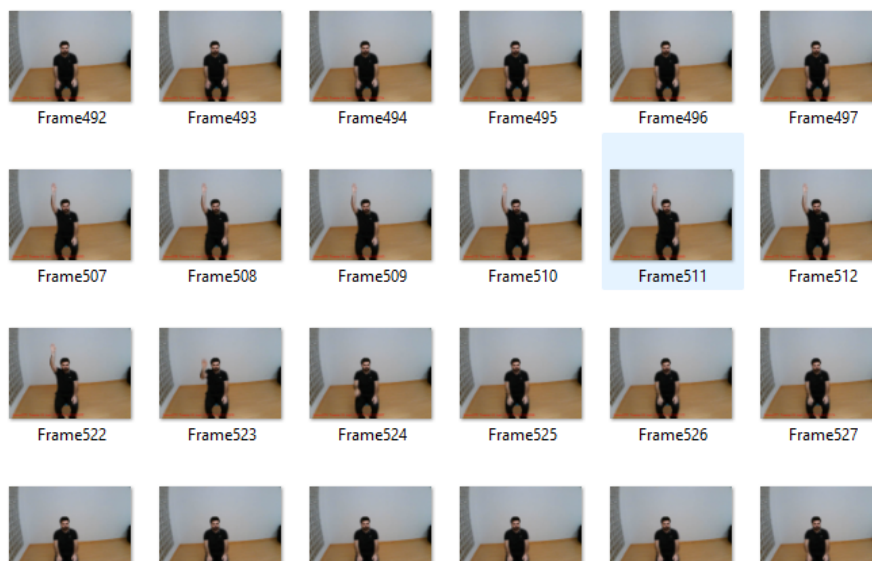
importante para uma análise visual ao algoritmo, pois à partida, caso os pontos batessem certo com o esperado em cada uma das imagens, então seria de imediato um bom indicador inicial.

Todos os valores das variáveis de saída (identificadas na subsecção 3.2.1), foram retirados e registados em formato *csv*, estando então prontos para a realização dos testes.

#### 4.3.1 Preparação dos *datasets*

Para um estudo conciso, claro e sequencial ao mesmo tempo, foi dado ênfase ao conjunto composto pela classe de estar sentado, e pela classe de levantar e baixar o braço, sendo esse o *dataset* presente em todos os testes que irão ser descritos.

Temos então um conjunto inicial vasto de imagens, ilustrado na figura 4.2, cuja enumeração está visível no nome das mesmas.

Figura 4.2: *Dataset* de imagem.

As duas classes referidas e que foram trabalhadas neste *dataset*, estão identificadas nas figuras 4.3 e 4.4.

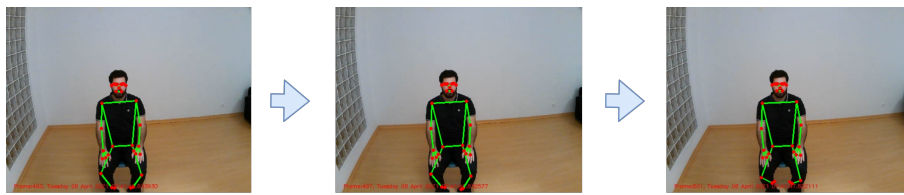


Figura 4.3: Indivíduo sentado.

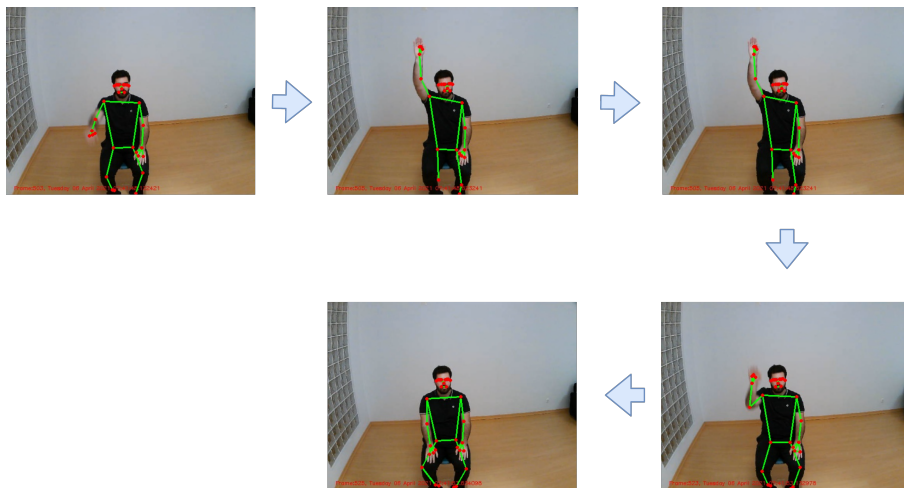


Figura 4.4: Indivíduo a levantar e baixar a mão.

Os resultados das duas classes integrantes são sucessivamente comparados, com objetivo de através das distribuições se conseguir identificar padrões que identificam o movimento.

No *dataset* no seu todo, as classes referidas anteriormente repetem-se várias vezes, intercalando entre si, e essas sub-classes são separadas levando a uma construção de resultados exemplificativa visível na figura 4.5.

- folder frame\_492.0\_503.0\_Seated
- folder frame\_503.0\_525.0\_Raising and Lowering a Hand
- folder frame\_526.0\_541.0\_Seated
- folder frame\_542.0\_566.0\_Raising and Lowering a Hand
- folder frame\_567.0\_582.0\_Seated
- folder frame\_583.0\_605.0\_Raising and Lowering a Hand
- folder frame\_606.0\_622.0\_Seated
- folder frame\_623.0\_645.0\_Raising and Lowering a Hand
- folder frame\_646.0\_661.0\_Seated
- folder frame\_662.0\_685.0\_Raising and Lowering a Hand

Figura 4.5: Exemplo da construção de resultados.

Dentro de cada conjunto de resultados, os mesmos estão organizados por articulação e saía ao algoritmo do Mediapipe Pose, tal como está presente na figura 4.6.

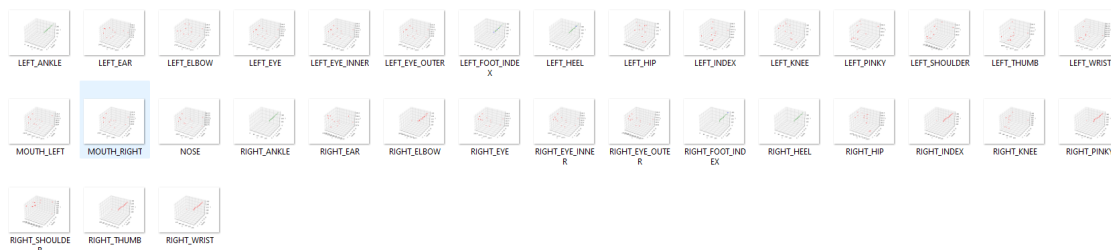


Figura 4.6: Distribuições retiradas por articulação.

Uma vez que o número de distribuições geradas em cada passo é elevado, sempre referentes a cada articulação e para uma determinada sub-classe se for o caso, apenas serão ilustradas algumas que sirvam para comprovar os objetivos mencionados.

#### 4.4 Visualização gráfica das saídas do modelo

O algoritmo do Mediapipe Pose para além das coordenadas  $x$  e  $y$  indicadoras da posição para cada articulação, possui um valor de visibilidade, entre 0 e 1, que indica a confiabilidade que podemos ter na identificação de determinada articulação numa dada *frame*. Foi definido que valores de visibilidade acima de 0.95 são pontos de grande visibilidade e por isso foram usados na continuação dos testes. Valores entre 0.5 e 0.95 são constituintes de visibilidade intermédia e, por fim, temos os valores menores que 0.5 indicadores de fraca visibilidade e confiança. Estes valores são fornecidos pelo próprio algoritmo.

De seguida, representam-se alguns dos *plots* tirados para identificação da visibilidade de cada ponto identificado em cada articulação. Nos *plots*, os pontos assumem cores de acordo com a tabela 4.3.

Tabela 4.3: Cor atribuída a cada ponto dos *plots* consoante a visibilidade.

Cor	Representação
Vermelha	Pontos de visibilidade superior a 0.95
Verde	Pontos de visibilidade entre 0.5 e 0.95
Azul	Pontos de visibilidade inferior a 0.5

Dependendo da articulação identificada, e do ângulo da captura de imagem, a maior parte das articulações possuem todos os pontos com elevada visibilidade. Contudo, existem outras mais críticas, as quais não irão ter qualquer influência nos resultados. Na generalidade dos casos, foi comprovado que os pontos com grande visibilidade mantêm o indicador ao longo de todo o conjunto em análise, sendo por isso confiável a utilização dos mesmos. Em sentido contrário e da mesma maneira, as articulações críticas que surgem em menor número, mantêm também os seus valores ao longo do *dataset*.

• Sub-classe identificada na figura 4.3 por "frame\_492.0\_503.0\_Seated"

Este sub-conjunto pertence à classe de uma pessoa estar sentada, em análise, e desta forma o número de pontos visíveis em cada um dos gráficos, da figura 4.7, corresponde ao número de *frames* da sub-classe.

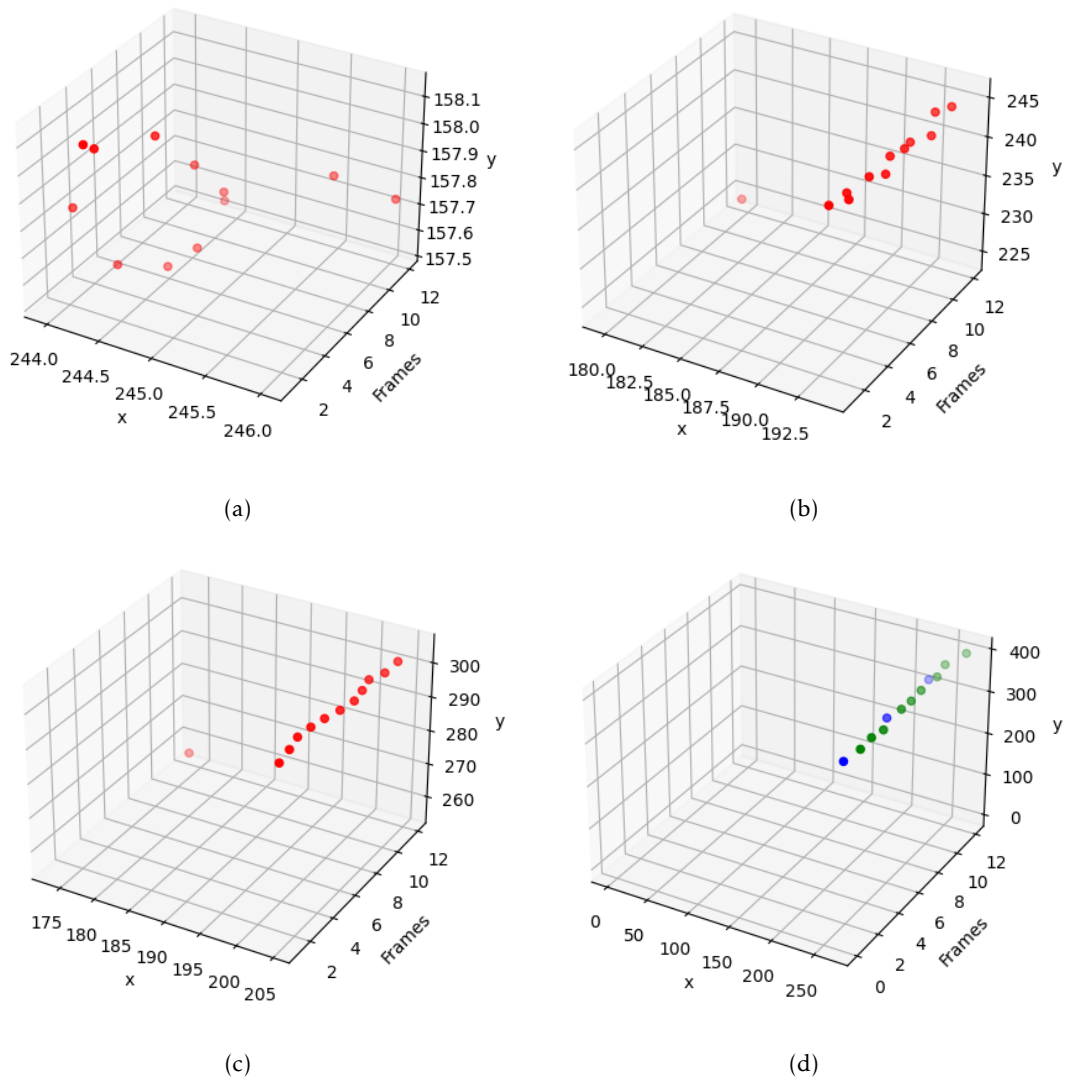


Figura 4.7: Em (a) temos o gráfico de pontos referente à orelha esquerda do indivíduo, em (b) referente ao cotovelo direito, em (c) ao indicador da mão direita e, por fim, o gráfico (d) correspondente ao indicador do pé esquerdo.

Pela figura acima, é de notar que para o cotovelo direito em (b), ou para o indicador da mão esquerda em (c), temos valores de coordenadas praticamente constantes ao longo das imagens, o que é normal, uma vez que o indivíduo está sentado com o braço em repouso. Por outro lado, se olharmos à orelha esquerda, apesar de os pontos permanecerem praticamente constantes, existem variações ligeiramente maiores do que o gráfico em (b), uma

vez que mesmo em repouso, a cabeça do indivíduo pode estar em ínfimos movimentos (parte do corpo mais instável do que um braço em repouso).

O gráfico que resta, em (d), estão visíveis pontos verdes e azuis, de pontos de visibilidade entre 0.5 e 0.95, e abaixo de 0.5, respetivamente, pelo que será uma articulação que desde logo é descartada na análise das distribuições para o conjunto de *frames* em questão.

- **Sub-classe identificada na figura 4.3 por "frame\_503.0\_525.0\_Raising and Lowering a Hand".**

Esta sub-classe engloba o conjunto de imagem desde que o indivíduo começa a levantar o braço direito, o pequeno tempo em que permanece com a mesma levantada e até ao momento em que volta a baixá-la, tal como ficou mostrado na figura 4.4.

Tal como ficou feito na classe anterior, foram analisados os *outputs* ao algoritmo do Mediapipe Pose para quatro articulações do corpo do indivíduo - figura 4.8 -, começando-se por analisar a visibilidade dos pontos e em seguida as coordenadas dos mesmos.

É desde logo observável que, relativamente à sub-classe anterior, os pontos apresentam níveis de visibilidade praticamente idênticos para as mesmas articulações. No entanto, há diferenças na distribuição dos pontos nos eixos das coordenadas, tal como era expectável.

Em (a), a análise mantém-se face ao gráfico (a) da figura 4.7. Da mesma maneira, o gráfico (d) volta a apresentar pontos de fraca visibilidade e tal como foi dito, estas articulações com fracos níveis de confiança na deteção dos pontos mantém-se, e por isso serão descartadas de qualquer tipo de análise.

As maiores diferenças surgem nos gráficos (b) e (c), relativos ao cotovelo direito e ao indicador direito, respetivamente. Os dois são semelhantes, já que apresentam parte dos pontos com valores de coordenadas  $x$  e  $y$  constantes, que dizem respeito ao tempo em que a pessoa permanece com o braço levantado. No entanto, veem-se cinco pontos mais dispersos dos outros, em cada um dos gráficos, que são indicadores das alturas em que se está a levantar ou a baixar o braço. Uma vez que estão em menor número, em relação aos outros pontos em que as coordenadas são praticamente iguais, é válido dizer que mesmo que não se soubesse que estávamos em pontos de levantar ou baixar o braço, estas ações poderiam ser identificáveis pois são repentinas.

##### 4.4.1 Gráficos diferenciais

É importante salientar, que todo o tipo de resultados a partir daqui, é feito apenas contabilizando os pontos em que a visibilidade é superior a 0.95. Se fossem considerados os outros pontos, iríamos estar a aplicar um algoritmo em que a fiabilidade dos resultados seria menor, onde não teríamos a certeza se a identificação dos pontos em questão, estaria a ser efetuada da forma mais precisa.

Tomando os mesmos, foram gerados gráficos para cada uma das articulações identificadas pelo Mediapipe Pose, com as distâncias efetuadas entre pares de *frames*. Estas

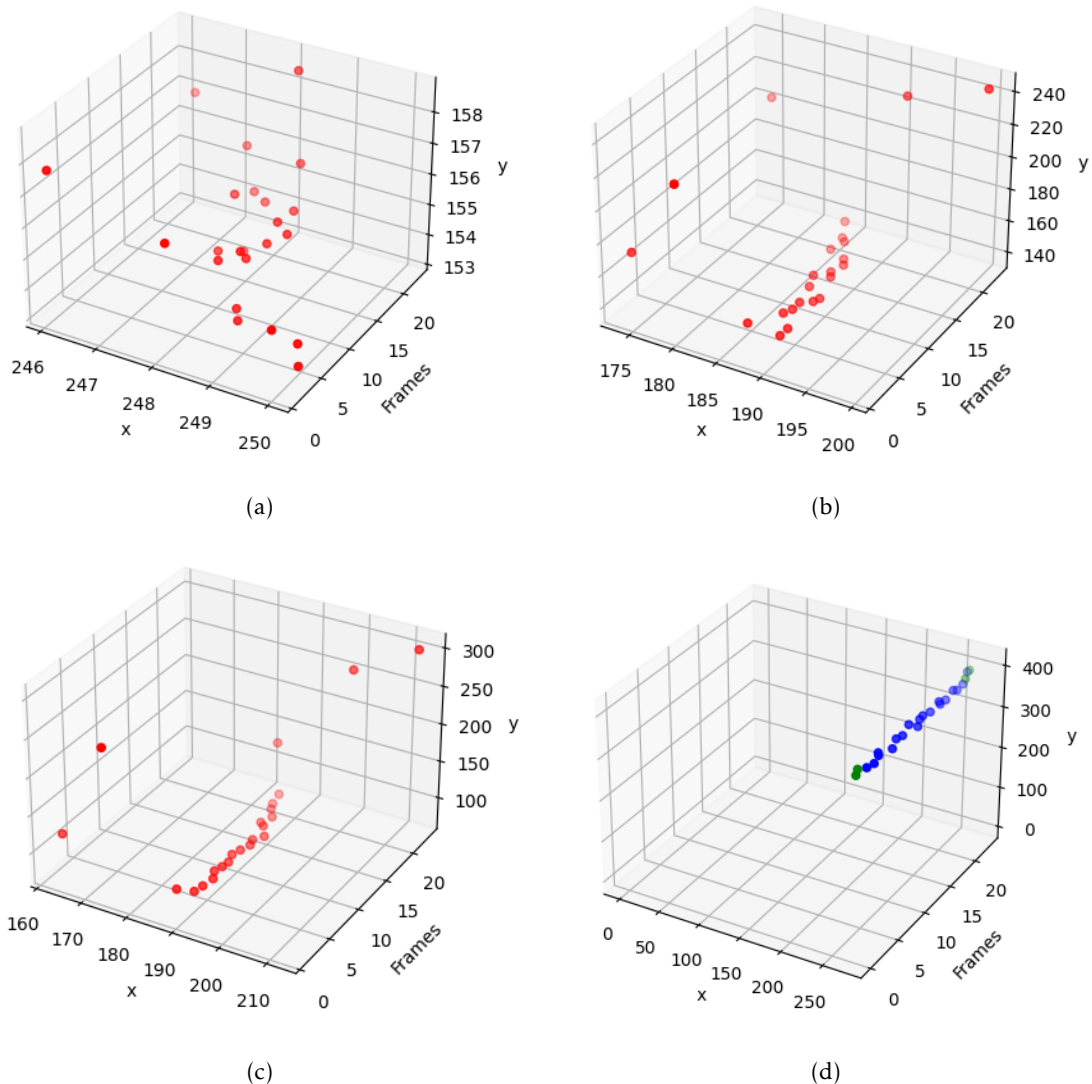


Figura 4.8: Em (a) temos o gráfico de pontos referente à orelha esquerda do indivíduo, em (b) referente ao cotovelo direito, em (c) ao indicador da mão direita e, por fim, o gráfico (d) correspondente ao indicador do pé esquerdo.

distribuições foram aplicadas ao *dataset* de imagem inteiro, englobando as duas classes em estudo, e podem ser úteis para o estudo da influência de cada uma das articulações na identificação do conjunto previamente etiquetado, assim como para concluir se no geral esta análise dos pontos é confiável para o estudo de todo o conjunto de imagem.

Começamos por ver as "distribuições de distância" para o indicador da mão esquerda na figura 4.9(a), e para o indicador da mão direita na figura 4.9(b). Para alguém que está a fazer um estudo a estes gráficos, com objetivo de etiquetar o *dataset*, pode-se afirmar que enquanto para a mão esquerda em (a), os pontos estão praticamente todos concentrados na mesma zona com um gráfico com pequenas variações na escala tanto no eixo  $x$  como no  $y$ , na mão direita o mesmo não acontece. A maioria dos pontos voltam a agrupar-se na mesma zona formando uma "mancha" mais visível que no gráfico em (a), isto porque cada

um dos eixos já referidos tem agora escalas maiores, resultado de termos também pontos com coordenadas maiores (distâncias efetuadas entre pares de imagens maiores). Assim, logo aqui pode-se constatar que no caso da mão direita, ela está e repouso na maioria do tempo - lembrando que pode ser tanto quando o braço está em baixo como quando está levantado -, tendo certos movimentos mais bruscos, que para alguém que esteja a ver os gráficos indica movimentos repentinos que são efetuados. São realizados não só uma vez, mas sucessivamente, informação dada pelo eixo das "Frames".

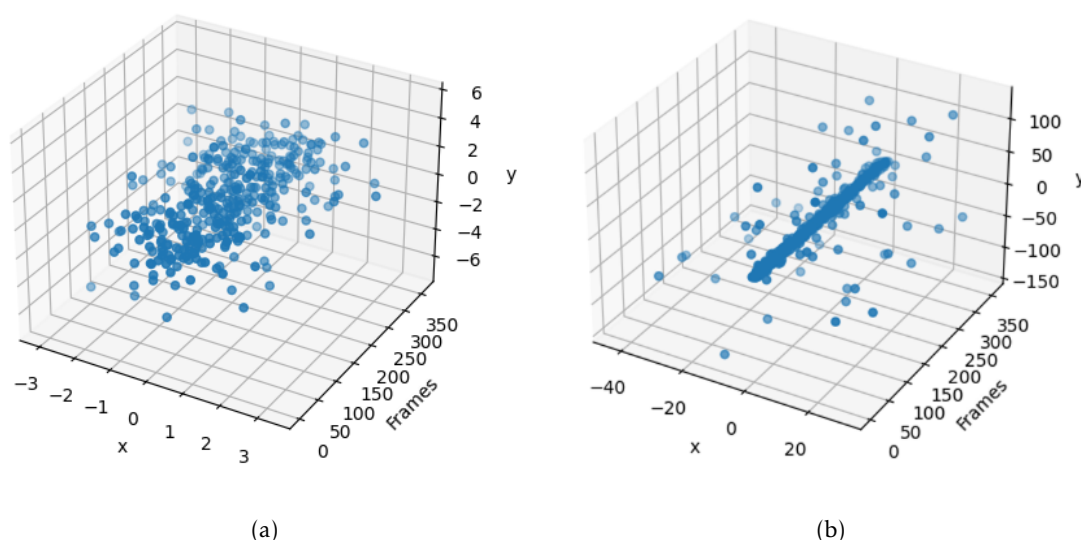


Figura 4.9: Gráfico de distâncias do indicador da mão esquerda, em (a); em (b) para o indicador da mão direita.

Na figura 4.10 são mostrados mais gráficos obtidos para outras articulações.

Desta vez em 4.10(a) e em 4.10(b) temos os gráficos para o ombro esquerdo e para o ombro direito, respetivamente. De forma auxiliar aos gráficos já mencionados da figura 4.9, o lado direito volta a apresentar distâncias efetuadas maiores em relação ao lado esquerdo, devido a ser levantado o braço direito. Contudo, naturalmente, estas distâncias são menores que distâncias presentes e evidenciadas na distribuição dos pontos para o indicador da mão esquerda, que faz um movimento maior em relação ao ombro.

Em 4.10(c) e em 4.10(d), gráficos relativos ao joelho esquerdo e ao joelho direito, respetivamente, as distâncias são muito idênticas entre si, o que é indicativo que os membros inferiores do corpo não estão a fazer nenhum tipo de movimento significativo no conjunto de imagem em estudo.

Em 4.11(a) temos o gráfico diferencial para o nariz da pessoa, mostrando que a sua maioria faz movimentos praticamente nulos, no entanto conseguimos ver pontos mais afastados do centro, que muito provavelmente se referem a alturas do movimento do braço, onde a cabeça acaba por acompanhar com um ligeiro movimento. Por fim, 4.11(b) foi tirado para o indicador do pé direito do indivíduo, que mostra que para esta articulação não existem pontos com visibilidade superior a 0.95.

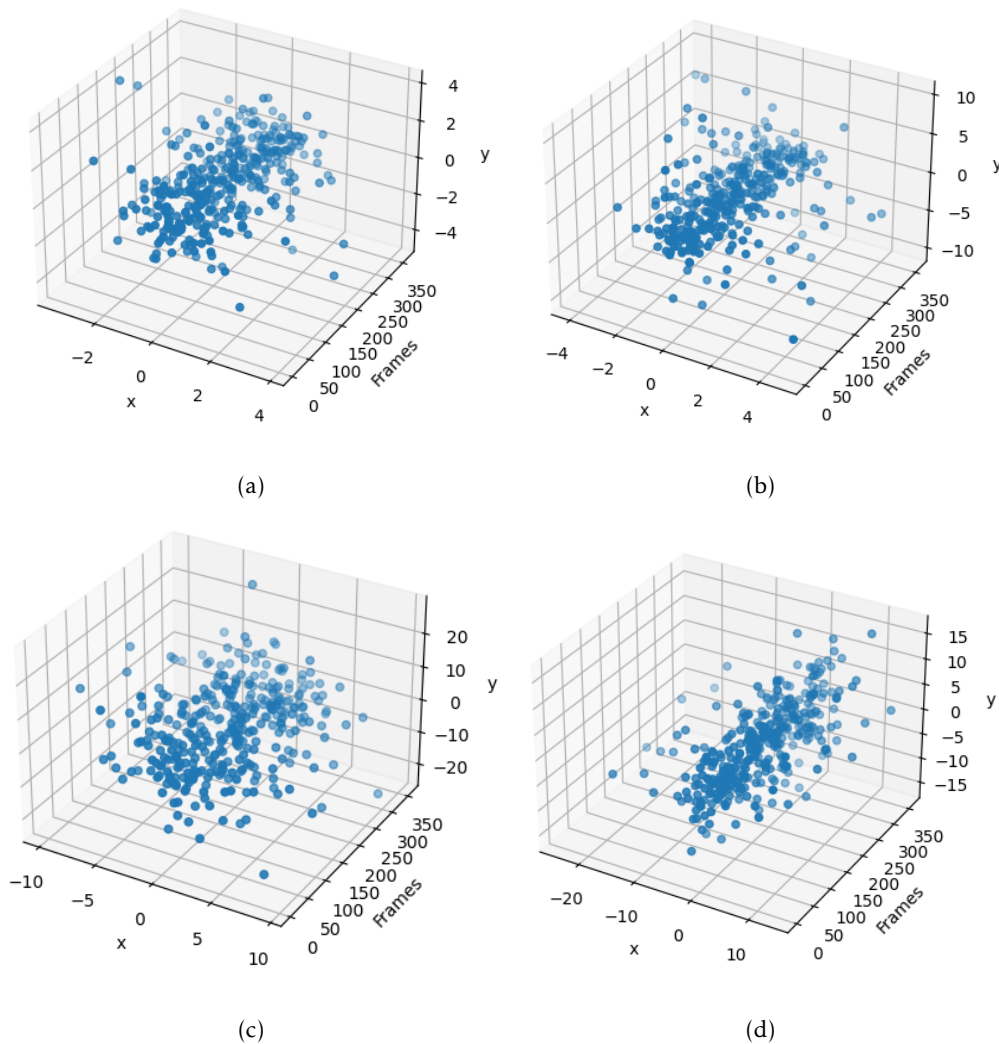


Figura 4.10: Gráfico de distâncias para os ombros e para os joelhos do indivíduo nas classes envolvidas.

#### 4.4.2 Influência geral de cada articulação

Antes de se partir para os próximos resultados tirados através dos vetores deslocamento e angular, é importante ter a percepção da influência geral das articulações no movimento respectivo. No caso atual, sabemos que estamos no *dataset* de um indivíduo estar sentado e levantar o braço direito sucessivamente ao longo de todo o conjunto de imagem, e cada vez que dada junção apresentou um deslocamento vetorial nas saídas  $x$  e  $y$  ao modelo acima de um *threshold* definido de 5 pixels, adicionou-se um a essa articulação. Essas "pontuações" foram contabilizadas no final, como o número de vezes que cada uma das articulações se moveu acima do *threshold*, dando origem ao gráfico de barras presente na figura 4.12.

Os números das articulações provêm da legenda da figura 3.1, no entanto para facilitar a leitura, os números encontram-se descritos na tabela 4.4.

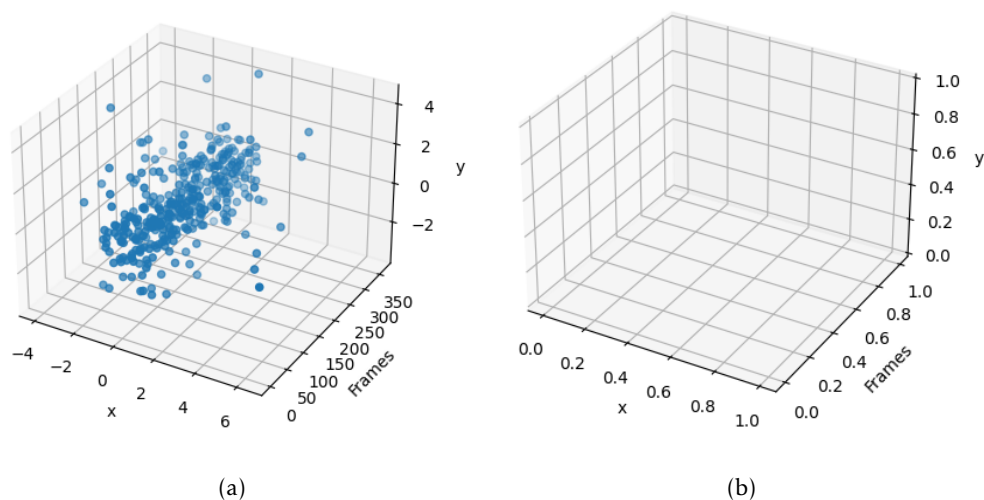


Figura 4.11: Gráfico de distâncias para o nariz e para o indicador do pé direito nas classes envolvidas.

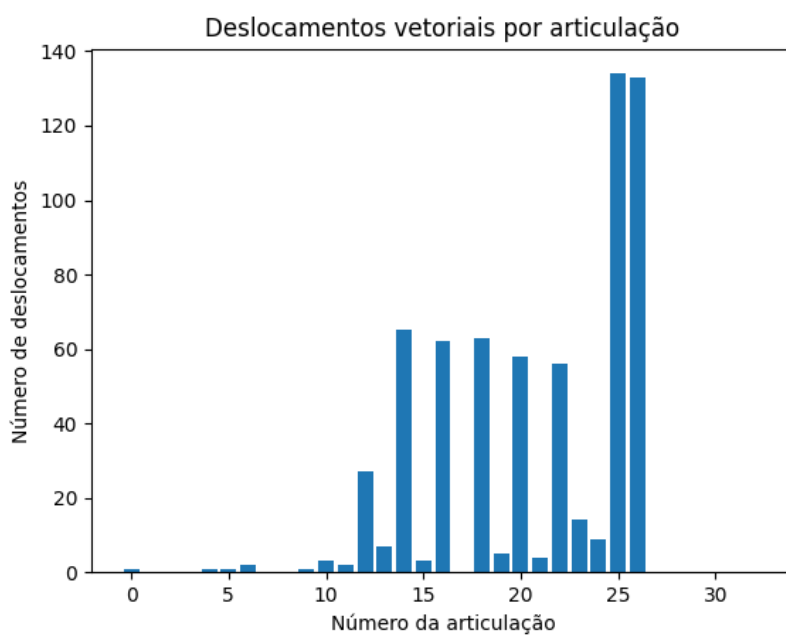


Figura 4.12: Deslocamentos vetoriais acima de um *threshold* definido de 5, para cada uma das articulações do Mediapipe Pose.

Pela figura 4.12, vemos que há uma predominância das articulações número 14, 16, 18, 20 e 22, tudo zonas pertencentes ao membro superior direito do indivíduo, pelo que os resultados são o esperado. Ainda assim, é de destacar a maior dominância para o joelho direito e para o joelho esquerdo. Isto deve-se ao facto de o *threshold*, neste caso, ainda não ser demasiado alto, uma vez que os joelhos representam zonas facilmente movíveis, que acabam por fazer ligeiros movimentos mesmo quando a ação não os envolve diretamente.

Na figura 4.13, foi aumentado o limite de deslocamento vetorial para 9 píxeis.

Tabela 4.4: Números associados a cada articulação.

Nº	Articulação	Nº	Articulação	Nº	Articulação
0	Nariz	11	Ombro esquerdo	22	Polegar direito
1	Interior olho esquerdo	12	Ombro direito	23	Anca esquerda
2	Olho esquerdo	13	Cotovelo esquerdo	24	Anca direita
3	Exterior olho esquerdo	14	Cotovelo direito	25	Joelho esquerdo
4	Interior olho direito	15	Pulso esquerdo	26	Joelho direito
5	Olho direito	16	Pulso direito	27	Tornozelo esquerdo
6	Exterior olho direito	17	Mindinho esquerdo	28	Tornozelo direito
7	Orelha esquerda	18	Mindinho direito	29	Calcanhar esquerdo
8	Orelha direita	19	Indicador esquerdo	30	Calcanhar direito
9	Lado esquerdo da boca	20	Indicador direito	31	Indicador pé esquerdo
10	Lado direito da boca	21	Polegar esquerdo	32	Indicador pé direito

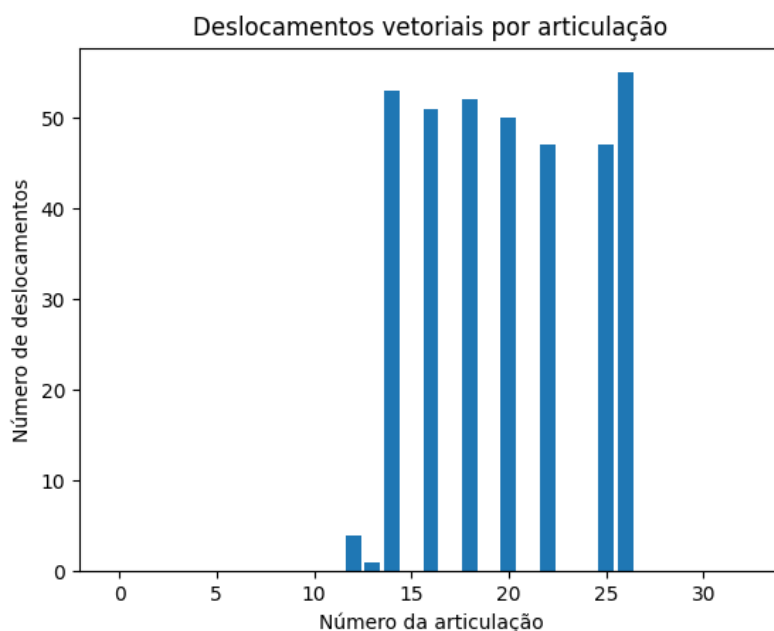


Figura 4.13: Deslocamentos vetoriais acima de um *threshold* definido de 9, para cada uma das articulações do Mediapipe Pose.

Vemos que os resultados são bastante mais fracos para as articulações 25 e 26 referentes aos joelhos. Enquanto que os valores das junções primeiramente referidas no braço direito permanecem com os mesmos valores, vemos que a escala das articulações 25 e 26 reduziu praticamente para metade.

## 4.5 Vetores deslocamento e angular

De forma a uma caracterização mais específica do movimento efetuado, que levará a uma análise mais pormenorizada do *dataset* de imagem, foram retirados histogramas em cada classe (sentado ou movimento do braço) acerca dos vetores de deslocamento (entre pares

de *frames*) e para a parte angular desses mesmos vetores, para cada uma das articulações em estudo. Esta análise de resultados foi feita em 2 partes, sendo uma primeira com a repartição nas sub-classes, isto é, por conjuntos de imagem situados entre as transições entre as classes, tal como já havíamos visto na figura 4.5. A segunda parte é composta pelo conjunto das sub-classes dentro de cada classe principal já referida, dando uma ideia global do movimento específico da classe.

#### 4.5.1 Componente $x$ e $y$ : análise por sub-classes

Tal como referido, neste tipo de análise e ao termos um *dataset* onde nove vezes a pessoa levanta o braço, o que faz termos dez outras vezes a pessoa estar simplesmente sentada, os resultados foram primeiramente tirados divididos pelas sub-classes mencionadas, ou seja, num *dataset* em que o indivíduo levanta e baixa o braço dez vezes, foram retirados os resultados em cada uma das dez sub-classes. O mesmo para o número de vezes em que a pessoa está meramente sentada (entre os movimentos do movimento do braço). Começou-se por fazer uma caracterização dos vetores deslocamento e angular através dos histogramas, para cada articulação, nos sub-conjuntos.

Uma vez que um vetor de deslocamento é composto pela componente em  $x$  e pela componente em  $y$ , o primeiro passo foi retirar os histogramas para cada componente em individual, servindo como parte necessária numa correlação futura com os histogramas para os vetores deslocamento, pois são os resultados em  $x$  e em  $y$  que conseguirão estimar a direção tomada pelo movimento. Para um estudo mais completo, juntaremos ainda a informação dos vetores angulares mais à frente, o que conciliado com tudo o resto dar-nos-á indicações sobre possíveis sentidos do movimento.

Esta análise das coordenadas em individual foi feita apenas ao nível das sub-classes. Escolhido um sub-conjunto em que a pessoa está apenas sentada, e outro sub-conjunto aquando o movimento do braço direito por parte da mesma, na figura 4.14, temos os gráficos para o deslocamento em  $x$  e em  $y$  do indicador direito do indivíduo, para cada uma das classes.

No cenário em que a pessoa está somente sentada, em 4.14(a) e em 4.14(b), o mesmo reflete-se no facto de tanto em  $x$  como em  $y$  não termos praticamente nenhum deslocamento na imagem.

Por outro lado, em 4.14(c) e em 4.14(d) relativos ao sub-conjunto formado pelo movimento do braço, vemos que em  $y$  já temos a presença de deslocamentos mais alargados, uma vez que o braço se está a mover de baixo para cima e depois de cima para baixo. Temos, contudo, uma maior concentração das barras do histograma junto ao 0, que dizem respeito à altura em que o braço permanece levantado. Em  $x$  temos também mais deslocamento que o gráfico em 4.14(a), o que é normal pois no movimento do braço em altitude, estamos ao mesmo tempo suscetíveis a ter maiores variações se olharmos a um eixo imaginário sob a largura de imagem. Tal como em 4.14(d), a maior concentração do gráfico acontece junto ao valor nulo, também devido à estaticidade quando o braço está

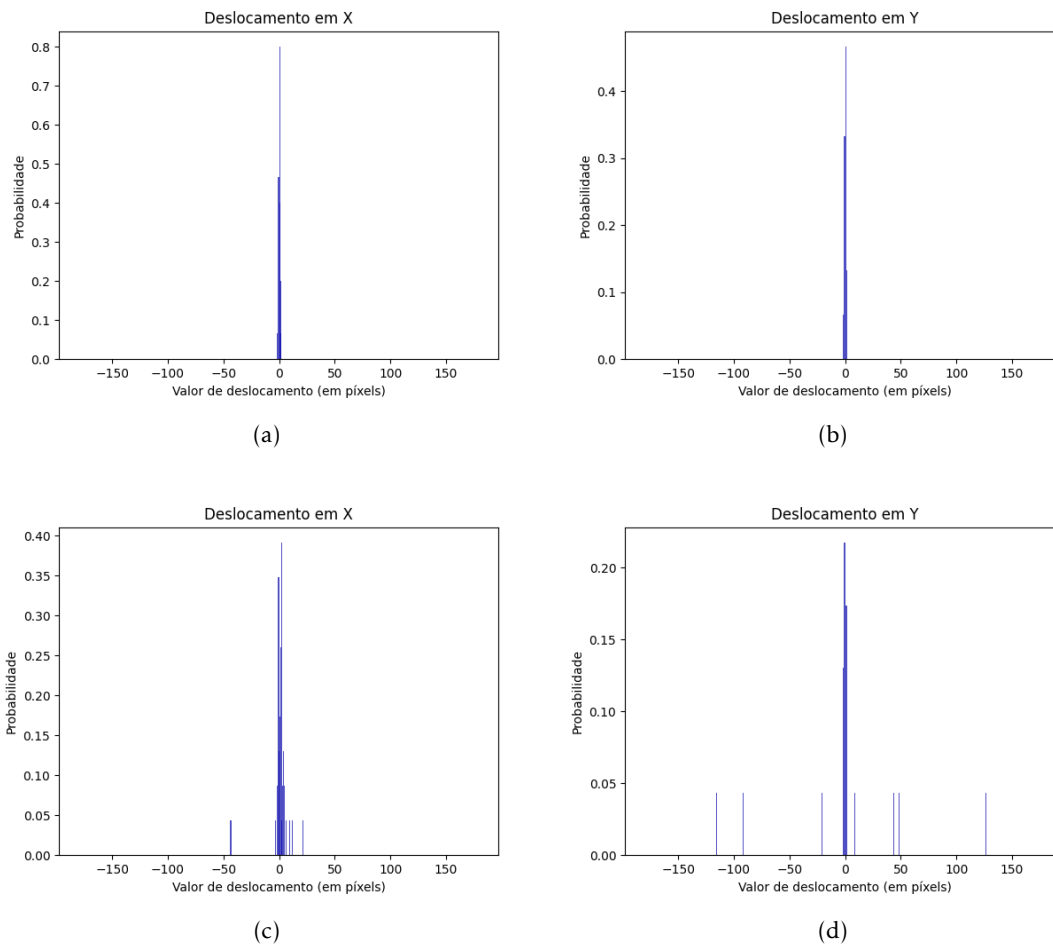


Figura 4.14: Em (a) e em (b), o deslocamento em  $x$  e em  $y$ , respetivamente, do indicador direito para uma altura em que o indivíduo está sentado. Em (c) e em (d), o deslocamento em  $x$  e em  $y$ , respetivamente, também do indicador direito e para o cenário em que o indivíduo está a levantar o braço direito.

levantado.

Analisando então de forma geral o comportamento do movimento assente no indicador direito da pessoa, é constatável que na classe em que a pessoa está a mover o braço, em  $y$ , vamos ter maiores variações de deslocamento do que em  $x$ . O mesmo acaba por ser normal uma vez que o movimento do braço está a ser feito para cima, isto é, para uma posição mais alta na imagem. Em  $x$  acabam por ser visíveis algumas pequenas barras que parecem querer afastar-se do centro, ainda que próximas do mesmo, relativas a pequenos deslocamentos realizados. Muito provavelmente são referentes a alturas em que o braço está levantado, o que naturalmente provoca pequenos movimentos ainda que quase impercetíveis, se olharmos a um eixo horizontal que atravessa a imagem.

Este tipo de análise poderia estar a ser feita a qualquer tipo de articulação, mas uma vez que temos a consciência enquanto investigadores, do movimento que se trata, foi selecionado o indicador direito que retratará bem as características das duas classes em

análise. Os gráficos foram normalizados, e por estarmos numa análise por sub-classes, cada gráfico contém um número reduzido de imagens e por isso as barras acabam por ser bastante visíveis e facilmente identificadas em cada valor específico.

#### 4.5.2 Vetores deslocamento e angular: análise por sub-classes

O deslocamento em  $x$  e o deslocamento em  $y$  são os constituintes do deslocamento vetorial. Ao termos o deslocamento em  $x$  e em  $y$  entre cada par de imagens, temos o deslocamento vetorial também, isto sempre para cada articulação do algoritmo, tal como já é sabido. O mesmo é sempre um vetor positivo, com um valor de distância, uma vez que é calculado pela equação 4.1:

$$d = \sqrt{d_x^2 + d_y^2}, \quad (4.1)$$

em que  $d_x$  é o deslocamento em  $x$  e  $d_y$  é o deslocamento em  $y$ .

Ao mesmo tempo, o deslocamento angular auxilia o nosso estudo sobre o movimento, representando a componente angular do vetor deslocamento referido anteriormente.

Para a mesma articulação (indicador direito) com grande influência no nosso *dataset*, e para os mesmos sub-conjuntos de imagem, na figura 4.15(a) está representado o histograma que caracteriza os vetores de deslocamento efetuados para o sub-conjunto em que está sentado, sendo que os mesmos vetores mas agora para o cenário do movimento do braço, estão representados na figura 4.15(b).

Nas figuras 4.15(c) e 4.15(d), temos na mesma o cenário sentado e o cenário de mexer o braço, respetivamente, mas agora com os histogramas que caracterizam os deslocamento angulares efetuados.

Os histogramas da figura 4.15 foram normalizados com o parâmetro booleano "*density*" ativado. Começando pelos dois primeiros gráficos mostrados, (a) e (b), é nos realçado logo à vista que para o cenário em que a pessoa está apenas sentada, temos barras grossas situadas próximas do zero. Significa que são muito reduzidos e praticamente nulos os deslocamentos do indicador esquerdo, no cenário em que o ser humano está sentado. Para a situação em que a pessoa move o braço, vemos barras já muito mais finas, o que quer dizer que temos uma maior amplitude na diferença entre o deslocamento mais baixo e o deslocamento mais elevado, no sub-conjunto. Mais uma vez, como na sub-classe o indivíduo acaba por estar mais tempo com o braço estático levantado, do que propriamente a movê-lo, então é normal que voltemos a ter uma maior densidade dos resultados do histograma junto a valores de deslocamento mais pequenos.

Nas figuras 4.15(c) e 4.15(d), temos os deslocamentos angulares dos vetores deslocamento analisados no parágrafo anterior. Por norma nos histogramas com os valores das componentes angulares dos vetores deslocamento, a escala no eixo da "*Frequência*" é menor, já que acabamos por ter maior variedade de ângulos, resultado de ser um sub-conjunto com mais movimento, ou mais instável.

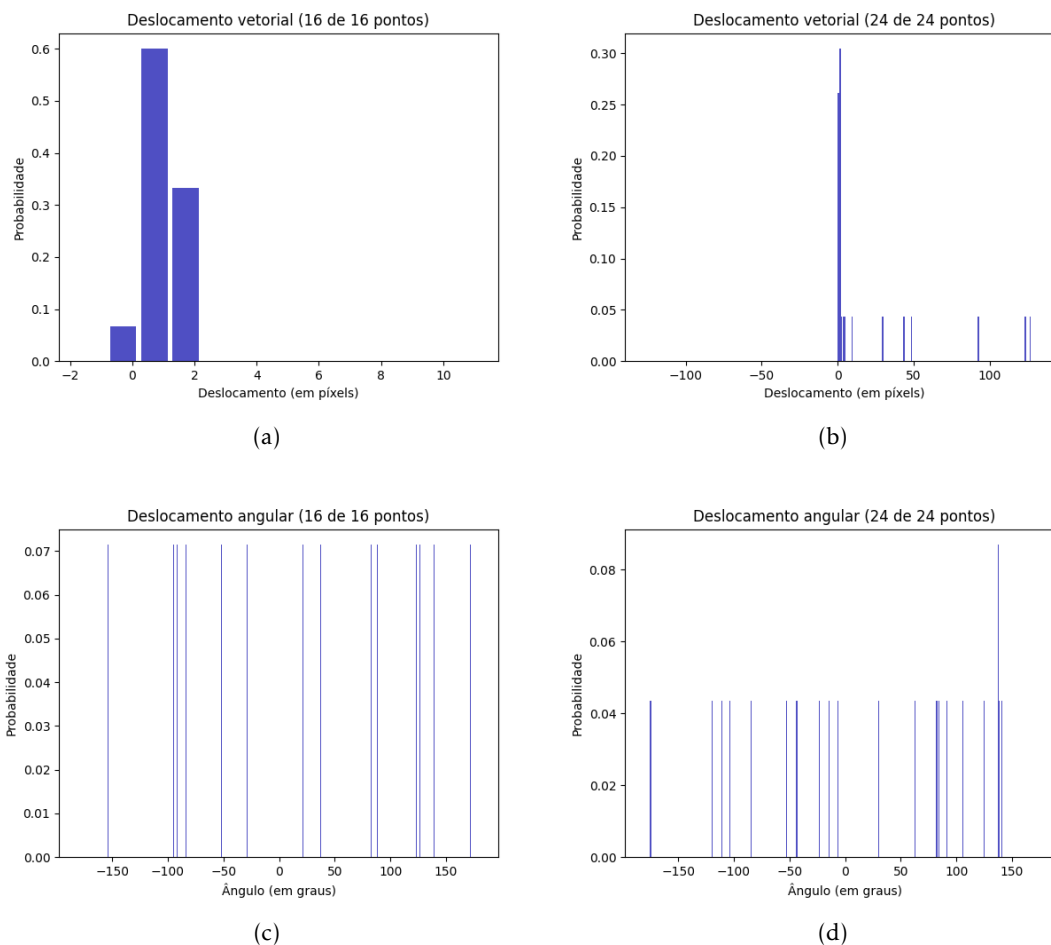


Figura 4.15: Análise por sub-classe: em (a) e em (b), os histogramas para os vetores deslocamento no caso em que o indivíduo está sentado e no caso em que está a mexer o braço, respetivamente. Em (c) e em (d) exatamente a mesma ordem de cenários que em (a) e em (b) mas agora para os deslocamentos angulares efetuados.

De destacar que para formar estes gráficos, temos sempre o total de pontos do sub-conjunto disponível, ou seja, já vimos que para qualquer resultado apenas os pontos com visibilidade acima de 0.95 seriam contabilizados, e visto que no indicador esquerdo, como na maioria das outras articulações oferecidas pelo algoritmo, temos sempre grande confiança nos pontos obtidos, então todos os pontos (com coordenadas  $x$  e  $y$ ) em cada *frame*, são utilizados para se formarem vetores deslocamento. Como o cenário sentado em análise, é composto por 16 imagens, então temos 16 pontos disponíveis. O mesmo para o cenário do movimento do braço, uma vez que estamos a analisar um conjunto de 24 imagens, temos então 24 pontos à disposição.

### 4.5.3 Vetores deslocamento e angular: análise pelas classes completas

Ao juntarmos todas as sub-classes em cada *dataset* temos o estudo feito sobre cada articulação no conjunto inteiro. Esta análise dará uma ideia geral sobre o estudo, e servirá de base

à formação do classificador final, assente sobre algumas métricas que irão ser aplicadas.

Seguindo o que foi feito até aqui, na figura 4.16(a) e 4.16(b) temos os histogramas que caracterizam todos os vetores deslocamento realizados no conjunto de imagem total em estudo, para o indicador direito do indivíduo, para as classes em que está sentado e em que está a movimentar o braço, respetivamente. Por outro lado, nas figuras 4.16(c) e 4.16(d) temos representadas as mesmas classes, pela mesma ordem, mas para os deslocamentos angulares efetuados pelo indicador direito e que acompanham os vetores de deslocamento.

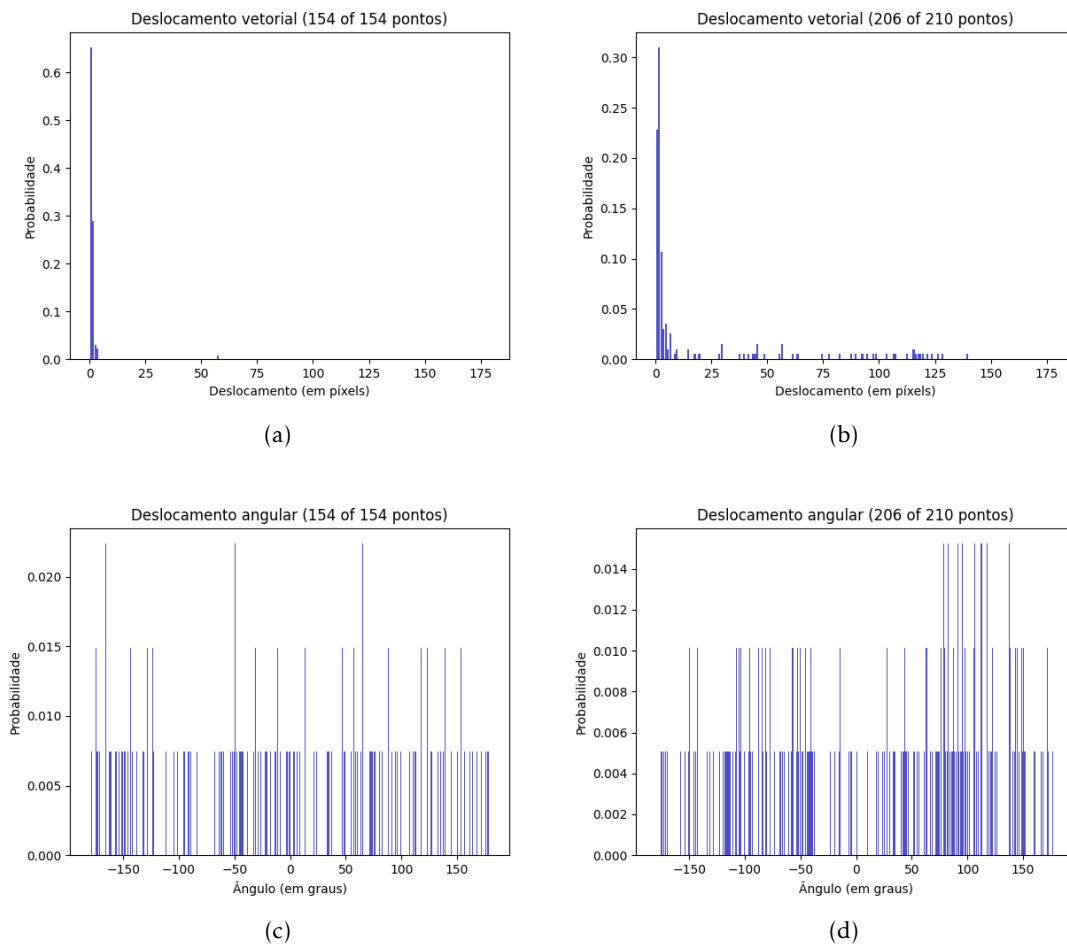


Figura 4.16: Análise na classe completa, para o indicador esquerdo da pessoa: em (a) e em (b), os histogramas para os vetores deslocamento no caso em que o indivíduo está sentado e no caso em que está a mexer o braço, respetivamente. Em (c) e em (d) exatamente a mesma ordem de cenários que em (a) e em (b) mas agora para os deslocamentos angulares efetuados.

De forma a clarificar mais a análise retirada na análise na sub-classe, estes histogramas já conseguem caracterizar melhor o *dataset* no seu todo. Enquanto que em 4.16(a) vemos que a amplitude do deslocamento em pixels é sempre bastante reduzida, pois estamos na classe em que o indivíduo está sentado, em 4.16(b) o mesmo volta a repetir-se e explicado

ao longo dos resultados pois o movimento do braço é sempre bastante "repentino". É por isso que temos barras até valores de deslocamento mais elevados, mas com probabilidade bastante mais reduzida que os deslocamentos praticamente nulos.

Em 4.16(c) e 4.16(d) é perceptível que a classe em que está sentado apesar de existirem ângulos provocados pelos deslocamentos bastante reduzidos, acabamos quase por ter um gráfico uniforme. Isto porque a pessoa está estática e não há grande instabilidade nos movimentos. Não tão uniforme temos o cenário em que levanta o braço, onde seria de esperar uma simetria do gráfico através do valor de deslocamento 0 uma vez que o ângulo efetuado pelo braço quando se está a levantar, acaba por ser simétrico ao ângulo que o mesmo faz quando se está a baixar. Ainda assim, movimentos mais rápido num dos sentidos, pode provocar ângulos maiores do que os movimentos se forem feitos de forma mais devagar. Esta simetria pode ser consolidada com a figura 4.17, onde temos os histogramas do deslocamento angular efetuado pelo mindinho direito, pelo pulso direito e pelo cotovelo direito, respetivamente.

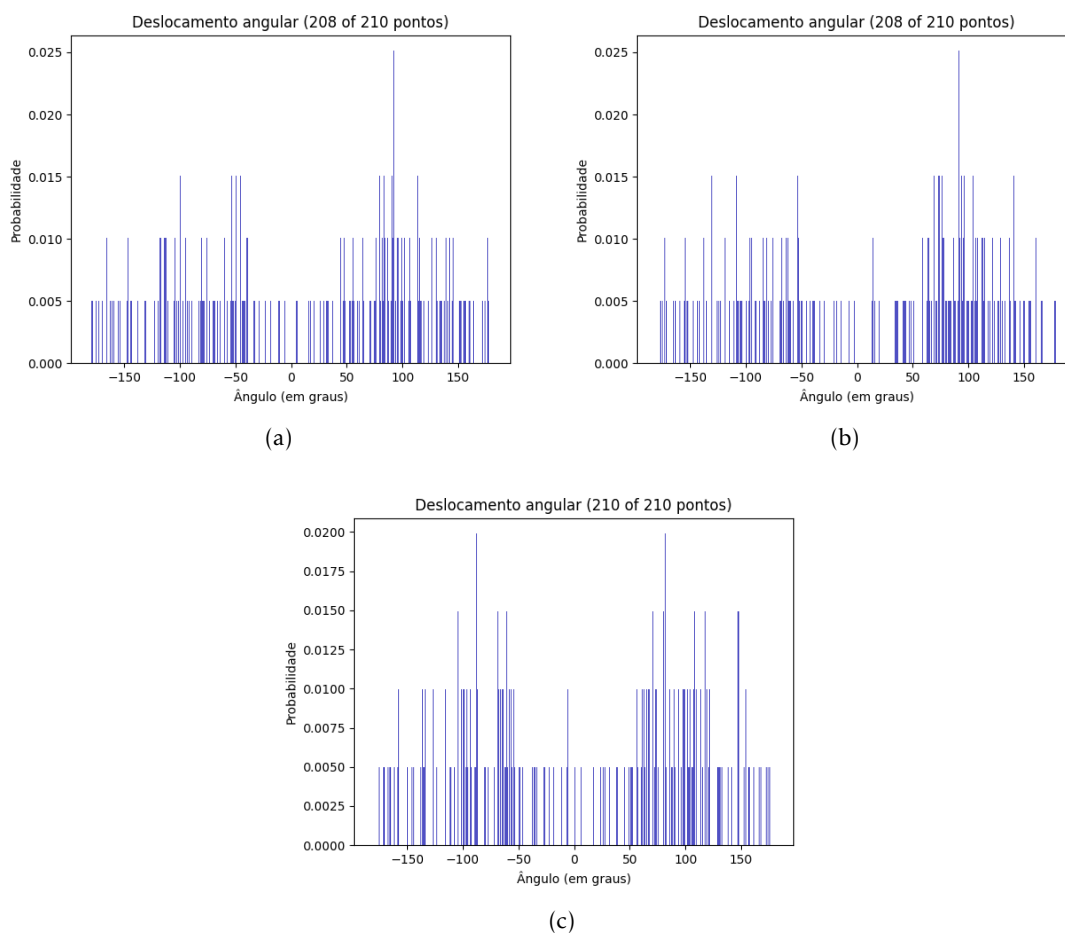


Figura 4.17: Análise na classe completa: em (a), (b) e (c), para os deslocamentos angulares efetuados do mindinho direito, do pulso direito e do cotovelo direito, respetivamente, na classe em que se está a mover o braço.

O que é pretendido ser transmitido aqui é a maneira como a análise pode ser feita ao movimento, e foi selecionado o indicador direito por se saber que é uma articulação de extrema influência no *dataset* que estamos a preparar. No entanto o trabalho foi realizado de forma a tirarmos sempre os histogramas para todas as articulações do algoritmo. Na figura 4.18, temos os histogramas para os vetores deslocamento do cotovelo esquerdo, em que em qualquer um dos cenários que estejamos a observar, são bastante idênticos com todos os deslocamentos muito próximos de zero. Devido a ser levantado o braço direito, é esse lado do corpo que irá estar com maior mobilidade. Conseguimos ainda assim verificar um ligeiro arrastar do gráficos para valores maiores no cenário em que se levanta o braço, o que se justifica pois todo o corpo acaba por se mexer mais, mesmo que esse aumento seja quase imperceptível.

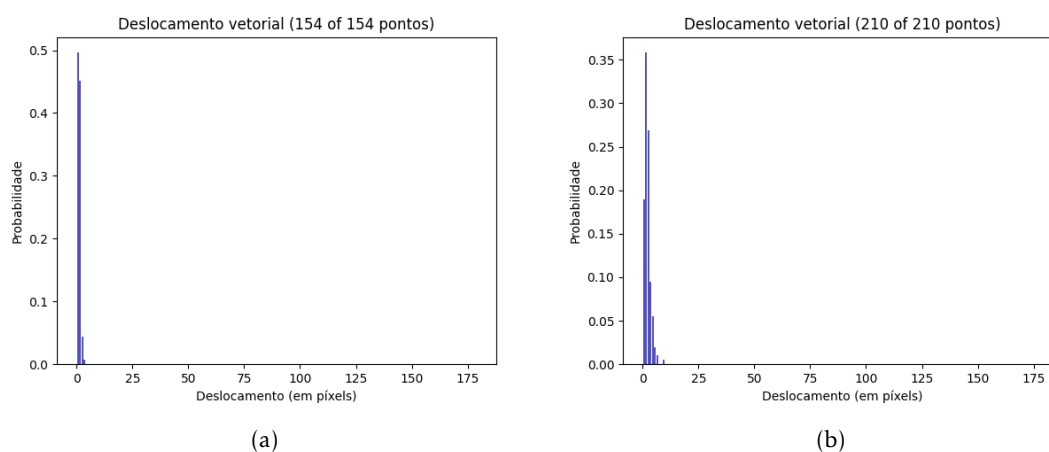


Figura 4.18: Histogramas dos vetores deslocamento nos dois cenários, para o cotovelo esquerdo.

## 4.6 Métricas e classificador final

Nas secções anteriores, foi feita uma análise ao comportamento de cada articulação no *dataset* em questão, maioritariamente ao indicador direito do indivíduo. O modelo foi preparado, e está agora pronto a ser construído um determinado classificador que trabalhe com essas valores e probabilidades retiradas em cada articulação, e que seja capaz de classificar movimentos futuros.

Naturalmente que este classificador foi treinado com base na análise sob as classes no seu todo. Começou-se, num primeiro passo, por calcular a média de cada histograma retirado para cada articulação, tal como foi introduzido na subsecção 3.4.2 do presente documento. Os valores obtidos estão presentes nas tabelas 4.5 e 4.6.

É visível pelos resultados que a classe de mexer o braço volta a confirmar valores superiores, no geral das articulações, para a média dos histogramas dos vetores deslocamento. Estes valores vêm confirmar as análises que podem ser feitas na análise a cada histograma

Tabela 4.5: Média dos histogramas de cada articulação, na classe de estar sentado.

Classe em que a pessoa está sentada					
Nariz	0.78148148	Ombro esquerdo	0.72222222	Polegar direito	1.27777777
Interior olho esquerdo	0.66296296	Ombro direito	0.78888888	Anca esquerda	1.12962962
Olho esquerdo	0.64074074	Cotovelo esquerdo	1.06296296	Anca direita	1.04814814
Exterior olho esquerdo	0.66296296	Cotovelo direito	1.34444444	Joelho esquerdo	5.00370370
Interior olho direito	0.67037037	Pulso esquerdo	0.72222222	Joelho direito	4.29259259
Olho direito	0.65555555	Pulso direito	1.20370370	Tornozelo esquerdo	NaN
Exterior olho direito	0.66296296	Míndinho esquerdo	0.79629629	Tornozelo direito	NaN
Orelha esquerda	0.61111111	Míndinho direito	1.28518518	Calcanhar esquerdo	NaN
Orelha direita	0.58888888	Indicador esquerdo	0.79629629	Calcanhar direito	NaN
Lado esquerdo da boca	0.69259259	Indicador direito	1.33703703	Indicador pé esquerdo	NaN
Lado direito da boca	0.68518518	Polegar esquerdo	0.74444444	Indicador pé direito	NaN

Tabela 4.6: Média dos histogramas de cada articulação, na classe de mexer o braço.

Classe em que a pessoa está a movimentar o braço					
Nariz	1.43532338	Ombro esquerdo	1.55472636	Polegar direito	17.67948717
Interior olho esquerdo	1.25621890	Ombro direito	2.18159203	Anca esquerda	1.95273631
Olho esquerdo	1.23134328	Cotovelo esquerdo	2.10199004	Anca direita	1.96766169
Exterior olho esquerdo	1.23631840	Cotovelo direito	10.73880597	Joelho esquerdo	5.36458333
Interior olho direito	1.29104477	Pulso esquerdo	1.5	Joelho direito	5.99746192
Olho direito	1.28606965	Pulso direito	17.37437185	Tornozelo esquerdo	NaN
Exterior olho direito	1.34079601	Míndinho esquerdo	1.67910447	Tornozelo direito	NaN
Orelha esquerda	1.18159203	Míndinho direito	20.01758793	Calcanhar esquerdo	NaN
Orelha direita	1.19154228	Indicador esquerdo	1.89303482	Calcanhar direito	NaN
Lado esquerdo da boca	1.26616915	Indicador direito	19.19543147	Indicador pé esquerdo	NaN
Lado direito da boca	1.32089552	Polegar esquerdo	1.78358208	Indicador pé direito	NaN

de cada articulação nas duas classes que englobam todo o conjunto de imagem, onde se destaca que na classe do movimento do braço direito, as articulações presentes no membro superior direito da pessoa irão apresentar valores maiores.

Também já mencionado na secção 3.4.2, o próximo passo foi calcular a distância para cada articulação através do desvio à média, que representa a diferença entre o valor obtido e os valores do histograma da classe que estão pesados pelas probabilidades. Mais uma vez, a fórmula usada está na equação 4.2, em que  $x$  é o valor do deslocamento considerado, e  $\mu_j$  o valor da média calculada no passo anterior.

$$d_j = \sqrt{(x_{amostra_j} - \mu_j)^2}. \quad (4.2)$$

Todos estes resultados foram sendo registados para uso no passo seguinte, através da criação e leitura de ficheiros de textos em *Python*. Em cada articulação, para cada deslocamento (que ocorre entre duas *frames* consecutivas em que os pontos em consideração têm visibilidade dada pelo algoritmo superior a 0.95), foi registado o valor, para que no fim fosse calculada a distância total do *dataset* com a soma de todas as distâncias anteriores, contabilizando-se todas as articulações cujo número de pontos com boa visibilidade em todo o conjunto fosse superior a um *threshold* de 20.

De seguida, repetiu-se o passo de calcular as distâncias das articulações dado o desvio à média, mas agora pegando na média da classe oposta (podemos denominar como tal, uma vez que estamos a testar apenas 2 classes. Caso houvesse três classes por exemplo,

iriam ser calculadas 3 distâncias para cada articulação em cada deslocamento). A ideia aqui foi depois poder classificar com a respetiva classe cada deslocamento no final, uma vez que ao termos duas distâncias calculadas (pelo desvio a cada média de classe) para cada deslocamento no conjunto de imagem, a etiquetagem irá ser feita pela menor distância obtida, isto é, o menor desvio à média irá indicar que se trata da classe em questão.

Foi feita a classificação para as duas classes em estudo, e os resultados estão visíveis na figura 4.19 para os deslocamentos que sabemos que se inserem na classe de estar sentado, e na figura 4.20 para os deslocamentos inseridos na classe do movimento do braço. Neste algoritmo de classificação supervisionada, estamos a aplicar o classificador sob as imagens que serviram para o treinar, e por isso sabíamos à partida a classe que teria de ser obtida. Estamos a verificar a eficácia na classificação, que servirá como uma medida de confiança na classificação de movimentos futuros.

Para melhor visualização do desempenho do nosso algoritmo de classificação, na figura 4.21 temos a matriz de confusão que contém os resultados acima obtidos.

É visível que a classificação de deslocamentos quando o indivíduo está sentado é muito mais eficaz que a classificação nas imagens em que o mesmo está a mover o braço. Isto porque na classe em que faz o movimento do braço, a mesma classe acaba por ter muitas imagens em que a pessoa está estática, com o braço levantado. Dessa forma, essas imagens acabam por ser classificadas pelo algoritmo como "Classe sentado", o que não deixa de ser incorreto pelo que foi dito, sendo que a classificação acaba por ser na classe de mover o braço nas alturas em que o movimento do braço está mesmo a ser realizado.

Pela matriz de confusão confirmamos o parágrafo anterior. Quando a classe prevista na classificação seria "MovimentoBraço", onde foram classificados 201 deslocamentos dentro da classe, 48 foram classificados corretamente. Contudo, pelo que também já foi mencionado, 153 acabaram por ser classificados incorretamente com a classe oposta. Por outro lado, quando a classe prevista seria "Sentado", foram classificados 144 movimentos dentro da mesma, em que apenas 2 obtiveram uma classificação incorreta, sendo que os restantes 142 foram classificados corretamente com a classe.

Assim sendo, tivemos um total de 345 movimentos classificados. Uma vez que o número dos que foram classificados de forma correta foi de 190 deslocamentos, este classificador obteve uma eficácia de 55% na classificação, o que se justifica pelo número de classificações incorretas, como já mencionado e justificado, na classe do movimento do braço. Como solução à percentagem apresentada, numa próxima experiência e usando o mesmo *dataset* para estudo, a sub-classe de mover o braço deve ser restringida somente às frames em que de facto há esse movimento (para cima ou para baixo), ocorrendo dessa forma a presença de 4 sub-classes principais: quando o braço está em baixo e o indivíduo está somente sentado, quando é feito o movimento do braço de forma ascendente, quando o braço permanece levantado e, por fim, quando o braço pratica o movimento descendente.





Valor real	MovimentoBraço	Sentado
Valor previsto		
MovimentoBraço	48	153
Sentado	2	142

Figura 4.21: Matriz de confusão da classificação do movimento em cada deslocamento.

Neste capítulo, é feita uma introdução aos sinais por radiofrequência, estudando de igual forma dois *datasets* tal como feito na análise da imagem, nos capítulos anteriores, comparando ambas os contextos através das métricas aplicadas sob os sinais em estudo. De igual forma, em cada caso são obtidos os sinais da antena de referência e da antena sensora, expondo as diferenças entre as duas.

A ideia dos dois cenários em estudo passa por fazer uma caracterização de cada um deles, para que futuramente a mesma já seja tida em conta na análise ao movimento que ocorreu na transição entre cada uma das classes.

## 5.1 Recolha de dados

O objetivo inicial partia da criação de um conjunto de amostras complexas, numa dada banda, quando um movimento humano ocorre na área circundante das antenas. Contudo, nos resultados a seguir expostos não é feita a análise ao próprio movimento entre dois cenários, como já foi dito anteriormente, mas a caracterização de cada um.

Para recolha das amostras, o *setup* foi montado numa das salas do departamento de Engenharia Eletrotécnica da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. Uma determinada pessoa foi colocada numa dada área, para recolha dos sinais para análise. Essas mesmas ações por parte do indivíduo, tiveram um máximo de duração de 20 segundos. A caracterização da ação em estudo e respetivo *dataset* de imagem é descrito na secção 5.2.

### 5.1.1 Rádio definido por *software*

O sistema usado traduziu-se num rádio definido por *software*, incluindo um conjunto de duas antenas omnidirecionais. Cada uma das antenas foi conectada a uma placa *bladeRF micro A4*, tal como ilustrado na figura 5.1.

Na figura, são visíveis as duas antenas de referência e de *sensing* identificadas pelos números 1 e 2, respetivamente. A antena 1 foi colocada numa posição superior à antena

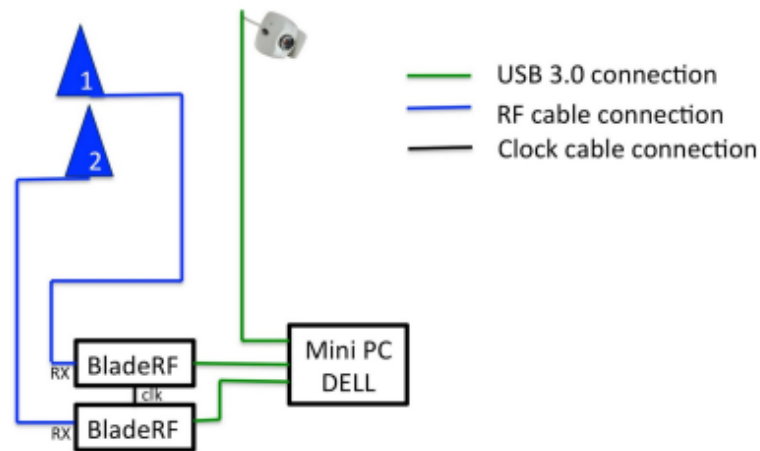


Figura 5.1: Câmera e *setup* do rádio definido por *software*.

2, a 2.20 metros de altitude, sendo que a antena de detecção foi colocada mais abaixo, a uma altura de 1.20 metros.

As duas placas foram conectadas a um computador, que antes de se começar o trabalho prático as configurou com a mesma frequência de operação e largura de banda, introduzindo valores relacionados com a banda de tecnologia de rádio a ser considerada no estudo. A câmera usada para gravar o indivíduo também foi conectada ao computador. Ambos os processos relacionados com a amostragem RF e com a gravação do vídeo foram realizados em paralelo.

### 5.1.2 Bandas de tecnologia

De entre as tecnologias rádio temos: a televisão digital terrestre (DTT), a modulação em amplitude (AM), a modulação em frequência (FM) e a norma IEEE 802.11. Para cada uma das tecnologias a frequência de operação e a largura de banda são escolhidas com base na distância mais próxima do transmissor da tecnologia de rádio em questão, em relação ao local do trabalho prático.

O *dataset* usado e exposto nos resultados descritos nas secções seguintes foi trabalhado com a recolha de sinais da tecnologia FM. Dentro da infraestrutura FM, na figura 5.2 temos representada a potência observada no local prático para as múltiplas estações de rádio. Por observação da figura, as amostras complexas foram adquiridas com uma frequência central de 102 Mhz e com largura de banda de 10 Mhz.

## 5.2 Caracterização dos *datasets*

O estudo foi realizado com recurso a transmissores FM. Como já introduzido, foram usadas duas antenas, uma de referência e outra de *sensing*, para recolha dos dados em cada cenário.

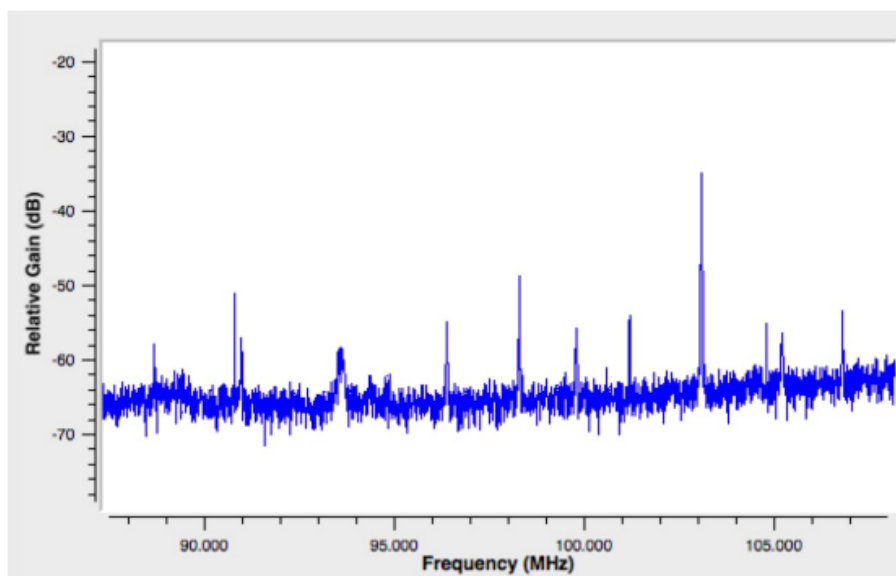


Figura 5.2: Potência de recepção de múltiplas estações de rádio FM.

Os sinais modulados na frequência irão ser usados para o estudo das duas classes em análise: quando temos um indivíduo em pé, e quando o mesmo está sentado, tal como mostrado nas capturas feitas ao vídeo em estudo, na figura 5.3.

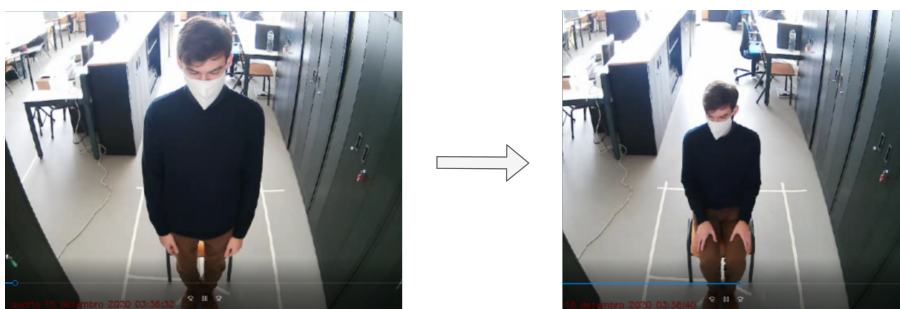


Figura 5.3: Contextos em estudo.

Uma vez que a frequência de amostragem de sinal RF é de 20 Mhz, sabemos que o tempo de amostragem é de 50 nanossegundos pela fórmula na equação 5.1:

$$T_s = \frac{1}{F_s}, \quad (5.1)$$

em que  $T_s$  é o tempo de amostragem e  $F_s$  a respetiva taxa de amostragem.

As amostras foram retiradas com recurso a um vídeo de 21 segundos, em que temos uma pessoa em pé que se senta de seguida, e guardadas como ficheiros binários na diretoria. Uma vez que temos duas antenas, foram guardados dois ficheiros, um para cada uma das antenas. Com recurso à linguagem de programação *python*, utilizada em todo o estudo até aqui, foi retirado que cada um dos ficheiros é composto por 300,000,000 amostras.

Uma vez que a frequência de amostragem usada é de 20 MHz o que equivale a um tempo de amostragem de 50 nanossegundos, esse mesmo tempo multiplicado pelas 300 000 000 amostras dá um total de 15 segundos. Uma vez que o vídeo respetivo tem 21 segundos, é sinal que o mesmo e a recolha de amostras RF não possuem a mesma duração. Isto deve-se ao facto de inicialmente quando é iniciado o processo de recolha no equipamento, o mesmo precisa de um determinado período para a sua própria configuração. Com isto, considerou-se os primeiros 5 segundos assim como os últimos 5 no tempo de recolha das amostras de forma a capturar a mudança de contexto relativamente à pose humana.

### 5.3 Resultados experimentais

Introduzidas as classes para análise na secção anterior, se formos abrir todas as amostras para estudo dentro de cada uma delas, o número continua a ser enorme e não é uma tarefa fácil pegar em todas elas para obtenção das métricas. Foi decidido dentro de cada intervalo de classe fazer o estudo sob pequenas partes dentro de cada um, pegando em cada 200000 amostras e fazer a caracterização sobre elas. A captura desse número de amostras correndo o vídeo em paralelo, corresponde a um intervalo de tempo de 0.01 segundos.

Posto isto, e iniciando do estudo do *dataset* com as classes mencionadas na secção anterior, foram definidos 20 intervalos dentro do tempo de recolha das amostras para análise, cada um contendo 200000 amostras.

Os sub-intervalos para análise estão definidos na tabela 5.1 assim como uma etiqueta para cada um deles (um número de 1 a 20) que irá ser usada para identificação do intervalo em alguns gráficos mais adiante. Estes intervalos correspondem à recolha de amostras, ou seja, assim que as antenas começam a receber sinal os tempos da tabela começam a ser contabilizados.

Apesar de haver tempo de vídeo em que não ocorre recolha de amostras, há garantias que nos intervalos de tempo selecionados estão presentes as duas classes do *dataset*. Apenas existem 6 segundos em que existe captura de imagem mas não de amostras, e no vídeo a transição de contexto ocorre aos 8 segundos. Ora por visualização dos valores de intervalo selecionados, e pelos valores de duração de cada classe no vídeo, o movimento estará presente algures entre os valores de intervalo da tabela .

#### 5.3.1 Métricas aplicadas

Para cada um dos intervalos de tempo da tabela 5.1, foram aplicadas métricas sobre a amplitude e fase do sinal recebido em cada antena, mais concretamente para os valores da média e variância das mesmas nos espaços de tempo deinidos. Por exemplo, para o intervalo 1 correspondente ao intervalo de tempo entre os 0.50 e os 0.51 segundos, foram retiradas a média e variância da amplitude do sinal recebido pela antena de referência

Tabela 5.1: Sub-intervalos de tempo em estudo, e identificação com um número de cada um.

Número do intervalo	Intervalo de tempo
1	[0.50; 0.51]s
2	[1.00; 1.01]s
3	[1.50; 1.51]s
4	[2.00; 2.01]s
5	[2.50; 2.51]s
6	[3.00; 3.01]s
7	[3.50; 3.51]s
8	[4.00; 4.01]s
9	[4.50; 4.51]s
10	[5.00; 5.01]s
11	[10.50; 10.51]s
12	[11.00; 11.01]s
13	[11.50; 11.51]s
14	[12.00; 12.01]s
15	[12.50; 12.51]s
16	[13.00; 13.01]s
17	[13.50; 13.51]s
18	[14.00; 14.01]s
19	[14.50; 14.51]s
20	[14.99; 15.00]s

entre esses instantes, o mesmo para a fase do sinal, e isto tudo repetido para a antena de *sensing*.

É conhecido que no conjunto de imagem o indivíduo está em pé e senta-se posteriormente, pelo que as zonas iniciais dos gráficos abaixo irão ser comparadas com as zonas finais dos mesmos, uma vez que há certeza que se tratará de contextos diferentes.

Nas figuras 5.4, 5.5, 5.6 e 5.7 temos a distribuição desses pontos, que foram interligados, ao longo dos intervalos, com a identificação da tabela, para cada uma das métricas e para cada uma das antenas.

Pela figura 5.4 é observável que os valores de amplitude para o sinal na antena de *sensing* são maiores que na antena de referência.

Na figura 5.5, analisando a variância nos valores de amplitude para as duas antenas, em cada intervalo, apesar das duas antenas apresentarem valores de variância muito idênticos numa fase inicial da recolha de amostras, a variância na antena de *sensing* acaba por aumentar assim que avançamos para espaços temporais mais avançados. Enquanto que para a antena de referência pode-se dizer que no contexto em pé os valores são ligeiramente maiores que no contexto sentado, na antena de *sensing* já é mais perceptível que os valores de variância na amplitude são superiores na classe em que a pessoa está sentada.

Nas figuras 5.6 e 5.7 temos, por fim, as métricas que foram aplicadas aos valores de

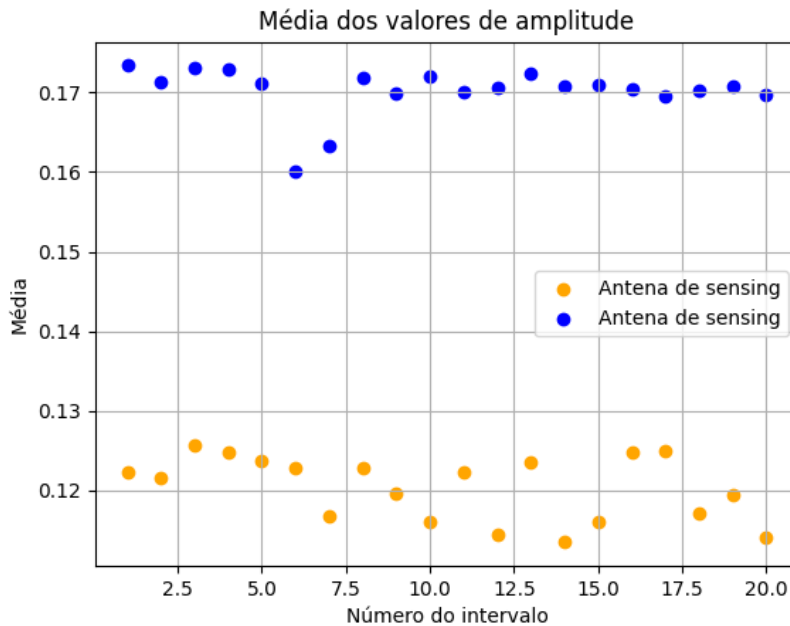


Figura 5.4: Média das amplitudes do sinal, nas duas antenas.

amplitude do sinal, mas agora para os valores da fase.

Pela figura 5.6 os valores entre contextos mostram-se muito parecidos nas duas antenas. Na antena de referência essa constante consegue ser mais perceptível com os valores praticamente constantes ao longo dos intervalos. Na antena de *sensing* a classe em que o indivíduo está sentado aparenta apresentar valores de fase ligeiramente maiores.

Na figura 5.7 vemos que a antena de *sensing* apresenta valores de variância na fase superiores à antena de referência. Ainda assim, a variância na fase não servirá para diferenciar os dois contextos em estudo, uma vez que não existe grande distinção dos valores ao longo dos intervalos, em cada antena.

Por análise geral aos gráficos obtidos, é possível concluir a alta probabilidade da transição de contexto ter ocorrido junto ao intervalo 7, que pela tabela 5.1 corresponde a cerca de 3.5 segundos depois de se ter iniciado a recolha de amostras, já que em todos os gráficos existem variações significativas junto ao mesmo.

### 5.3.2 Correlação do valor das amplitudes entre as duas antenas

Sendo  $I_1 = i_1, i_2, \dots, i_{200000}$  e  $J_1 = j_1, j_2, \dots, j_{200000}$ , as duas séries temporais de 200000 amostras da amplitude da antena de referência e da antena de *sensing*, capturadas em cada sub-intervalo estudado. Cada par das mesmas foi submetido a uma função de correlação cruzada que é representativa de uma distribuição da subtração das duas magnitudes.

Os gráficos das correlações são bastante idênticos para todos os intervalos de tempo em estudo, tal como é mostrado por alguns deles na figura 5.8.

As rampas quase lineares mostram a uniformidade dos sinais. É visível que inclusive

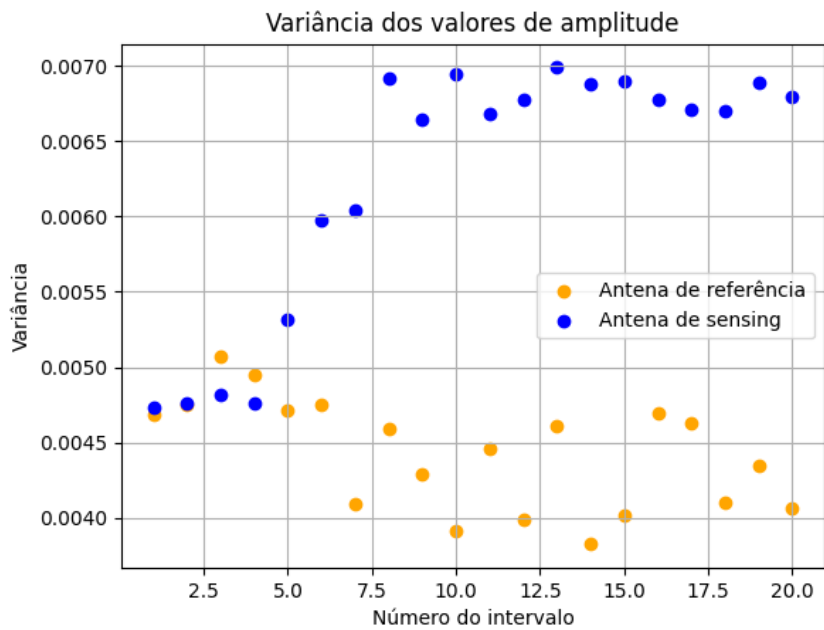


Figura 5.5: Variância dos valores da amplitude do sinal, para as duas antenas.

no ponto máximo de correlação a mesma é muito baixa, em todos os casos. É normal a forma triangular apresentada, com valores máximos para a probabilidade da diferença a meio dos vetores de amostras, pois estamos numa situação de correlação cruzada com janela deslizante, e é quando os dois vetores são correlacionados integralmente que podem existir mais diferenças entre os mesmos.

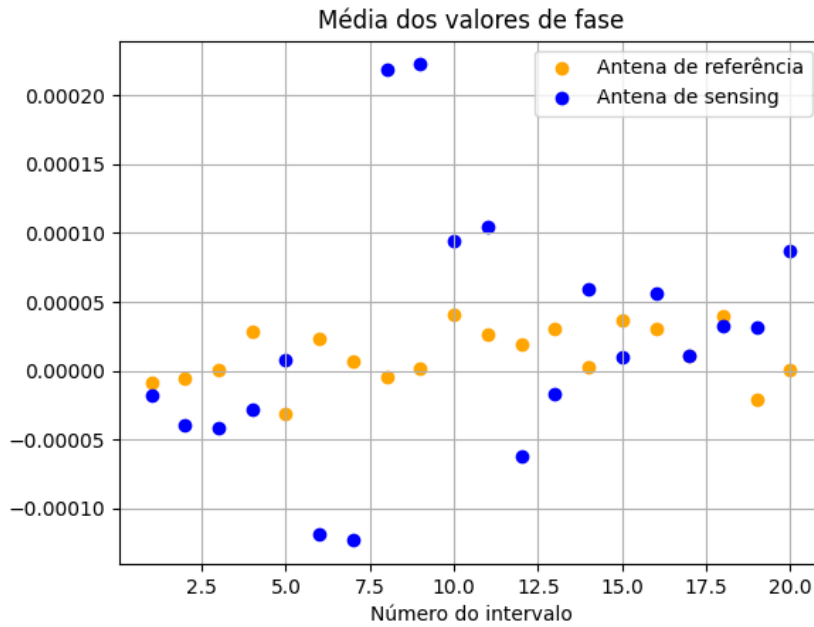


Figura 5.6: Média dos valores de fase dos sinais das duas antenas.

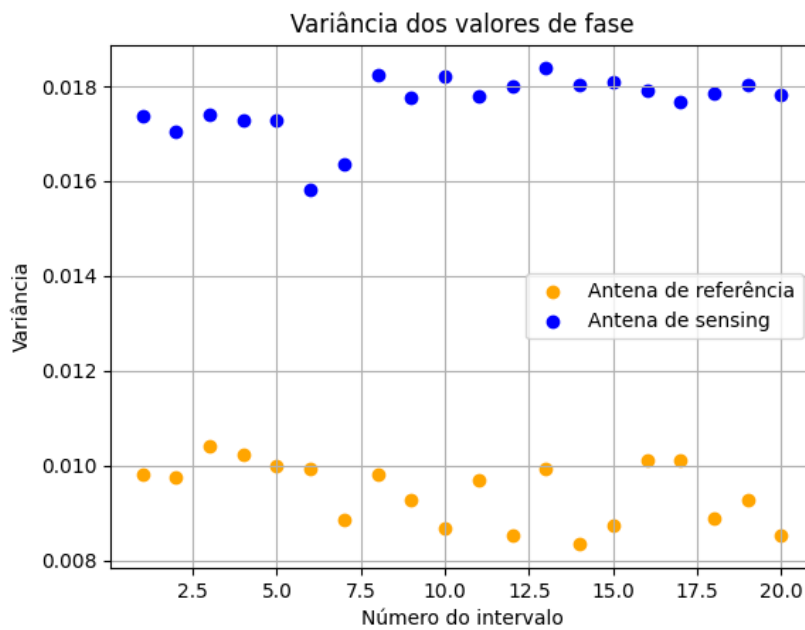


Figura 5.7: Variância dos valores de fase, para as duas antenas.

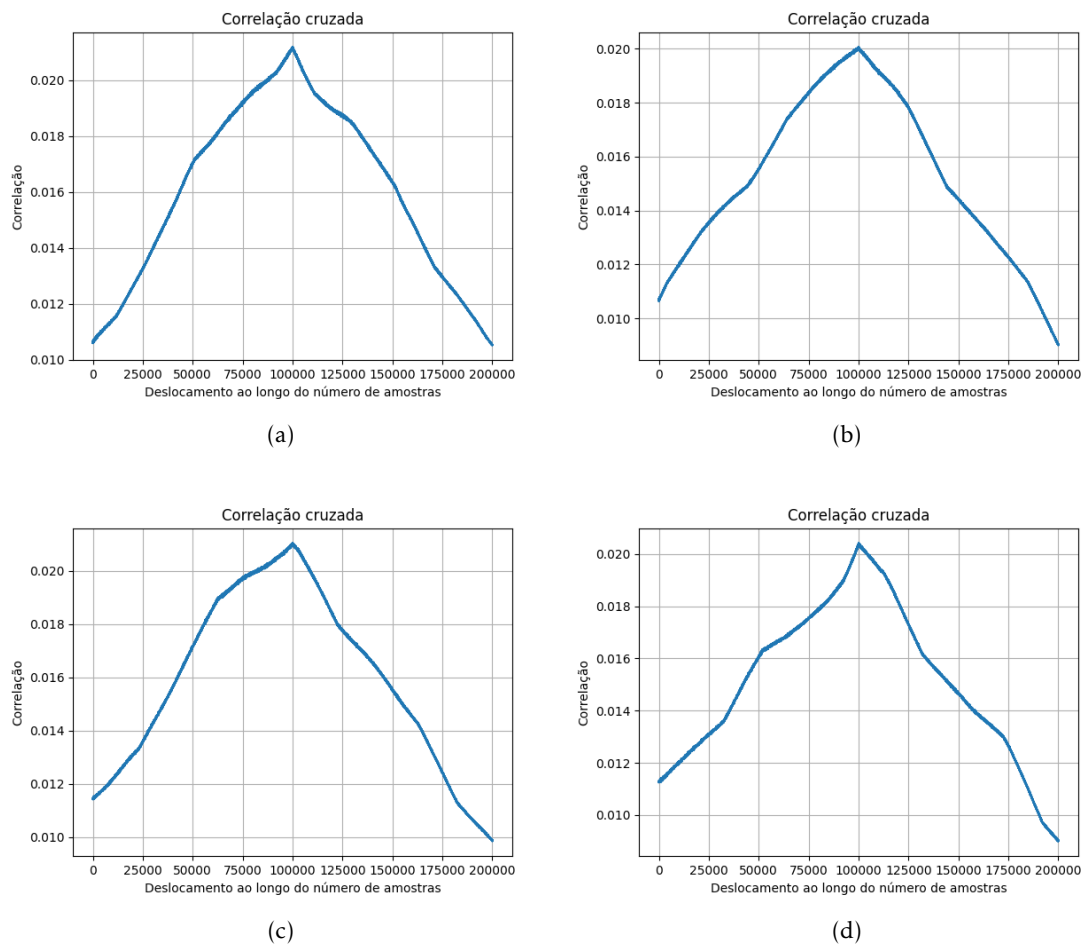


Figura 5.8: Correlação cruzada entre as amplitudes do sinal da antena de referência com a antena de *sensing*, para quatro intervalos de tempo em estudo.

## CONCLUSÕES

Saber prever o movimento humano é um dos tópicos mais interessantes e explorados na interação humano-computador. Nesta dissertação foi feito um estudo intensivo ao domínio de imagem, aplicando um modelo de estimação da pose para um determinado alvo identificado, o que levou à criação de um modelo, obtido por aprendizagem, treinado e capaz de ser usado sobre conjuntos de imagem futuros. A detecção RF foi introduzida com a aplicação de um sistema de recolha de sinais de rádio, dentro da tecnologia FM, afetados por determinada ação no caminho até às antenas.

Desta forma, no capítulo 3 foi realizada uma descrição sobre o domínio de imagem e sobre o algoritmo aplicado para a identificação da pose do indivíduo. Ficou definido um modelo estatístico bem como as características fundamentais que levam a uma análise correta ao movimento realizado entre um determinado conjunto de imagem. Adicionalmente, foram descritas as métricas a usar nos resultados obtidos e que levaram à construção do classificador final adotado. No capítulo 4 foram expostos os resultados obtidos sobre o domínio de imagem. As métricas utilizadas assim como o modelo de classificação foram comprovados ao serem aplicados aos dados de treino. Consoante o contexto em análise, existiram articulações com maior influência para o estudo que outras, indicativas de igual forma de qual seria o movimento esperado que estivesse a ocorrer.

Por outro lado, o capítulo 5 introduziu o domínio de RF, comprovando-se a influência de movimentos ocorridos em determinado cenário sobre as ondas RF que atravessam o mesmo. Foram usadas duas antenas dispostas em localizações e alturas diferentes, e foi possível comparar as duas assim como detetar as diferenças no sinal obtido por cada uma, perante determinado contexto.

## BIBLIOGRAFIA

- [1] J. M. Lourenço. *The NOVAthesis L<sup>A</sup>T<sub>E</sub>X Template User's Manual*. NOVA University Lisbon. 2021. URL: <https://github.com/joaomlourenco/novathesis/raw/master/template.pdf> (ver p. ii).
- [2] J. .-. Zhao e M. .-. Liu. "A Color HSV Image Edge Detection Method Based on Gradient Extreme Value". Em: *2008 Second International Symposium on Intelligent Information Technology Application*. Vol. 3. 2008, pp. 381–384. DOI: [10.1109/IITA.2008.9](https://doi.org/10.1109/IITA.2008.9) (ver p. 4).
- [3] R. Mancuso, S. Smorfa e M. Olivieri. "A novel high-quality YUV-based image coding technique for efficient image storage in portable electronic appliances". Em: *IEEE Transactions on Consumer Electronics* 54.2 (2008), pp. 695–702. DOI: [10.1109/TCE.2008.4560149](https://doi.org/10.1109/TCE.2008.4560149) (ver p. 4).
- [4] T. Moeslund, A. Hilton e V. Krüger. "A survey of advances in vision-based human motion capture and analysis". Em: *Computer Vision and Image Understanding* 104 (nov. de 2006), pp. 90–126. DOI: [10.1016/j.cviu.2006.08.002](https://doi.org/10.1016/j.cviu.2006.08.002) (ver pp. 4, 9, 10).
- [5] G. Stockman e L. G. Shapiro. *Computer Vision*. 1st. USA: Prentice Hall PTR, 2001. ISBN: 0130307963 (ver p. 6).
- [6] G. R. Bradski e J. W. Davis. "Motion Segmentation and Pose Recognition with Motion History Gradients". Em: *Mach. Vision Appl.* 13.3 (jul. de 2002), pp. 174–184. ISSN: 0932-8092. DOI: [10.1007/s001380100064](https://doi.org/10.1007/s001380100064). URL: <https://doi.org/10.1007/s001380100064> (ver pp. 6, 7).
- [7] P. Viola, M. J. Jones e D. Snow. "Detecting pedestrians using patterns of motion and appearance". Em: *International Journal of Computer Vision* 63.2 (2005), pp. 153–161 (ver p. 7).
- [8] Y. Kameda e M. Minoh. "A human motion estimation method using 3-successive video frames". Em: *International conference on virtual systems and multimedia*. 1996, pp. 135–140 (ver p. 8).

- [9] C. R. Wren et al. "Pfinder: real-time tracking of the human body". Em: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.7 (1997), pp. 780–785. DOI: [10.1109/34.598236](https://doi.org/10.1109/34.598236) (ver p. 9).
- [10] R. Rosales e S. Sclaroff. "Learning and synthesizing human body motion and posture". Em: *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. 2000, pp. 506–511. DOI: [10.1109/AFGR.2000.840681](https://doi.org/10.1109/AFGR.2000.840681) (ver p. 9).
- [11] N. Howe, M. Leventon e W. Freeman. "Bayesian Reconstruction of 3D Human Motion from Single-Camera Video". Em: (abr. de 2000) (ver p. 9).
- [12] M. Brand. "Shadow puppetry". Em: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, 1237–1244 vol.2. DOI: [10.1109/ICCV.1999.790422](https://doi.org/10.1109/ICCV.1999.790422) (ver p. 9).
- [13] I. Haritaoglu, D. Harwood e L. S. Davis. "Ghost: a human body part labeling system using silhouettes". Em: *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*. Vol. 1. 1998, 77–82 vol.1. DOI: [10.1109/ICPR.1998.711084](https://doi.org/10.1109/ICPR.1998.711084) (ver p. 9).
- [14] C. R. Wren e A. P. Pentland. "Dynamic models of human motion". Em: *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*. 1998, pp. 22–27. DOI: [10.1109/AFGR.1998.670920](https://doi.org/10.1109/AFGR.1998.670920) (ver p. 10).
- [15] Z. Cao et al. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields". Em: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1 (2021), pp. 172–186. DOI: [10.1109/TPAMI.2019.2929257](https://doi.org/10.1109/TPAMI.2019.2929257) (ver p. 10).
- [16] P. F. Felzenszwalb e D. P. Huttenlocher. "Pictorial Structures for Object Recognition". Em: *Int. J. Comput. Vision* 61.1 (jan. de 2005), pp. 55–79. ISSN: 0920-5691. DOI: [10.1023/B:VISI.0000042934.15159.49](https://doi.org/10.1023/B:VISI.0000042934.15159.49). URL: <https://doi.org/10.1023/B:VISI.0000042934.15159.49> (ver p. 10).
- [17] R. Gu et al. "Multi-Person Hierarchical 3D Pose Estimation in Natural Videos". Em: *IEEE Transactions on Circuits and Systems for Video Technology* 30.11 (2020), pp. 4245–4257. DOI: [10.1109/TCSVT.2019.2953678](https://doi.org/10.1109/TCSVT.2019.2953678) (ver p. 10).
- [18] T. W. Mathumo, T. G. Swart e R. W. Focke. "Implementation of a GNU radio and python FMCW radar toolkit". Em: *2017 IEEE AFRICON*. 2017, pp. 585–590. DOI: [10.1109/AFRCON.2017.8095547](https://doi.org/10.1109/AFRCON.2017.8095547) (ver p. 11).
- [19] Sundaresan S et al. "Real time implementation of FMCW radar for target detection using GNU radio and USRP". Em: *2015 International Conference on Communications and Signal Processing (ICCSP)*. 2015, pp. 1530–1534. DOI: [10.1109/ICCSP.2015.7322772](https://doi.org/10.1109/ICCSP.2015.7322772) (ver p. 11).

- [20] C. Will et al. “Human Target Detection, Tracking, and Classification Using 24-GHz FMCW Radar”. Em: *IEEE Sensors Journal* 19.17 (2019), pp. 7283–7299. DOI: [10.1109/JSEN.2019.2914365](https://doi.org/10.1109/JSEN.2019.2914365) (ver p. 11).
- [21] Q. Zhu e Y. Wang. “FMCW radar implemented with GNU Radio Companion”. Em: 2016 (ver p. 12).
- [22] B.-s. Kim et al. “Low-Complexity Joint Range and Doppler FMCW Radar Algorithm Based on Number of Targets”. Em: *Sensors* 20.1 (2020). ISSN: 1424-8220. DOI: [10.3390/s20010051](https://doi.org/10.3390/s20010051). URL: <https://www.mdpi.com/1424-8220/20/1/51> (ver p. 13).
- [23] K. Ramasubramanian. “Using a complex-baseband architecture in FMCW radar systems”. Em: 2017 (ver pp. 13, 15–17).
- [24] G. Wang et al. “Application of Linear-Frequency-Modulated Continuous-Wave (LFMCW) Radars for Tracking of Vital Signs”. Em: *IEEE Transactions on Microwave Theory and Techniques* 62.6 (2014), pp. 1387–1399. DOI: [10.1109/TMTT.2014.2320464](https://doi.org/10.1109/TMTT.2014.2320464) (ver pp. 13, 14, 30).
- [25] E. Hyun, Y.-S. Jin e J.-H. Lee. “A Pedestrian Detection Scheme Using a Coherent Phase Difference Method Based on 2D Range-Doppler FMCW Radar”. Em: *Sensors* 16 (jan. de 2016), p. 124. DOI: [10.3390/s16010124](https://doi.org/10.3390/s16010124) (ver pp. 14, 15).
- [26] L. Piotrowsky e N. Pohl. “Spatially Resolved Fast-Time Vibrometry Using Ultrawideband FMCW Radar Systems”. Em: *IEEE Transactions on Microwave Theory and Techniques* 69.1 (2021), pp. 1082–1095. DOI: [10.1109/TMTT.2020.3038080](https://doi.org/10.1109/TMTT.2020.3038080) (ver p. 15).
- [27] K. Siddiq et al. “On Phase Measurement in FMCW Radar Systems”. Em: *2017 Sensor Signal Processing for Defence Conference (SSPD)*. 2017, pp. 1–4. DOI: [10.1109/SSPD.2017.8233250](https://doi.org/10.1109/SSPD.2017.8233250) (ver p. 16).
- [28] B. Fulcher. “Feature-based time-series analysis”. Em: (set. de 2017) (ver pp. 19, 22).
- [29] B. Fulcher e N. Jones. “Highly Comparative Feature-Based Time-Series Classification”. Em: *IEEE Transactions on Knowledge and Data Engineering* 26 (jan. de 2014). DOI: [10.1109/TKDE.2014.2316504](https://doi.org/10.1109/TKDE.2014.2316504) (ver p. 19).
- [30] X. Wang, K. Smith-Miles e R. Hyndman. “Characteristic-Based Clustering for Time Series Data”. Em: *Data Min. Knowl. Discov.* 13 (set. de 2006), pp. 335–364. DOI: [10.1007/s10618-005-0039-x](https://doi.org/10.1007/s10618-005-0039-x) (ver pp. 19, 22).
- [31] D. J. Weller-Fahy, B. J. Borghetti e A. A. Sodemann. “A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection”. Em: *IEEE Communications Surveys Tutorials* 17.1 (2015), pp. 70–91. DOI: [10.1109/COMST.2014.2336610](https://doi.org/10.1109/COMST.2014.2336610) (ver p. 20).
- [32] F. Mörchen. “Time series feature extraction for data mining using DWT and DFT”. Em: 2003 (ver p. 23).

- [33] P. Gupta e T. Dallas. “Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer”. Em: *IEEE Transactions on Biomedical Engineering* 61.6 (2014), pp. 1780–1786. DOI: [10.1109/TBME.2014.2307069](https://doi.org/10.1109/TBME.2014.2307069) (ver p. 23).
- [34] J. Liu et al. “Wireless Sensing for Human Activity: A Survey”. Em: *IEEE Communications Surveys Tutorials* 22.3 (2020), pp. 1629–1645. DOI: [10.1109/COMST.2019.2934489](https://doi.org/10.1109/COMST.2019.2934489) (ver p. 24).
- [35] Z. Zeng, M. G. Amin e T. Shan. “Automatic Arm Motion Recognition Based on Radar Micro-Doppler Signature Envelopes”. Em: *IEEE Sensors Journal* 20.22 (2020), pp. 13523–13532. DOI: [10.1109/JSEN.2020.3004581](https://doi.org/10.1109/JSEN.2020.3004581) (ver p. 25).
- [36] B. Jokanović e M. Amin. “Suitability of Data Representation Domains in Expressing Human Motion Radar Signals”. Em: *IEEE Geoscience and Remote Sensing Letters* 14.12 (2017), pp. 2370–2374. DOI: [10.1109/LGRS.2017.2765341](https://doi.org/10.1109/LGRS.2017.2765341) (ver pp. 25, 26).
- [37] J. Yan et al. “Phase-Based Human Target 2-D Identification With a Mobile FMCW Radar Platform”. Em: *IEEE Transactions on Microwave Theory and Techniques* 67.12 (2019), pp. 5348–5359. DOI: [10.1109/TMTT.2019.2939523](https://doi.org/10.1109/TMTT.2019.2939523) (ver p. 26).
- [38] P. Bahl e V. N. Padmanabhan. “RADAR: an in-building RF-based user location and tracking system”. Em: *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*. Vol. 2. 2000, 775–784 vol.2. DOI: [10.1109/INFCOM.2000.832252](https://doi.org/10.1109/INFCOM.2000.832252) (ver p. 27).
- [39] M. O. Padar, A. E. Ertan e Ç. ğ. Candan. “Classification of human motion using radar micro-Doppler signatures with hidden Markov models”. Em: *2016 IEEE Radar Conference (RadarConf)*. 2016, pp. 1–6. DOI: [10.1109/RADAR.2016.7485201](https://doi.org/10.1109/RADAR.2016.7485201) (ver pp. 27, 32).
- [40] X. Li et al. “Hierarchical radar data analysis for activity and Personnel recognition”. Em: *Remote Sensing* 12.14 (2020), p. 2237 (ver p. 28).
- [41] F. Adib et al. “3D Tracking via Body Radio Reflections”. Em: *NSDI*. 2014 (ver pp. 27, 32).
- [42] Q. Pu et al. “Whole-home gesture recognition using wireless signals”. Em: *Proceedings of the 19th annual international conference on Mobile computing networking* (2013) (ver pp. 28, 29, 32).
- [43] G. Malysa et al. “Hidden Markov model-based gesture recognition with FMCW radar”. Em: *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. 2016, pp. 1017–1021. DOI: [10.1109/GlobaSIP.2016.7905995](https://doi.org/10.1109/GlobaSIP.2016.7905995) (ver pp. 29, 32).
- [44] J. Lien et al. “Soli: ubiquitous gesture sensing with millimeter wave radar”. Em: *ACM Trans. Graph.* 35 (2016), 142:1–142:19 (ver pp. 29, 32).

- 
- [45] H. R. Lee, J. Park e Y.-J. Suh. “Improving Classification Accuracy of Hand Gesture Recognition Based on 60 GHz FMCW Radar with Deep Learning Domain Adaptation”. Em: *Electronics* 9 (2020), p. 2140 (ver pp. 30, 32).
- [46] J. Yan et al. “The Development of Vital-SAR-Imaging with an FMCW Radar System”. Em: *2019 IEEE MTT-S International Microwave Biomedical Conference (IM-BioC)*. Vol. 1. 2019, pp. 1–4. DOI: [10.1109/IMBIOC.2019.8777881](https://doi.org/10.1109/IMBIOC.2019.8777881) (ver pp. 30, 32).
- [47] F. Adib et al. “Smart Homes that Monitor Breathing and Heart Rate”. Em: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015) (ver pp. 30–32).
- [48] L. Anitori, A. de Jong e F. Nennie. “FMCW radar for life-sign detection”. Em: *2009 IEEE Radar Conference*. 2009, pp. 1–6. DOI: [10.1109/RADAR.2009.4976934](https://doi.org/10.1109/RADAR.2009.4976934) (ver pp. 31, 32).
- [49] A. S. B. Pauzi et al. “Movement Estimation Using Mediapipe BlazePose”. Em: *International Visual Informatics Conference*. Springer. 2021, pp. 562–571 (ver pp. 34, 43).

