



FACULDADE DE CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE COIMBRA

MESTRADO INTEGRADO EM ENGENHARIA  
ELECTROTÉCNICA E DE COMPUTADORES

---

# **Interfaces Inovadoras e de Baixo Custo para Aplicações de Software**

**Mecanismos de Interacção  
Baseados no Reconhecimento Gestual**

---

Gustavo José Videira Quintela

**Membros do Júri:**

Professor Doutor Nuno Miguel Mendonça da Silva Gonçalves

Professor Doutor Paulo Jorge Carvalho Menezes

Professor Doutor Marco Alexandre Cravo Gomes

Coimbra, Setembro de 2012



# Agradecimentos

*Em primeiro lugar dedico um agradecimento especial a toda a minha família, aos meus pais e irmãos, pelo forma como sempre me apoiaram e guiaram ao longo da minha vida.*

*Quero também agradecer toda a orientação, disponibilidade e incentivo transmitidos pelo Professor Doutor Paulo Menezes, sem o qual não seria possível a realização desta dissertação de mestrado. Agradeço também ao Eng. Luís Almeida por toda a contribuição e apoio técnico.*

*Aos meus amigos David, Miguel, Adriano, Bruno, Filipe, por todo o apoio e bons momentos passados durante toda esta etapa.*

*Deixo um enorme obrigado à minha namorada Eliana pelo apoio e incentivo, pela amizade e amor.*

*Não podia deixar de agradecer a todos os colegas e amigos que me apoiaram desde a minha entrada na Universidade e me permitirão assim relembrar talvez uma das melhores fases da minha vida.*

*A todos o meu profundo agradecimento,*

*Gustavo Quintela*



# Resumo

Cada vez mais, os sistemas de Interação Homem-Máquina são considerados uma mais valia para a comunidade. Esta interação quando feita natural e intuitivamente, torna-se mais simples para o operador executá-la, pois não necessita de qualquer aprendizagem para a fazer. Desta forma, no presente trabalho são desenvolvidos mecanismos para o reconhecimento de gestos simples, como os periódicos e deícticos ou demonstrativos, através de uma interface de programação de aplicações inovadora e de baixo custo, designada por OpenNI.

São explorados e comparados dois métodos de processamento de sinal, utilizados para a detecção dos gestos periódicos, um deles com recurso a *Fast Fourier Transform* e outro com recurso a Auto-Correlação.

É também tratada uma experiência que valida a forma como é feito o reconhecimento dos gestos periódicos, envolvendo a entoação que determinado usuário tem em indicar uma acção, explorando a intensidade da velocidade com que este executa o gesto.

Neste trabalho, é reconhecido um dos gestos mais apetecíveis na interação com o robô. Trata-se do gesto de apontar, que está dentro do conjunto de gestos deícticos ou demonstrativos, tendo sido obtidos resultados de grande precisão. Para além disso, é também explorada uma técnica de redução de ruído nos dados utilizados, tornando o sistema de apontar mais exacto.

Através dos resultados analisados, é possível concluir que os mecanismos utilizados são robustos, mostrando que estes são fonte de uma potencial utilização por parte de diversas aplicações de interação, quer na vertente social quer na vertente industrial.

**Palavras-chave:** Interação homem-máquina, Reconhecimento gestual, Gestos periódicos, Gestos deícticos ou demonstrativos.



# Abstract

Increasingly, the systems for Human-Robot Interaction have been demonstrating that are an asset to the community. This interaction when done naturally and intuitively, it makes it easier for the operator to perform it, and it does not require any learning. Thus, in the present work are developed mechanisms for recognizing simples gestures, such as periodic and deictics or demonstratives, through an application programming interface innovative and low cost, called OpenNI.

Are explored and compared two methods of signal processing used to detect periodic gestures, one using Fast Fourier Transform and another using Autocorrelation.

It is also treated an experiment that validates the way the recognition of periodic gestures is done, involving the intention that a particular user has on state an action, by exploring the intensity of the velocity with which it performs the gesture.

Still, this work is acknowledge one of the most desirable gestures in dealing with the interaction with the robot. It's pointing gesture, which is within the range of deictics or demostrative gestures. Where on these, the results were obtained with great precision. Moreover, it is also explored a technique for reducing noise in data used which makes the system more accurate.

Through analysis of the results we conclude that the mechanisms used are robust, showing that they are a potential source for use by various applications in social interaction and also in the industrial side.

**Keywords:** Human-Robot Interaction, Gesture recognition, Periodic gestures, Deictics or demostrative gestures.





*"If you're not failing every now and again,  
it's a sign you're not doing anything very innovative."*

Woody Allen



# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação, objectivos e contribuição . . . . .	2
1.2	Estado de arte . . . . .	4
1.3	Estrutura da tese . . . . .	6
<b>2</b>	<b>Gestos Periódicos</b>	<b>9</b>
2.1	Detecção de Gestos Periódicos . . . . .	10
2.1.1	Método com recurso à FFT . . . . .	10
2.1.2	Método com recurso à auto-correlação . . . . .	13
2.2	Comparação dos métodos e seus resultados . . . . .	17
2.2.1	Gesto circular . . . . .	17
2.2.2	Gesto horizontal com movimento periódico sobre o eixo de coordenadas $X$ . . . . .	18
2.2.3	Gesto horizontal com movimento periódico sobre o eixo de coordenadas $Z$ . . . . .	19
2.2.4	Movimento aleatório . . . . .	20
<b>3</b>	<b>Análise dos gestos periódicos</b>	<b>21</b>
3.1	Apresentação de um método proposto para o reconhecimento de gestos periódicos . . . . .	23
3.2	Confirmação do método proposto recorrendo a uma experiência. . . . .	31
<b>4</b>	<b>Gestos demonstrativos</b>	<b>35</b>
4.1	Gesto apontar através da intersecção com plano do chão . . . . .	35
4.2	Método para redução de ruído . . . . .	37
4.3	Resultados . . . . .	39
<b>5</b>	<b>Conclusão</b>	<b>43</b>
5.1	Trabalho futuro . . . . .	46
	<b>Referências Bibliograficas</b>	<b>48</b>
<b>A</b>	<b>Resultados Gestos Periódicos</b>	<b>53</b>
A.1	Gestos Periódicos . . . . .	53
A.2	Detecção dos Gestos Periódicos . . . . .	55

<b>B Reconhecimento de Gesto Periódicos</b>	<b>57</b>
B.1 Gestos periódicos a reconhecer . . . . .	57

# Lista de Figuras

1.1	Seguimento do esqueleto humano adquirido através da interface de programação de aplicações OpenNI complementada com a tecnologia do Xbox Kinect. . . . .	3
2.1	Progressão no tempo da coordenada X do um gesto com direcção horizontal e a respectiva análise espectral com a frequência corresponde de 1.641Hz. (a) - Coordenada X de um gesto com direcção horizontal; (b) - Análise espectral. (c) - Coordenada Y de um gesto com direcção vertical; (d) - Análise espectral . . . . .	10
2.2	Diagrama de blocos que representa o método implementado. São descritos os sucessivos passos necessários ao funcionamento do mesmo. . . . .	11
2.3	Passo do método onde é feita a selecção de uma sequência periódica e não periódica, entre a máxima amplitude da $f_s$ (representado por uma linha vermelha) e a gama de frequências 0.45 Hz e 2 Hz. (a) - Sequência de 60 amostras de uma trajectória não periódica sem componente DC; (b) - Espectro de frequência correspondente a ao sinal (a); (c) - Sequência de 60 amostras de uma trajectória periódica sem componente DC; (d) - Espectro de frequência correspondente a ao sinal (c). . . . .	12
2.4	Diagrama de blocos que representa o algoritmo implementado.São descritos os vários passos necessários para o processamento do mesmo. . . . .	13
2.5	Processo de recorte aplicado a um sinal conhecido $x[n]$ . (a) - Sinal conhecido $x[n]$ com frequência de 0.8Hz. (b) - Passo do corte pelo nível $cL$ . . . . .	15
2.6	Comparação entre os resultados obtidos após auto-correlação e os limites definidos aplicada a uma sequência não periódica e a uma sequência periódica conhecida $x[n]$ . (a) - Sequência periódica conhecido $x[n]$ de frequência 0.8 Hz; (b) - Auto-correlação do sequência periódica $x[n]$ , onde estão representados o primeiro mínimo a verde e o limiar -0.5 com linha tracejada; (c) - Sequência não periódica; (d) - Auto-correlação da sequência não periódica, com limiar 0.3 do AUTOC representado com linha a preto tracejada e novo limiar definido -0.5 a linha vermelha tracejada; . . . . .	16
2.7	Ilustração das componentes X, Y e Z de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto circular; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação. . . . .	18

2.8	Ilustração das componentes $X$ , $Y$ e $Z$ de um gesto periódico para a direita e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas $X$ ; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação. . . . .	18
2.9	Ilustração das componentes $X$ , $Y$ e $Z$ de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas $Z$ ; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação. . . . .	19
2.10	Ilustração das componentes $X$ , $Y$ e $Z$ de um gesto não periódico e resultados obtidos pelos métodos implementados. (a) - Movimento aleatório; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação. . . . .	20
3.1	Alguns gestos periódicos. (a) - Gesto que dá a indicação de movimento para cima; (b) - Gesto que dá a indicação de movimento para a direita; (c) - Gesto circular; (d) - Gesto de acenar. . . . .	22
3.2	Representação da transformação linear $\mathbb{R}^3$ para $\mathbb{R}^2$ . (a) - Representação de um gesto para a direita num sistema de coordenadas cartesiano; (b) - Representação do gesto em (a) num sistema de coordenadas em componentes principais; (c) - Representação de um gesto circular num sistema de coordenadas cartesiano; (d) - Representação do gesto em (c) num sistema de coordenadas em componentes principais; . . . . .	24
3.3	Representação da distribuição e da contribuição dos vectores próprios $v_1$ e $v_2$ nos dados. . . . .	25
3.4	Representação gráfica do vector próprio $v$ , indicando sua direcção. . . . .	27
3.5	Gesto para cima e para baixo com a representação dos vectores velocidade, onde os vectores representados por cor azul indicam baixa velocidade e os de cor vermelha elevada velocidade. (a) - Gesto para cima; (b) - Gesto para baixo. . . . .	27
3.6	Resultado obtido após aplicada a derivada aos dados que representam a trajectória do gesto. (a) - Resultado da derivada aplicado ao gesto para a esquerda; (b) - Resultado da derivada aplicado ao gesto para a direita. . . . .	28
3.7	Esquema que representa o algoritmo efectuado para a contagem dos picos negativos e positivos da derivada. . . . .	29
3.8	Representação do passo de recorte do algoritmo de contagem de picos negativos e positivos da derivada. (a) - Representação gráfica da velocidade antes do recorte; (b) - Representação gráfica da velocidade depois do recorte. . . . .	30
3.9	Esquema que ilustra todo o processo de reconhecimento de gestos periódicos. . . . .	30
3.10	Esquema montado de maneira a realizar a experiência. (1) - Sujeito que indica direcções; (2) - Sujeito que recebe as direcções e percorre caminho; (3) - Ilustração do mapa; (4) - Sensor Xbox Kinect; (5) - Sensor Optotrak. . . . .	31
3.11	Ilustração do caminho efectuado para testar experiência. (a) - Ilustração do caminho; (b) - Forma como o caminho era mostrado. . . . .	32

3.12	Trajectórias obtidas pelo Xbox Kinect e pelo Optotrak. (a) - Trajectória obtida pelo Xbox Kinect; (b) - Trajectória obtida pelo Optotrak. . . . .	34
4.1	Ilustração do método implementado para o gesto apontar. . . . .	36
4.2	Comparação entre os dados originais e os filtrados após a intervenção do algoritmo. . . . .	38
4.3	Resultado do gesto apontar. O ponto branco representa a localização para onde o gesto está direccionado. . . . .	39
4.4	Ilustração do ponto marcado para testar a precisão do sistema. . . . .	40
4.5	Resultados obtidos para a medição da precisão. Comparação do entre ponto médio de 30 amostras e o ponto marcado a 3 m de profundidade e a 1 m para a direita do sensor, tendo este como referencia. . . . .	40
4.6	Resultados obtidos depois de traçada a forma circular. (a) - Trajectória sem recorrer à filtragem; (b) - Trajectória com filtragem. . . . .	41
5.1	Representação de dois gestos naturais e intuitivos. (a) - Gesto acenar; (b) - Gesto apontar. . . . .	44
A.1	Progressão no tempo da coordenada Y de um gesto com movimento vertical e a respectiva análise espectral com a frequência corresponde de 1.406Hz. (a) - Coordenada Y de um gesto com direcção vertical; (b) - Análise espectral. . . . .	53
A.2	Progressão no tempo da coordenada X de um gesto com movimento circular a respectiva análise espectral com a frequência corresponde de 0.479Hz. (a) - Coordenada X de um gesto com movimento circular; (b) - Análise espectral. . . . .	54
A.3	Progressão no tempo da coordenada X do gesto acenar a respectiva análise espectral com a frequência corresponde de 0.7Hz. (a) - Coordenada X do gesto acenar; (b) - Análise espectral. . . . .	54
A.4	Progressão no tempo da coordenada Z de um gesto com movimento horizontal e a respectiva análise espectral com a frequência corresponde de 0.937Hz. (a) - Coordenada Z de um gesto com movimento horizontal; (b) - Análise espectral. . . . .	54
A.5	Ilustração das componentes X, Y e Z de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas X; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação. . . . .	55
A.6	Ilustração das componentes X, Y e Z de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas Y; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação. . . . .	55
B.1	Alguns gestos periódicos. (a) - Gesto que dá a indicação de movimento para baixo; (b) - Gesto que dá a indicação de movimento para a esquerda; (c) - Gesto que dá a indicação de movimento para a frente; (d) - Gesto que dá a indicação de movimento para a trás. . . . .	57





# Lista de Tabelas

3.1	Resultados obtidos pela experiência através dos dados do Xbox Kinect, sendo "Sim" a indicação que o método funcionou e "Não" caso contrário. .	32
3.2	Comparação entre os resultados obtidos pelo Optotrack e pelo Xbox Kinect. Sim - indica que o método funcionou; Não - indica que o método não funcionou. . . . .	33



# Abreviaturas

IHM	Interacção Homem-Máquina
AUTOCL	<i>Modified Autocorrelation Method Using Clipping</i>
ACP	Análise das Componentes Principais
DVS	Decomposição em Valores Singulares
FFT	<i>Fast Fourier Transform</i>
DFT	<i>Discrete Fourier Transform</i>



# Capítulo 1

## Introdução

Nas relações interpessoais, a comunicação efectua-se maioritariamente sob forma verbal. No entanto, esta forma de comunicação não se pode considerar padrão devido aos seguintes factores: na existência de dialectos diferentes, as pessoas não podem comunicar verbalmente entre si se não falam a mesma língua; na existência de ruído sonoro de elevada intensidade, levando à incapacidade de comunicação de forma verbal; quando estamos na presença de pessoas com deficiência auditiva ou oral, existindo a necessidade de uso da linguagem gestual.

Neste sentido, a utilização de gestos surge como outra forma de comunicação, mas tal como a forma verbal existem cenários específicos onde esta necessita de uma aprendizagem prévia para os aplicar. Exemplo disso, na regulação de tráfego aéreo no solo, são executados gestos que requerem um conhecimento da sua semântica. Para além destes, existem gestos simples e versáteis feitos intuitivamente e de fácil interpretação, tais como a orientação, a saudação. Estes são utilizados para descrever situações da nossa rotina diária, onde é mais importante apontar para a posição de um objecto do que descrevê-la oralmente, sendo intuitiva a forma como indicamos a direcção para qual outra pessoa tem de se deslocar. São também gestos que, mediante o contexto, têm o mesmo significado em qualquer parte do mundo.

Contudo, a comunicação não existe apenas entre seres vivos. No último século, tem vindo a verificar-se a necessidade de interagir com todo o tipo de meios tecnológicos. Estes aparecem do mais antigo, como a manipulação de máquinas no ambiente fabril, até ao mais simples, como o pressionar de um botão num controlo remoto. Estes requerem um tipo de comunicação adaptativa por parte do utilizador. Na actualidade, cada vez mais o Homem precisa de conviver com a Máquina, e esta necessita de uma adaptação à forma como comunicamos com ela. É da natureza humana evitar novas e complexas formas de comunicação que tornem a interacção com a máquina penosa, dando origem a novas

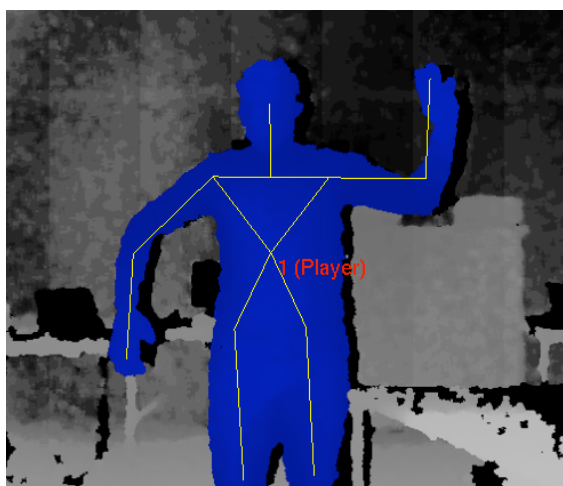
tecnologias com o objectivo de facilitar a adaptação na comunicação, tais como: câmaras, monitores capacitivos e resistivos, sensores de aproximação, entre outros.

A interacção com a máquina não surge apenas no meio industrial mas também em situações quotidianas ou triviais. Temos o caso dos designados *care robots* que vêm colmatar as necessidades de pessoas carentes de cuidados médicos, de efectuar tarefas diárias a pessoas com dificuldades motoras, e até mesmo diminuir a exclusão social que existe, maioritariamente, nas pessoas de terceira idade. No entanto, tem vindo a verificar-se um grande desafio na Interacção Homem-Máquina (IHM) por forma a tornar a comunicação mais simples, pois no caso das pessoas de terceira idade com dificuldades de movimentos e com atrofia nos membros torna-se difícil o uso de um rato ou teclado.

Desta maneira, esta dissertação propõe fornecer uma componente de interacção gestual intuitiva e natural entre o Homem e a Máquina de forma a facilitar a comunicação entre eles. Assim, o utilizador não necessita de um conhecimento prévio de gestos específicos para efectuar tarefas simples, como direccionar um robô para uma posição ou mesmo desviá-lo de uma trajectória. Esta componente pode ser aplicada num ambiente industrial onde existem plataformas de transporte, por forma a proceder ao seu controlo. Caso a rotina desta seja interrompida por uma causa inesperada, o operador tem a possibilidade de ordenar à plataforma o retorno à sua tarefa através de um simples gesto de orientação que surge de forma natural, sem recorrer à utilização de quaisquer objectos periféricos. Ou seja, o que se pretende é utilizar gestos simples que nos são característicos de modo a comunicar com máquinas tecnologicamente evoluídas.

## **1.1 Motivação, objectivos e contribuição**

Toda a tecnologia sofre um processo evolutivo, e como tal a IHM tem de acompanhar esta evolução. A simplicidade e a naturalidade que se pede na sua realização é a premissa necessária para que este tema esteja em constante desenvolvimento. Isto mostra que, para além de se querer evoluir nas formas de interacção, há uma tentativa de as tornar o mais naturais e intuitivas possíveis, tornando-se o principal motivo para o estudo deste trabalho. São necessários mecanismos de interacção gestual natural e intuitiva que não obriguem os utilizadores a ter nenhum conhecimento "à priori" para a interacção simples com a Máquina. Outro motivo que levou ao desenvolvimento deste trabalho, foi o aparecimento de sensores RGB-D, como sejam a Xbox Kinect, PrimeSense PS1080 e Asus Xtion, e a disponibilidade de uma biblioteca [2] que usando os dados destes sensores efectua a estimação dos parâmetros cinemáticos de um modelo simplificado do esqueleto humano (ver Fig. 1.1).



**Figura 1.1** – Seguimento do esqueleto humano adquirido através da interface de programação de aplicações OpenNI complementada com a tecnologia do Xbox Kinect.

Como veremos mais à frente, muitos foram os investigadores que, nas últimas décadas, se têm dedicado ao seguimento dos gestos humanos. Este tipo de sensores veio resolver parcialmente este problema, dando a possibilidade de nos dedicarmos ao estado seguinte, ou seja a análise e interpretação dos gestos.

Por forma a desenvolver a componente gestual de interacção desejada, o objecto de estudo recaiu sobre dois tipos de gestos: gestos periódicos e gestos deícticos ou demonstrativos.

Por gestos periódicos entendem-se todos aqueles que são feitos com a mão horizontal ou verticalmente, em qualquer direcção com uma determinada repetição. Obviamente que falamos de gestos com significado, que representam indicação de uma acção. Estes utilizados para indicar manobras, mandar parar e avançar, podem ser muito úteis para uma comunicação simples com os robôs.

No caso dos gestos deícticos, são todos aqueles que se caracterizam por serem gestos demonstrativos ou direccionais. Estes são também úteis, pois podem contextualizar um pedido que se faça a um robô, indicando-lhe um objecto, uma pessoa ou um lugar.

Em termos objectivos, esta dissertação visa criar um sistema capaz de processar e reconhecer os gestos descritos, de uma forma eficaz e sem utilizar muitos recursos computacionais. Recorrendo ao uso destes gestos, um potencial utilizador pode operar o robô mesmo que não conheça o mecanismo de interacção. Pretende-se que não seja necessário utilizar nenhum tipo de acessório, como por exemplo, luvas coloridas, marcadores, etc. Tomando partido do uso da biblioteca e dos sensores referidos, é possível oferecer um custo monetário reduzido na sua utilização. A análise gestual é feita com gestos naturais e intuitivos, conferindo dinamismo ao sistema e não se tornando penosa a sua utilização.

São estes os requisitos necessários para o bom funcionamento do sistema apresentado.

Os resultados alcançados mostram a viabilidade da utilização deste tipo de interacção, tendo os mesmos sido publicados nas actas da conferência *Ro-MAN 2012* [3].

## 1.2 Estado de arte

Na última década, foram vários os projectos de investigação dedicados ao reconhecimento gestual, especialmente em tentativas de criar linguagens gestuais para se comunicar com robôs [4, 5, 6, 7]. Na verdade, o reconhecimento da linguagem gestual é um problema muito complexo, com base apenas no processamento de imagem, e a sua utilização é limitada a um pequeno grupo de pessoas.

Em [4] é usada a análise das componentes principais (ACP) para o reconhecimento de 25 gestos do alfabeto gestual, uma técnica comum para encontrar padrões em dados de grandes dimensões. O processamento que a ACP vai executar neste tipo de sistemas é, num grupo de imagens da pose de cada gesto, retirar delas as componentes principais que evidenciam a mesma, reduzindo as características das poses dos gestos a analisar. Assim, usando as componentes principais é possível criar padrões de classificação de forma a reconhecer o gesto. Podendo assim considerar que a ACP é o primeiro passo para o processamento do reconhecimento de gestos, onde este encontra as características necessárias para a criação de padrões. Em [8, 9] é também utilizada a ACP para reduzir a dimensão dos dados a processar, classificando de seguida as poses do gestos em questão. Estes classificadores podem utilizar métodos como os *Hidden Markov Models*, ou *Support Vector Machines*, estes têm como função reconhecer os padrões criados pela ACP e atribuindo-lhes assim o significado desejado. Um dos inconvenientes deste tipo de sistemas são as fases de treino necessárias para a criação de padrões de reconhecimento. São necessários vários treinos, pois só desta forma é possível encontrar um padrão robusto para que na fase de classificação o processo seja feito de forma adequada. A implementação destes sistemas está dependente das fase de treino feitas para a criação de padrões, tornando-o pouco dinâmico quando este sistema é utilizado por diferentes pessoas, isto é quando o padrão é feito por uma determinada pessoa não significa que este vá funcionar para todas [4, 8, 9].

Outro sistema para o reconhecimento de postura das mãos, de forma a criar uma linguagem gestual, consiste em achar correspondência em contornos e silhuetas em imagens de poses de mãos, relacionando assim um modelo correspondente de modo a identificar o gesto pretendido. São várias as pesquisas no sentido de utilizar esta técnica para o



reconhecimento de gestos [7, 12]. No entanto, este tipo de técnica é utilizada, maioritariamente, na parte do seguimento. Em [10, 11] é empregado um alinhamento de silhueta e uma técnica de correspondência para reconhecer a silhueta de uma mão numa imagem e, posteriormente, fazer o seu seguimento.

Os *Hidden Markov Models* (HMM) são modelos estatísticos em que um conjunto de parâmetros ocultos é determinado a partir de um conjunto de parâmetros relacionados, observáveis. Nestes, o estado não é directamente observável, mas as variáveis que envolvem o estado são. Cada estado tem uma probabilidade de distribuição sobre as possíveis símbolos de saídas. Portanto, a sequência de símbolos geradas pelos HMM fornece informação sobre as sequências de estados. No contexto do reconhecimento de gestos, os parâmetros observáveis são estimados pelo reconhecimento das posturas dos gestos nas imagens [13, 14, 15]. Por esta razão e porque os gestos podem ser reconhecidos como sequências de posturas, HMM têm sido muito utilizados para o reconhecimento de gestos. HMM são utilizados para representar os gestos e os seus parâmetros são adquiridos a partir de sessões de treino. Baseado no critério "*the most likely performance*", os gestos podem ser reconhecidos através da avaliação dos HMM treinados.

Os métodos, até agora referidos, utilizam um processamento sobre um referencial duas dimensões (2D), ou seja estão limitados a um processamento num sistema de eixos  $X$  e  $Y$ . No entanto, tem-se vindo a verificar um desenvolvimento na tecnologia de sensores com visão em três dimensões (3D), onde é adicionado para o seu processamento um terceiro eixo,  $Z$ , também designada como profundidade.

Existem, na actualidade, três tipos de tecnologias que podem adicionar informação de profundidade às cenas de análise: visão estereoscópica, padrões de luz codificada e tempo de voo [16].

O método de aquisição da componente de profundidade mais comum é o sistema de visão estereoscópica, que utiliza duas câmeras para obter uma imagem *stereo*. Estas imagens estão ligeiramente deslocadas da mesma forma que os olhos humanos estão. O processamento é feito comparando as duas imagens, resultando numa imagem de disparidade que relaciona o deslocamento de objectos na imagem, e desta forma obtido um mapa de profundidade [17]. São alguns métodos [18, 19] que usam a imagem de disparidade para proceder ao reconhecimento de gestos. Contudo, esta tecnologia apresenta alguma dificuldade em conseguir uma imagem com boa qualidade para extrair as características necessárias para o seu reconhecimento.

As câmaras com tecnologia de tempo de voo tornaram-se populares nos últimos anos dadas as suas capacidade de adquirir um mapa de profundidade. No entanto, a tecnologia obtém mapas de profundidade de baixa resolução e os dados apresentam-se corrompidos com ruído, para além disso, são sensores que têm um elevado custo. Estes englobam já

um elevado número de aplicações [20, 21, 22, 23] servindo de controlador para jogos, reconhecimento de gestos, navegação de robôs, entre outros.

Dentro da gama de sensores que utilizam padrões de luz codificada, surgiram, recentemente, um conjunto de câmaras, que apresenta mapas de profundidade com uma maior qualidade, e com um custo mais reduzido quando comparadas com as de tempo de voo, estamos a falar de câmaras como a Xbox Kinect, PrimeSense PS1080, Asus Xtion.

A Xbox Kinect é o modelo utilizado nesta dissertação, o seu funcionamento baseia-se em imagens de luz codificadas para o cálculo do mapa de profundidade. Onde este resulta da deformação de um padrão de luz predefinido adquirido pela câmara, quando este é projectado sobre um objecto. Este usa um projector de luz de infravermelhos com um padrão em manchas e um sensor de captura, o cálculo do mapa de profundidade é feito sem qualquer uso de uma câmara RGB. Para o processo de seguimento do esqueleto humano, foi feito um treino de mais de 1 milhão de exemplos de poses, onde as partes do corpo são adquiridas através do mapeamento das imagens de profundidade [24].

Existem já alguns investigadores que tentaram explorar este no tipo de sensores, mas sempre com o intuito de criar uma linguagem gestual onde determinam as poses das mãos [25, 26].

Nos gestos onde esta dissertação se debruça é do nosso conhecimento que, nos últimos anos, apenas alguns investigadores se interessaram nos gestos deícticos [27, 28] e periódicos [29, 30].

### **1.3 Estrutura da tese**

No presente trabalho são desenvolvidas mecânicas de reconhecimento de gestos periódicos e deícticos ou demonstrativos. Inicialmente é introduzido o tema que levou à exploração dos mesmos, e nos capítulos seguintes é descrito qual o desenvolvimento efectuado bem como os seus resultados. No final, foi elaborado um capítulo onde mostra a viabilidade dos métodos, as suas práticas e futuras melhorias e aplicações.

No Cap. 2 é feito o desenvolvimento para a detecção dos gestos periódicos, onde são demonstrados dois métodos para a fazer. São dadas as explicações do funcionamento do método com recurso à FFT e com recurso à Auto-correlação na secção 2.1. Na secção 2.2, são analisados os resultados obtidos pelos dois métodos aplicados a vários gestos.

Um método para o reconhecimento dos gestos referidos é elaborado no Cap. 3, onde é mostrado o seu desenvolvimento na subsecção seguinte. Para validar o mesmo foi feita uma experiência, onde é descrita e analisada na secção 3.2.

No Cap. 4 é desenvolvido um mecanismo que recria um gesto demonstrativo, onde nas suas secções seguintes é descrito o desenvolvimento do mesmo e analisados os seus resultados. Também é referido e aplicado um método de suavização de ruído nos parâmetros utilizados, de forma a tornar o método mais robusto, na secção 4.2.

Por fim, no Cap. 5 é feita uma conclusão concisa acerca de todo o trabalho realizado nesta dissertação de mestrado, indicando os benefícios que os sistemas de IHM terão com a inclusão dos gestos referidos. São também apresentados alguns casos de trabalho futuro na secção 5.1, descritas alterações de forma a melhorar os resultados deste trabalho, e outras possíveis aplicações.



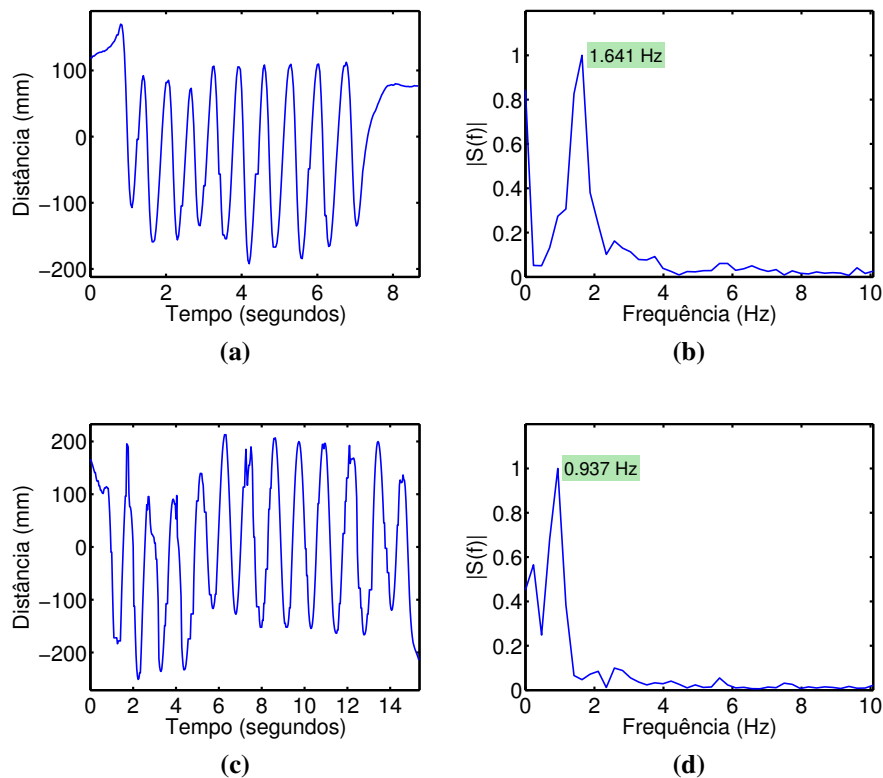
# Capítulo 2

## Gestos Periódicos

Um gesto periódico pode ser definido como o movimento repetitivo da mão de uma pessoa. Dependendo da direcção, significa olá, mover para a direita, mover para a esquerda, mover para a frente, entre outros. Esta classe de gestos simples é utilizada intuitivamente por nós quando queremos indicar acções como a orientação, chamar a atenção ou cumprimento, estes como surgem naturalmente na nossa rotina diária não necessitam de qualquer aprendizagem. O movimento repetitivo traduz-se matematicamente na seguinte equação:

$$S[n] = S[n + mN], \forall n \in \mathbb{Z}, m \in \mathbb{Z}, \quad (2.1)$$

onde  $S[n]$  representa um sinal discreto, onde os seus valores de coordenadas são obtidos ao ritmo da frequência de amostragem, que neste caso é 30 Hz, e  $N$  corresponde ao período fundamental do mesmo. Analisando o sinal no domínio da frequência, obtido pela transformada de Fourier, é possível concluir, pelo espectro de frequência que estes gestos são normalmente executados, por diferentes pessoas, numa gama de frequências entre 0.45 Hz a 2 Hz. Por consequência para detectar estes gestos é necessário detectar a presença de uma trajectória periódica da mão com a gama de frequência dadas. Na Fig. 2.1 encontram-se exemplos de gestos periódicos e as suas frequências correspondentes. Assim como no anexo A.1, onde estão representados mais exemplos de gestos periódicos com as suas respectivas frequências dentro da gama referida.



**Figura 2.1** – Progressão no tempo da coordenada X do um gesto com direcção horizontal e a respectiva análise espectral com a frequência corresponde de 1.641Hz. (a) - Coordenada X de um gesto com direcção horizontal; (b) - Análise espectral. (c) - Coordenada Y de um gesto com direcção vertical; (d) - Análise espectral

Após a detecção do gesto periódico é necessário classificá-lo, pois apenas com a sua detecção não é possível demonstrar uma acção a quem o está a interpretar. Posto isto, temos que lhes atribuir um significado para que seja classificado nas várias intervenções. Significado que pode ser caracterizado como sentidos para a direita, para a esquerda, para a frente, para baixo, entre outros. Na secção seguinte serão explicados dois procedimentos, os quais, foram utilizados para a detecção dos gestos periódicos, bem como os resultados obtidos.

## 2.1 Detecção de Gestos Periódicos

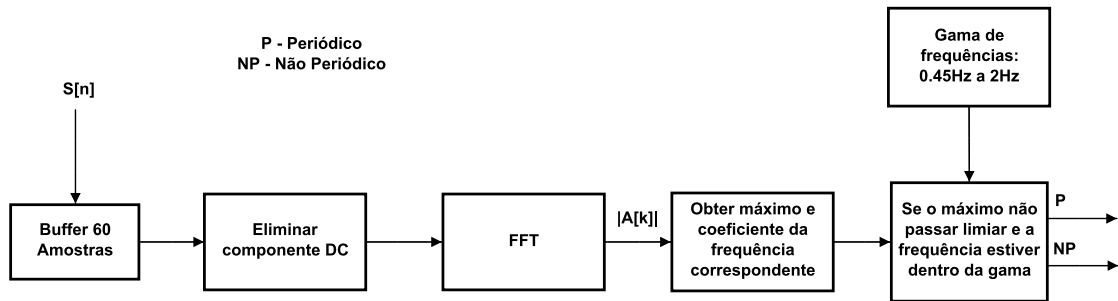
### 2.1.1 Método com recurso à FFT

Em seguimento ao referido anteriormente, sobre a detecção de gestos periódicos, utiliza-se como abordagem inicial lógica a análise da gama de frequência na qual os gestos se inserem. Como os dados que representam as trajectórias da mão são periódicas e uma

grande maioria da frequência fundamental destes está dentro da gama referida (0.45 Hz a 2 Hz), foi usada *Fast Fourier Transform* (FFT) de forma a analisar estas características.

Assim, ao aplicar a FFT, vamos explorar a informação que esta nos dá em função do domínio da frequência. Informação que, no domínio do tempo não é muito importante, mas no domínio da frequência tem um papel de, no problema corrente, permitir a classificação do gesto como periódico, falando assim da amplitude e frequência.

Na Fig. 2.2 está representado sucintamente o diagrama do procedimento utilizado. Como se está a lidar com uma implementação que recorre a processamento em tempo real foi necessário criar uma sequência de amostras na qual são retidas para proceder à sua análise.



**Figura 2.2** – Diagrama de blocos que representa o método implementado. São descritos os sucessivos passos necessários ao funcionamento do mesmo.

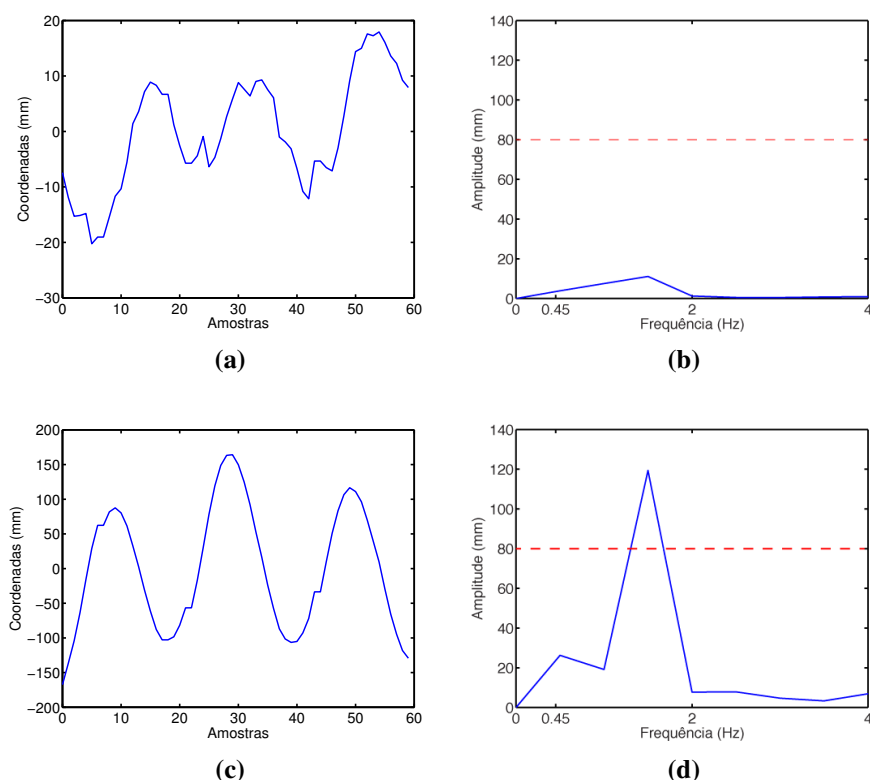
Foi escolhido um tamanho de 60 amostras, pois verificou-se que este consegue conter informação, de pelo menos, um período de sinal do gesto periódico, e assim suficiente para a detecção da periodicidade do gesto. Inicialmente é removida a componente contínua (componente DC) da sequência de amostras, pois é útil para uma mais fácil análise do espectro de frequência. Esta provém do facto de os dados serem amostrados de acordo com o referencial (0,0,0) do Xbox Kinect. De seguida, a esta sequência de 60 amostras, com uma frequência de amostragem de 30 Hz, é aplicada a FFT, que não é mais que um método computacionalmente eficiente de calcular a DFT,

$$A[k] = \sum_{n=0}^{N-1} S[n] e^{\frac{(-j2\pi)(kn)}{N}}, k = 0, \dots, N-1, \quad (2.2)$$

onde  $A[k]$  correspondem aos coeficientes de Fourier da sequência de amostras  $S[n]$ . O valor máximo destes coeficientes corresponderá à frequência fundamental da sequência de amostras,  $f_s$ .

Dada a  $f_s$ , é então verificado se esta se encontra na gama de frequências, definida manualmente, sendo escolhido neste caso 0.45 Hz a 2 Hz, pois é nesta gama que se encontram os gestos que se querem identificados. Caso esta se encontre dentro da gama

referida e a sua amplitude estiver acima de um limiar previamente definido, a sequência é então rotulada como periódica, caso contrário, é rotulada como não periódica. Este limiar é inserido de maneira a prevenir gestos mal executados ou ruído, e corresponde à metade da amplitude do gesto em milímetros nessa componente. Desta forma, definiu-se um limiar de 80 mm para amplitude da frequência fundamental, isto é, uma amplitude pico a pico de 160 mm do gesto executado. A razão para a escolha deste valor deve-se ao facto de ter sido aquele que mostrou ter melhores resultados. Na Fig. 2.3 está ilustrado um exemplo do fenómeno referido, onde a Fig. 2.3 (b) corresponde ao espectro de frequências de um sinal não periódico, mas apresenta uma frequência fundamental dentro da gama desejada. No entanto, a sua amplitude apresenta um valor abaixo daquele que é considerado para validar como gesto periódico.



**Figura 2.3** – Passo do método onde é feita a selecção de uma sequência periódica e não periódica, entre a máxima amplitude da  $f_s$  (representado por uma linha vermelha) e a gama de frequências 0.45 Hz e 2 Hz. (a) - Sequência de 60 amostras de uma trajectória não periódica sem componente DC; (b) - Espectro de frequência correspondente a ao sinal (a); (c) - Sequência de 60 amostras de uma trajectória periódica sem componente DC; (d) - Espectro de frequência correspondente a ao sinal (c).

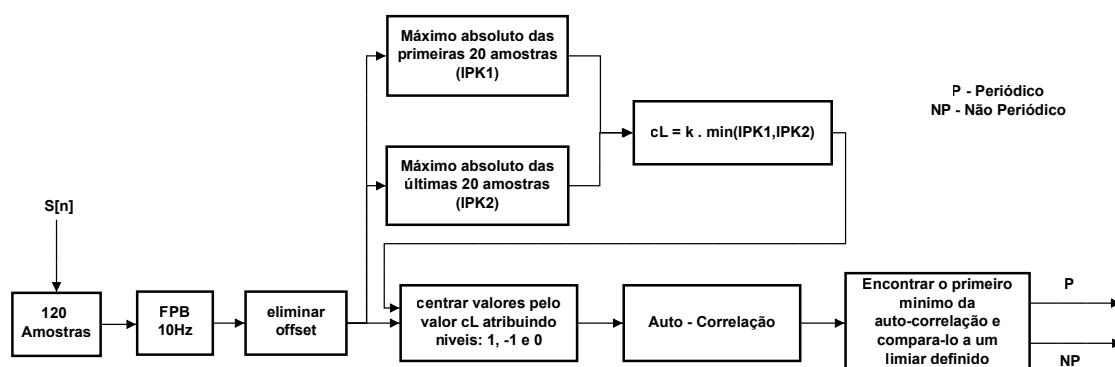
Neste momento é apresentado um método capaz de detectar a presença de um gesto periódico, renegando gestos com movimentos aleatórios. No entanto, é um método que requer alguma afinação dos limiares escolhidos para que este tenha resultados válidos.



## 2.1.2 Método com recurso à auto-correlação

Um outro método implementado nesta dissertação é semelhante ao utilizado para descobrir o período de sinais de voz em streams de áudio. Ambos, os gestos periódicos e os sinais de áudio, em caso de detecção de voz, têm uma frequência fundamental e uma determinada amplitude. Para além disso, ambos estão normalmente corrompidos com ruído, que necessita de ser eliminado para que a detecção se concretize mais facilmente.

Dados estes factores, é usada uma adaptação de um algoritmo de detecção de voz, mais concretamente o *Modified Autocorrelation Method Using Clipping* (AUTOC) [32]. Este algoritmo tem um princípio de funcionamento bastante simples, que teve que ser adaptado para o problema corrente. Isto deve-se ao facto de estar a lidar com gamas de frequência muito baixas e sequências com número de amostras muito inferiores quando comparados com os dos sinais de voz.



**Figura 2.4** – Diagrama de blocos que representa o algoritmo implementado. São descritos os vários passos necessários para o processamento do mesmo.

Na Fig. (2.4) está representado o diagrama de blocos onde mostra os passos do algoritmo. Em cada 120 amostras ( $S[n], n = 1, \dots, 120$ ), a sequência é submetida a um filtro passa-baixo (filtro FIR de ordem 5) com uma frequência de corte de 10Hz de modo a eliminar o ruído de alta frequência, tendo em conta a gama de frequência dos gestos. Em seguida, foi necessário eliminar o *offset*, centrando a sequência de amostras o recorte é feito de forma a não ter em conta o deslocamento das posições em relação à origem do referencial. Neste sentido, foi determinado o  $S_{max}$  (2.3) e  $S_{min}$  (2.4) e usando a expressão (2.5) a componente contínua da sequência é eliminada.

$$S_{max} = \max(S[n]), n = 1, \dots, 120 \quad (2.3)$$

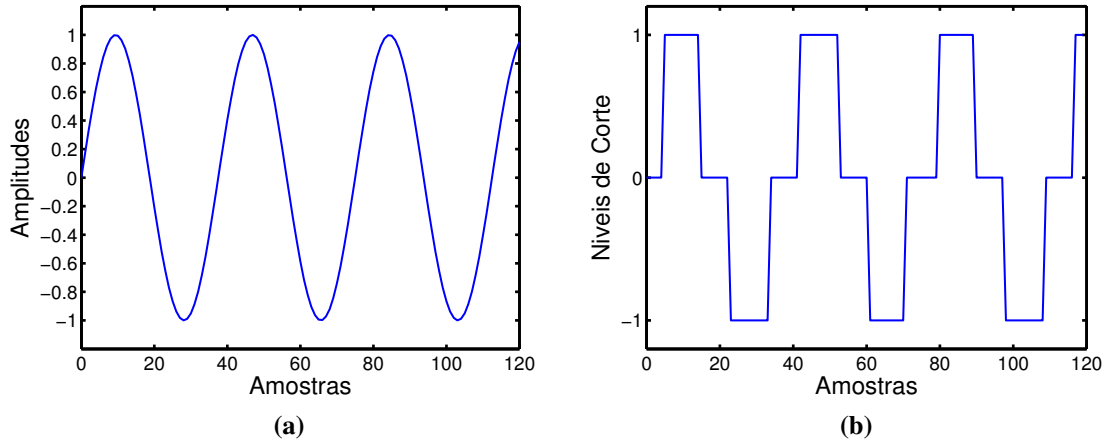
$$S_{min} = \min(S[n]), n = 1, \dots, 120 \quad (2.4)$$

$$S_c[n] = S[n] - \frac{S_{min} + S_{max}}{2}, n = 1, \dots, 120 \quad (2.5)$$

No próximo passo é feito o recorte pelo nível  $cL$  para cada sequência de 120 amostras. Após a computação do máximo absoluto para ambas as primeiras, e últimas 20 amostras das 120,  $IPK1$  e  $IPK2$  respectivamente. O  $cL$  é determinado da seguinte forma:  $cL = k \times \min(IPK1, IPK2)$ , onde  $k = 0.64$ . Usando este nível de corte (ver Fig. (2.5b)), uma sequência binária de valores centrados em 0 é gerada a partir do sinal de entrada como se segue:

$$S_c[n] = \begin{cases} 1 & , \text{ se } S_c[n] \geq cL \\ -1 & , \text{ se } S_c[n] \leq -cL, \\ 0 & , \text{ outros casos.} \end{cases} \quad n = 1, \dots, 120 \quad (2.6)$$

A auto-correlação normalizada da sequência binária centrada é de seguida obtida, e da sequência resultante é encontrado o primeiro mínimo. Se este mínimo estiver abaixo de  $-0.5$ , que define o limiar de selecção, então a sequência de 120 amostras é considerada periódica, caso contrário é rotulada como não periódica.

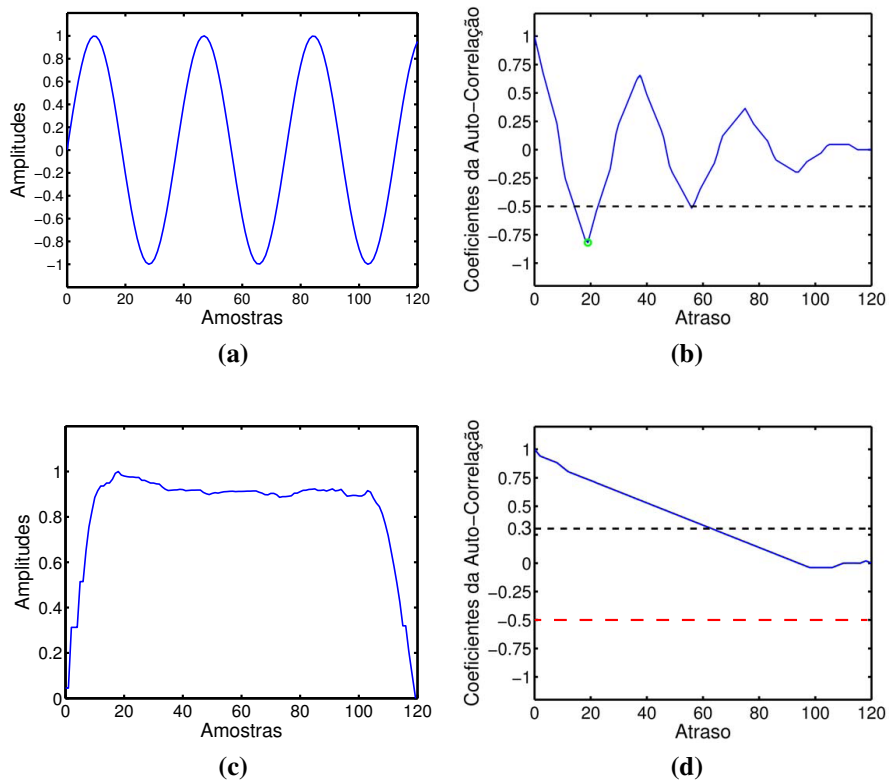


**Figura 2.5** – Processo de recorte aplicado a um sinal conhecido  $x[n]$ . (a) - Sinal conhecido  $x[n]$  com frequência de 0.8Hz. (b) - Passo do corte pelo nível  $cL$ .

É importante notar que este limiar escolhido é de facto diferente do algoritmo AUTOC, neste é considerado o primeiro valor máximo da auto-correlação normalizada, onde para um atraso ( $\tau$ ) nulo está localizado o pico principal da auto-correlação e o primeiro valor máximo é o pico a seguir. Na realidade, o primeiro mínimo, no atraso equivalente, corresponde a metade do período do sinal (2.7) (atingido quando, ao longo da auto-correlação do sinal, este se encontra em oposição de fase), enquanto que o primeiro máximo corresponde, no atraso equivalente, ao período do sinal (ver Fig. 2.6 (a) e (b)).

$$r[\tau_{min}] = \min r[\tau], \quad \frac{f_s}{2} = \frac{1}{\tau_{min}} \quad (2.7)$$

A variação deste algoritmo implementado para o AUTOC deve-se ao facto de este, na presença de uma sequência de amostras não periódicas, a auto-correlação gerar um resultado que impossibilita a escolha do primeiro máximo (ver Fig. 2.6 (d)) para a selecção de uma sequência periódica e não periódica, impossibilitando o uso do limiar 0.3. Nas Fig 2.6 (c) e (d) é mostrado uma situação onde este fenómeno ocorre, onde ao utilizar o limiar 0.3 do algoritmo AUTOC, referido em [32], ia-se gerar uma sequência periódica quando na verdade não é. Desta forma, o algoritmo aqui implementado utiliza o limiar  $-0.5$ , como já foi referido, evitando o fenómeno anteriormente verificado.



**Figura 2.6** – Comparação entre os resultados obtidos após auto-correlação e os limiares definidos aplicada a uma sequência não periódica e a uma sequência periódica conhecida  $x[n]$ . (a) - Sequência periódica conhecido  $x[n]$  de frequência 0.8 Hz; (b) - Auto-correlação do sequência periódica  $x[n]$ , onde estão representados o primeiro mínimo a verde e o limiar -0.5 com linha tracejada; (c) - Sequência não periódica; (d) - Auto-correlação da sequência não periódica, com limiar 0.3 do AUTOC representado com linha a preto tracejada e novo limiar definido -0.5 a linha vermelha tracejada;

Agora este algoritmo permite, assim, detectar se um gesto é periódico ou não, e de seguida é possível a análise e identificação da semântica associada. Há que referir que este é um método que requer apenas um parâmetro de afinação, o limiar de selecção, tornando-o bastante robusto.

## 2.2 Comparação dos métodos e seus resultados

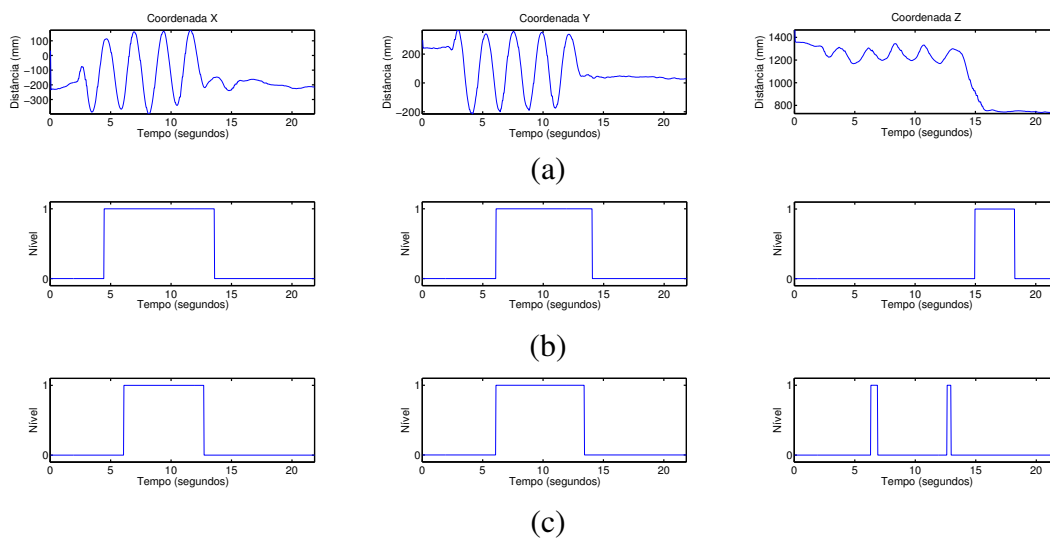
O que se pretende com estes resultados é avaliar se os algoritmos implementados, de facto, conseguem detectar quando um gesto periódico ocorre e se perante movimentos aleatórios que não são periódicos o sistema não apresenta comportamentos não esperados, como a detecção de falsos positivos. Este último é um ponto importante, pois como estamos a lidar com sistema de IHM, estes requerem que o sistema não tenha falhas no procedimento de detecção de gestos, para que não considere um movimento aleatório como um gesto periódico. Para proceder à análise dos métodos foram feitas aquisições de dados referentes a vários gestos, e estas adquiridas directamente através do sistema de seguimento fornecida pela biblioteca utilizada [2].

Nas subsecções seguintes estão ilustrados as trajectórias de 4 gestos, e sua posterior análise dos resultados obtidos depois de aplicado o método com recurso a FFT e a auto-correlação. Estão ilustrados as componentes  $X$ ,  $Y$  e  $Z$  de cada gesto e os resultados estão representados através de um gráfico de dois níveis, 1 e 0, que indicam se o gesto apresenta comportamento periódico ou não, respectivamente. Estes níveis são atribuídos após o processamento sucessivo das sequências de amostras respectivas a cada método, caso a sequência apresente comportamento periódico confere-se o valor 1, ou 0 caso contrário.

### 2.2.1 Gesto circular

Quando analisada a Fig. 2.7 (a) é possível verificar que existem duas componentes onde a periodicidade se destaca com maior intensidade, como se trata de um gesto circular existem duas coordenadas que determinam o gesto,  $X$  e  $Y$ , e é expectável que a periodicidade seja detectada primeiro num destes valores de coordenadas. Este facto verifica-se, quando ambos os métodos com recurso a FFT (Fig. 2.7 (b)) e a auto - correlação (Fig. 2.7 (c)) detectam a periodicidade nos eixo  $X$  e  $Y$ .

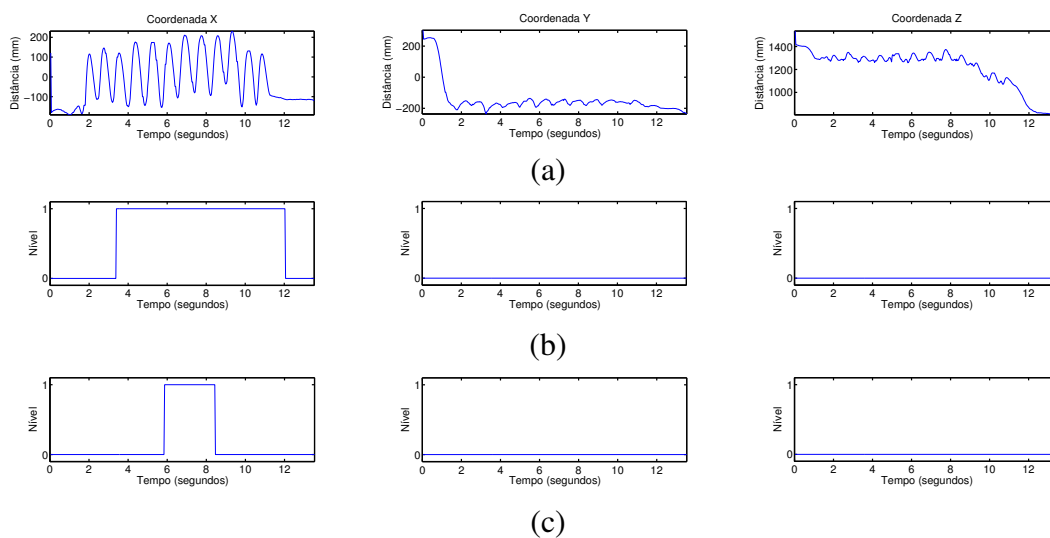
No entanto, também é possível visualizar que na componente  $Z$  do gesto são detectáveis, em ambos os métodos, fragmentos de periodicidade. Isto deve-se ao facto de existir um comportamento periódico de baixa intensidade no eixo  $Z$  do gesto, resultante da execução de gesto circular não perfeitamente paralelo ao plano  $XY$ . É importante referir que o gesto circular não sendo executado de forma natural e intuitiva, no entanto periódico, a sua análise foi feita, pois este apresenta duas componentes periódicas que ajudaram a testar a robustez dos métodos implementados.



**Figura 2.7** – Ilustração das componentes  $X$ ,  $Y$  e  $Z$  de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto circular; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação.

## 2.2.2 Gesto horizontal com movimento periódico sobre o eixo de coordenadas $X$

Como se trata de um gesto horizontal no eixo  $X$  (Fig. 2.8 (a)) o esperado é que esteja presente um comportamento periódico nesta componente. Este facto verifica-se, quando ambos os métodos com recurso a FFT (Fig. 2.8 (b)) e a auto - correlação (Fig. 2.8 (c)) detectam a periodicidade apenas no eixo  $X$ .

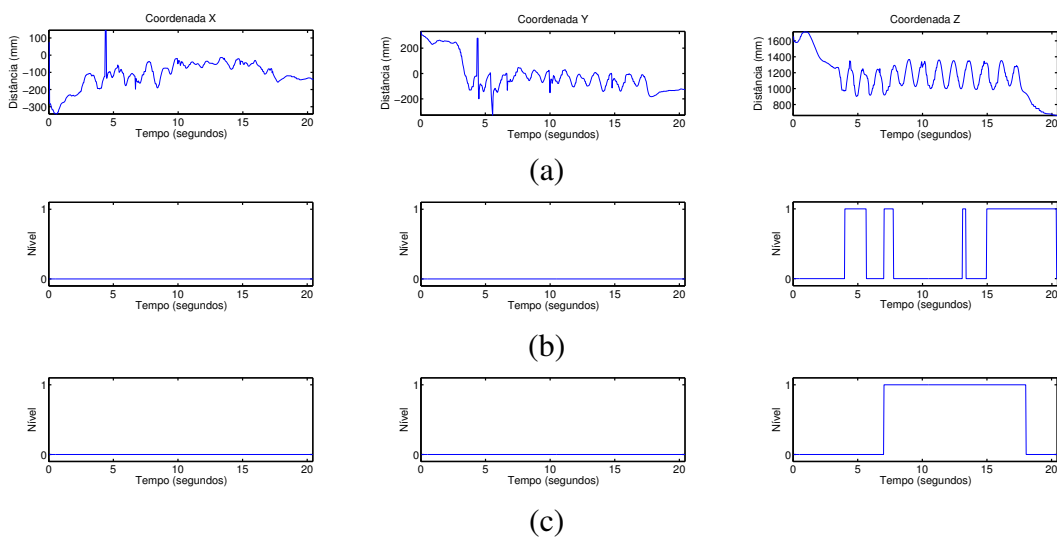


**Figura 2.8** – Ilustração das componentes  $X$ ,  $Y$  e  $Z$  de um gesto periódico para a direita e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas  $X$ ; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação.

Contudo, quando comparados os tempos de resposta dos dois métodos, verifica-se que o com recurso à FFT apresenta um tempo inferior ( $\approx 2$  segundos) em relação ao com recurso à auto-correlação ( $\approx 4$  segundos). O tempo de resposta é definido através do início da execução do gesto até este ser detectado como periódico. Assim, o método com recurso à FFT necessita que o gesto seja executado durante, pelo menos, 2 segundos, enquanto que o método com recurso à auto-correlação é necessário, pelo menos, 4 segundos. Este está dependente do tamanho da sequência de amostras pela qual a computação é feita, já que estamos a falar de uma sequência de 60 amostras contra uma sequência de 120 amostras, respectivamente.

### 2.2.3 Gesto horizontal com movimento periódico sobre o eixo de coordenadas Z

Como é um gesto executado horizontalmente com incidência no eixo Z (Fig. A.6 (a)) é expectável que esteja presente um comportamento periódico nesta componente. Este facto verifica-se, quando ambos os métodos com recurso a FFT (Fig. A.6 (b)) e a auto-correlação (Fig. A.6 (c)) detectam a periodicidade apenas no eixo Z.

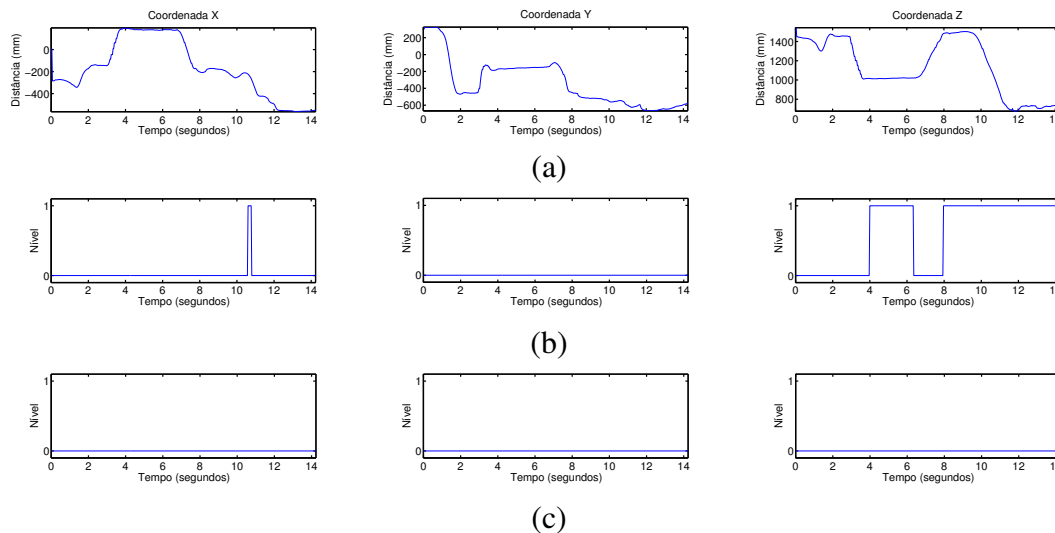


**Figura 2.9** – Ilustração das componentes X, Y e Z de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas Z; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação.

Apesar dos resultados mostrarem que o método com recurso à FFT consegue detectar periodicidade no eixo desejável, na verdade, desde que seja possível detectar que existe um comportamento periódico é possível classificar o gesto como tal. No entanto, é visível que o seu comportamento é inconsistente, não mostrando a mesma performance quando comparado com o resultado obtido pelo método com recurso a auto-correlação.

## 2.2.4 Movimento aleatório

Aqui é apresentado um movimento que não apresenta nenhum comportamento periódico (Fig. A.6 (a)), prevê-se que os resultados não apresentem qualquer ocorrência de periodicidade. Esta facta verifica-se no caso nos resultados obtidos pelo método com recurso à auto-correlação (Fig. A.6 (c)), já o método com recurso à FFT é notória a inconsistência do mesmo.



**Figura 2.10** – Ilustração das componentes  $X$ ,  $Y$  e  $Z$  de um gesto não periódico e resultados obtidos pelos métodos implementados. (a) - Movimento aleatório; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação.

Como no processamento do método com a FFT se utilizam três limiares para a selecção de sequências periódicas e não periódicas, a gama de frequências de execução do gesto (mínima e máxima) e o limiar da amplitude máxima de  $f_s$ , não foi possível encontrar um conjunto que mostrasse a consistência pretendida. Ou seja, encontrar uma relação entre a gama de frequências, em que os gestos periódicos estavam inseridos, e um limiar de selecção da amplitude máxima da  $f_s$  onde a inconsistência, aqui demonstrada, fosse dissipada. O facto de este método não apresentar os resultados que esperávamos ter, é a razão para a implementação do segundo método com recurso à auto-correlação. Neste aspecto, este segundo método apresentou melhorias significativas quanto à consistência do mesmo, e desta forma recaiu neste a escolha do método a utilizar para o desenvolvimento do mecanismo de reconhecimento de gestos.

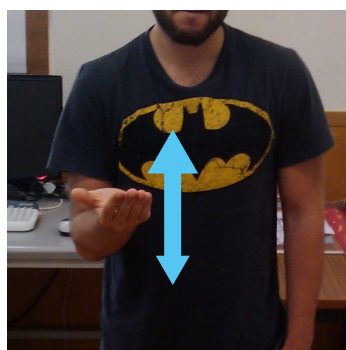


# Capítulo 3

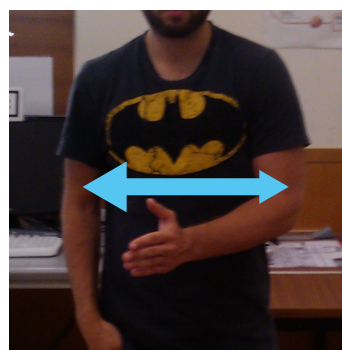
## Análise dos gestos periódicos

Após a detecção da presença de um gesto periódico, este é analisado de forma a atribuir-lhe um significado para quem o está a interpretar. Desta forma, o reconhecimento passa a ser o passo seguinte. Esta classificação vai incidir num grupo de gestos que têm como acções a orientação, chamar a atenção ou cumprimento. Gestos que se inserem no paradigma deste trabalho, que consiste no reconhecimento de gestos naturais e intuitivos, e que não necessitem de um conhecimento prévio para prestarem uma interacção. Nas Fig. 3.1 estão ilustrados alguns dos gestos que podem ser reconhecidos.

Também no Anexo B.1 estão ilustrados mais exemplos de gestos. Neste capítulo irá ser mostrado um método para o reconhecimento dos gestos periódicos, descrevendo-se a experiência feita como prova do mesmo.



(a)



(b)



(c)



(d)

**Figura 3.1** – Alguns gestos periódicos. (a) - Gestos que dá a indicação de movimento para cima; (b) - Gestos que dá a indicação de movimento para a direita; (c) - Gestos circulares; (d) - Gestos de acenar.

### 3.1 Apresentação de um método proposto para o reconhecimento de gestos periódicos

O método pelo qual foi seguido o reconhecimento dos gestos tem como primeiro passo determinar qual a orientação principal do gesto. Esta, vai indicar a direcção do gesto com que foi executado, ou seja, determinar qual o eixo de coordenadas onde o gesto se incidiu relativo ao referencial do Xbox Kinect, se no eixo  $X$ ,  $Y$  ou  $Z$ . Desta forma, o processamento do sentido do gesto, que passa a ser o passo seguinte, fica limitado a dois sentidos diferentes por direcção. Isto é, quando estamos na presença de um gesto com incidência no eixo de coordenadas  $X$ , é verificado se o sentido é para a esquerda e direita. Caso o eixo de coordenadas predominante for  $Y$ , os sentidos passam a ser cima e baixo. Por último, se a incidência for no eixo de coordenadas  $Z$ , temos os sentidos frente e trás. Desta forma, foi utilizado o método Análise da Componente Principal (ACP) segundo a Decomposição em Valores Singulares (DVS), de forma a determinar a direcção do gesto, e foi feita uma análise na velocidade de execução do gesto para processar o sentido.

A ACP é uma forma de identificar padrões de dados, reduzindo assim, a dimensão dos mesmos de forma a preservar o máximo de informação contida neles [33]. Este, utiliza transformações lineares que transformam os dados para um novo sistema de coordenadas de tal forma que a direcção com maior variância situa-se no primeiro eixo de coordenadas (chamada de primeira componente principal), a segunda maior variância sobre o segundo eixo de coordenadas. Desta forma, o que se pretende é reduzir a dimensão de coordenadas a ser avaliadas na qual o gesto está inserido. Podemos então definir a matriz de dados,  $[A]$ , a analisar da seguinte forma:

$$[A]_{m \times n} = \begin{bmatrix} [X]_{m \times 1} & [Y]_{m \times 1} & [Z]_{m \times 1} \end{bmatrix}, \quad (3.1)$$

onde  $n$  indica a linha da matriz onde estão as posições referentes aos eixos  $X$ ,  $Y$  e  $Z$  na qual a trajectória do gesto foi efectuada e  $m$  o número de amostras a analisar em cada eixo.

Assim, a matriz  $[A]$  definida em (3.1) é factorizada da seguinte forma:

$$[A] = [U] \cdot [\Sigma] \cdot [V]^T, \quad (3.2)$$

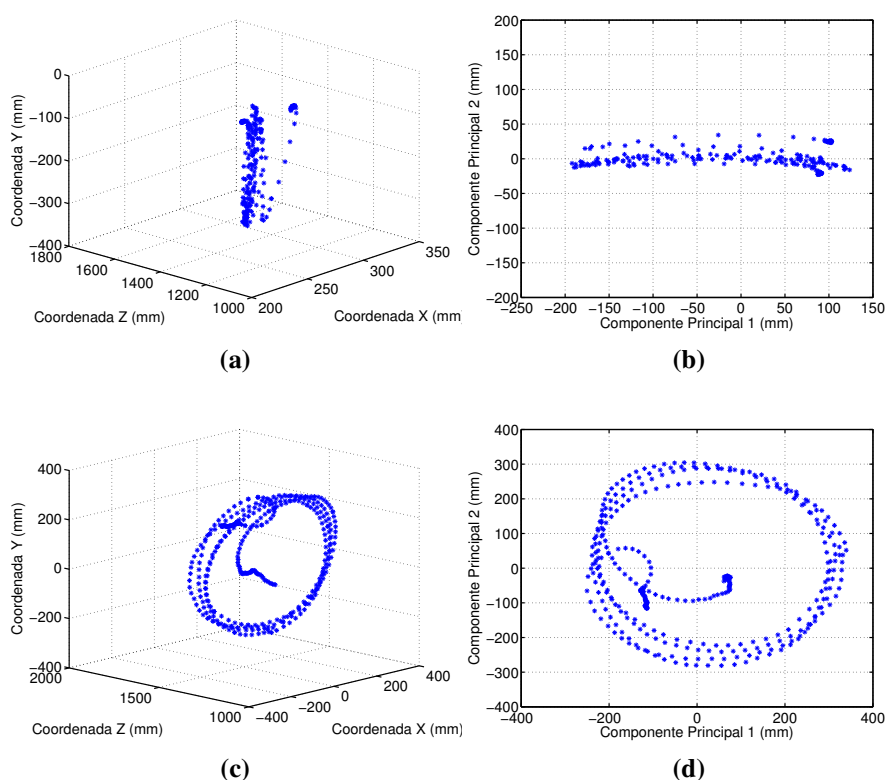
chamada de decomposição em valores singulares (DVS). Com esta, é possível obter o conjunto de vectores próprios ortogonais,  $[V]_{3 \times 3}^T$ , ( $\mathbf{v}_1$ ,  $\mathbf{v}_2$ ,  $\mathbf{v}_3$ ) com os valores próprios associados  $[\Sigma] = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ . O vector próprio com o maior valor próprio é definido

como a componente principal dos dados. A DVS, uma vez encontrados os vectores próprios ordena-os pela ordem dos valores próprios, de forma decrescente. Isto dá-nos as componentes por ordem de importância.

Agora, é importante efectuar uma redução da dimensão do sistema de coordenadas, ou seja, fazer uma transformação de um sistema  $\mathbb{R}^3$  para  $\mathbb{R}^2$ , isto porque ao reduzir a dimensão dos dados estamos a facilitar a análise dos gestos. Desta forma, aplicando a seguinte expressão aos dados originais,  $[A]$ ,

$$[DadosTransformados] = [\mathbf{v}_1 \ \mathbf{v}_2]^T \cdot [A]^T, \quad (3.3)$$

vamos obter uma matriz de duas dimensões,  $[DadosTransformados]_{2 \times m}$ , segundo os vectores que definem as duas primeiras componentes principais,  $\mathbf{v}_1$  e  $\mathbf{v}_2$ . É possível visualizar esta transformação na Fig. 3.2, onde se verifica a mudança do sistema de eixo de coordenadas.



**Figura 3.2** – Representação da transformação linear  $\mathbb{R}^3$  para  $\mathbb{R}^2$ . (a) - Representação de um gesto para a direita num sistema de coordenadas cartesiano; (b) - Representação do gesto em (a) num sistema de coordenadas em componentes principais; (c) - Representação de um gesto circular num sistema de coordenadas cartesiano; (d) - Representação do gesto em (c) num sistema de coordenadas em componentes principais;

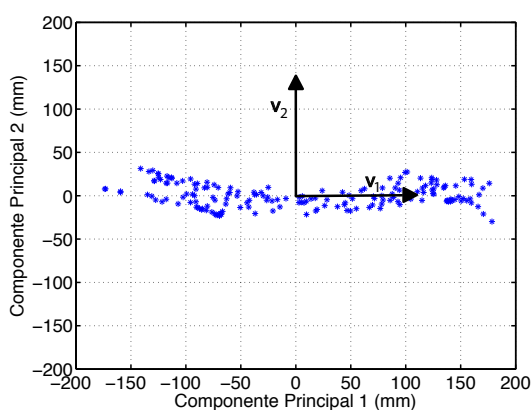
Pela amplitude dos valores próprios é possível definir a dimensão do comportamento do gesto. Estas, no conjunto de gestos que se vai analisar, comportam-se de duas maneiras

diferentes:

1.  $\lambda_1 \gg \lambda_2$ , indica que estamos perante um gesto apenas numa dimensão;
2.  $\lambda_1 \approx \lambda_2$ , indica que existe um gesto com duas dimensões.

É assim possível, logo à partida, diferenciar dois tipos de gestos, os que incidem num só eixo de coordenada ( $X$ ,  $Y$  ou  $Z$ ), que são os do nosso interesse, como também gestos dispostos sobre dois eixos de coordenadas, como por exemplo o gesto circular (ver Fig. 3.2 (d)).

No entanto, no primeiro caso, apesar do elevado valor apresentado pelo valor próprio,  $\lambda_1$ , indicar que a maior variância dos dados se apresentam no vector próprio  $\mathbf{v}_1$  (ver Fig. 3.3), não é possível de todo, determinar qual a direcção que o gesto tem no eixo de coordenadas original, sendo necessário avaliar esta para que o processo de reconhecimento prossiga.



**Figura 3.3** – Representação da distribuição e da contribuição dos vectores próprios  $\mathbf{v}_1$  e  $\mathbf{v}_2$  nos dados.

Visualmente, é possível determinar que a direcção do gesto da Fig. 3.2 (a) é em torno de  $Y$ , ou seja há uma maior variância de valores sobre este eixo, enquanto que no novo sistema de eixos (ver Fig. 3.2 (b)) é impossível determinar tal situação. Assim, ao retornar os dados de dimensões reduzidas ao eixo de coordenadas original utilizando os vectores próprios, é possível determinar a direcção que estes vão incutir neste mesmo referencial, e consequentemente determinar a direcção do gesto. Como estamos perante pontos em coordenadas cartesianas no espaço ( $X, Y, Z$ ), as componentes do vector próprio também o são. Desta forma, pode-se escrevê-lo da seguinte forma:

$$\mathbf{v} = v_x \mathbf{i} + v_y \mathbf{j} + v_z \mathbf{k}, \quad (3.4)$$

onde  $\mathbf{i}$ ,  $\mathbf{j}$  e  $\mathbf{k}$  são os versores unitários e  $v_x$ ,  $v_y$  e  $v_z$  são as componentes escalares do vector próprio,  $\mathbf{v}$ . No caso de atribuir valores às componentes escalares do mesmo,  $v_x \approx 0$ ,

$v_y \approx 1$  e  $v_z \approx 0$ , ao fazer uma redução de dimensão de dados,  $\mathbb{R}^3 \rightarrow \mathbb{R}^1$ , podemos através do produto escalar seguinte:

$$[DadosTrasnformados] = \mathbf{v}^T \cdot [DadosOriginais]_{3 \times m} = \begin{bmatrix} \approx 0 & \approx 1 & \approx 0 \end{bmatrix} \cdot \begin{bmatrix} [X]_{1 \times m} \\ [Y]_{1 \times m} \\ [Z]_{1 \times m} \end{bmatrix}, \quad (3.5)$$

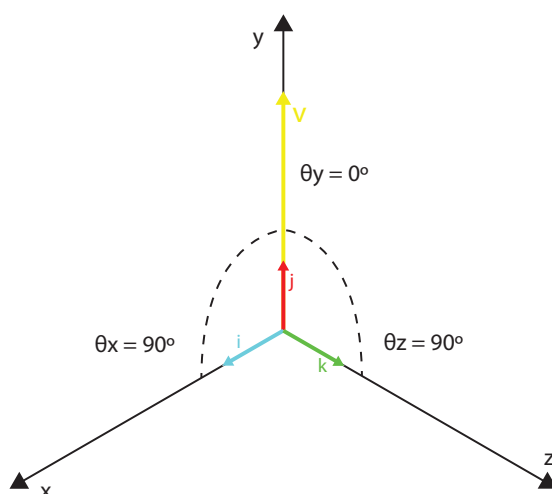
obter um vector de 1 dimensão,  $[DadosTransformados]_{m \times 1}$  segundo o vector próprio,  $\mathbf{v}$ , que define a primeira componente principal. De seguida, ao retornar aos eixos de coordenadas que correspondem ao referencial do Xbox Kinect, aplicando um produto escalar entre os  $[DadosTransformados]^T$  e o  $\mathbf{v}$ :

$$[DadosOriginais]_{3 \times m} = \mathbf{v} \cdot [DadosTransformados]^T, \quad (3.6)$$

obtemos os  $[DadosOriginais]_{3 \times m}$  segundo a direcção do vector próprio  $\mathbf{v}$ . Assim, através da análise vectorial deste, é possível determinar qual o eixo de coordenadas em que o gesto se incidiu. É importante referir que estes vectores próprios são vectores unitários, isto é, o modulo do seu comprimento é 1, ou seja, os dados que por este são projectados apenas vão ser afectados pela direcção que este vector apresenta. Desta forma, ao escolhermos a componente escalar ( $v_x, v_y, v_z$ ) com o valor mais elevado do vector próprio que caracteriza a componente principal, estamos a escolher a componente escalar onde o ângulo direccional resultante (ver equação 3.7) define a maior sobreposição da direcção com um eixo de coordenadas do Xbox Kinect, escolhendo assim a orientação primordial do gesto. No caso dos valores que atribuímos às componentes escalares do vector  $\mathbf{v}$ , podemos determinar os ângulos direccionais do vector da seguinte forma:

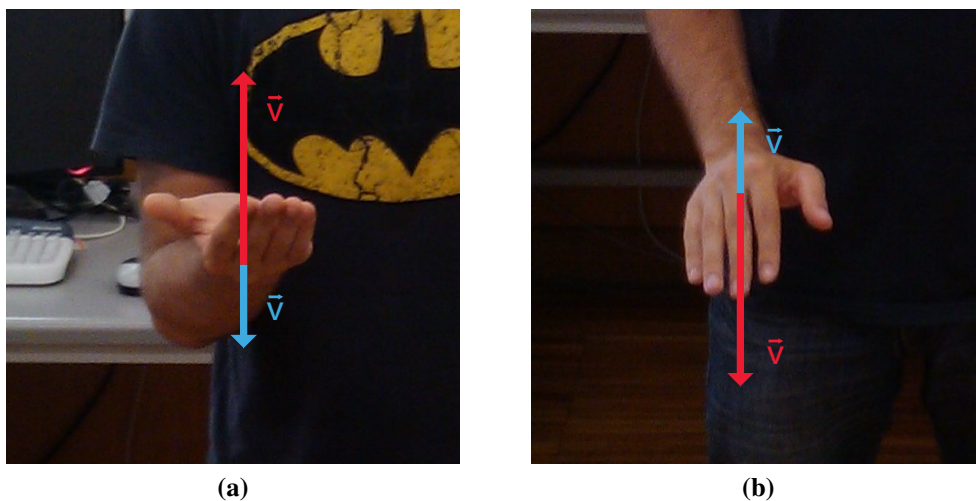
$$\theta = \cos^{-1} \frac{v}{\|\mathbf{v}\|}, \quad (3.7)$$

obtendo, assim, os ângulos  $\theta_x \approx 90^\circ$ ,  $\theta_y \approx 0^\circ$  e  $\theta_z \approx 90^\circ$ . Estes, ao serem descritos no sistema de coordenadas do Xbox Kinect (ver Fig. 3.4), verifica-se que a direcção do vector  $\mathbf{v}$  está sob o eixo de coordenadas  $Y$ , concluindo que o gesto tem como orientação este eixo.



**Figura 3.4** – Representação gráfica do vector próprio  $v$ , indicando sua direcção.

Sabendo a direcção do gesto, através da determinação do seu sentido é possível classificar o tipo de gesto, sendo este o passo seguinte. Ao executar um gesto periódico, a pessoa impõe intuitivamente uma maior velocidade no sentido que pretende indicar (ver Fig. 3.5).



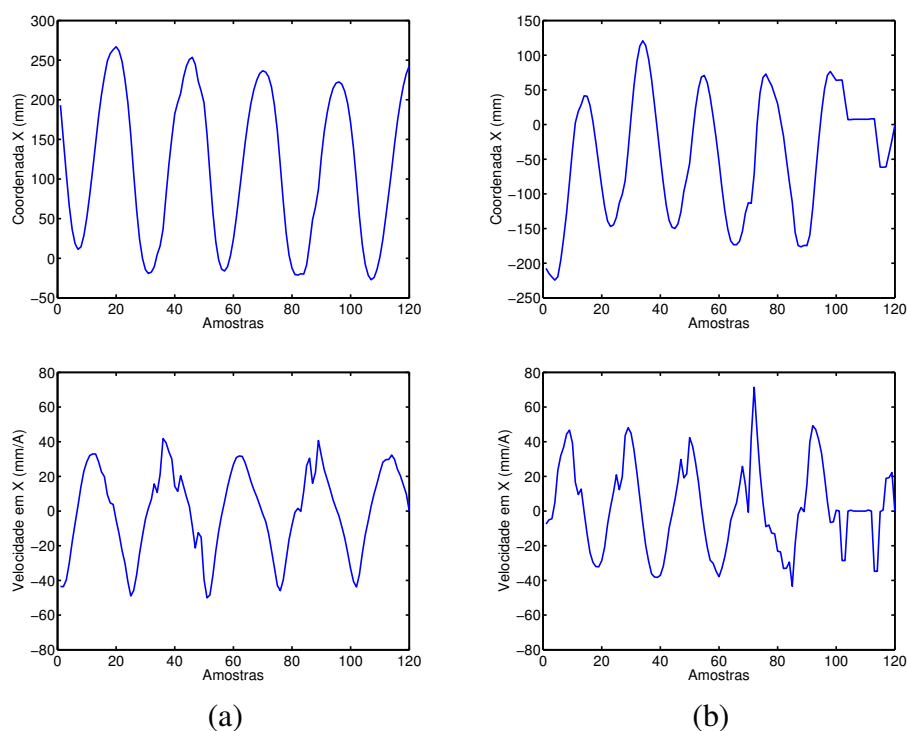
**Figura 3.5** – Gesto para cima e para baixo com a representação dos vectores velocidade, onde os vectores representados por cor azul indicam baixa velocidade e os de cor vermelha elevada velocidade. (a) - Gesto para cima; (b) - Gesto para baixo.

Tendo isto como um pressuposto característico, tomou-se partido dele e, aplicou-se o cálculo da derivada,

$$v(t) = \frac{dx}{dt}, \quad (3.8)$$

ao deslocamento da coordenada principal determinada da trajectória do gesto, e analisou-se a velocidade resultante.

De forma a demonstrar o método foram obtidos os resultados ilustrados na Fig. 3.6 (a) e (b), onde estão representados um gesto horizontal com orientação no eixo de coordenadas  $X$  para esquerda e para a direita, respectivamente. É visível que no caso da Fig. 3.6 (a) o resultado da derivada apresenta valores máximos superiores aos valores mínimos em absoluto, e, o caso contrário acontece para a Fig. 3.6 (b). O que indica que existe uma maior velocidade quando estamos a deslocar a mão para o sentido que é pretendido. Posteriormente dá-se o abrandamento para o sentido inverso. Podendo desta maneira diferenciar um gesto com a mesma direcção mas com sentidos diferentes.



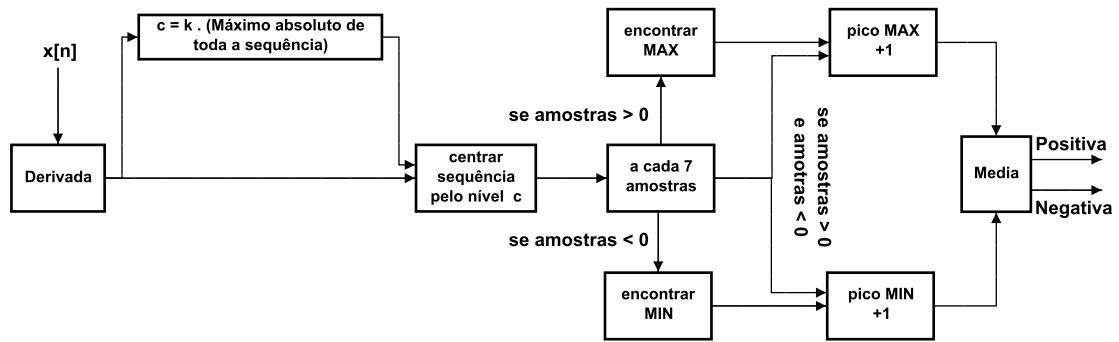
**Figura 3.6** – Resultado obtido após aplicada a derivada aos dados que representam a trajectória do gesto. (a) - Resultado da derivada aplicado ao gesto para a esquerda; (b) - Resultado da derivada aplicado ao gesto para a direita.

Desta forma, passamos a contabilizar quantos máximos e mínimos existem na sequência de dados consequentes da derivada, calculando-se, posteriormente, a média dos mesmos. Caso a média resulte num valor positivo é considerado um sentido, caso seja negativa temos o sentido contrário.

No entanto, existe uma excepção, estamos a falar do gesto de acenar, que quando é executado de forma natural, o que se vai verificar é que a velocidade dos dois sentidos é praticamente igual. Neste caso, o resultado do processamento dará uma media perto de 0,



classificando-o desta forma. A Fig. 3.7 ilustra sucintamente como é feita a contabilização dos máximos e mínimos da sequência.

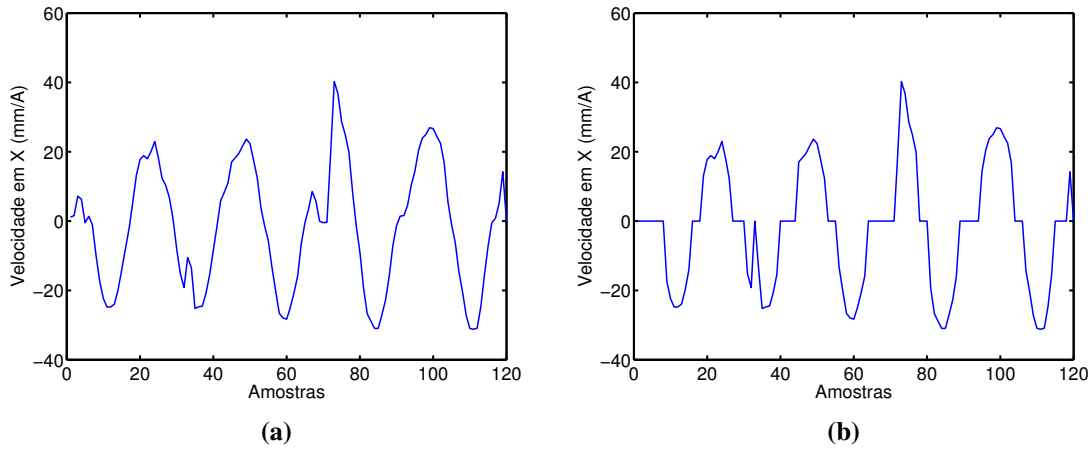


**Figura 3.7** – Esquema que representa o algoritmo efectuado para a contagem dos picos negativos e positivos da derivada.

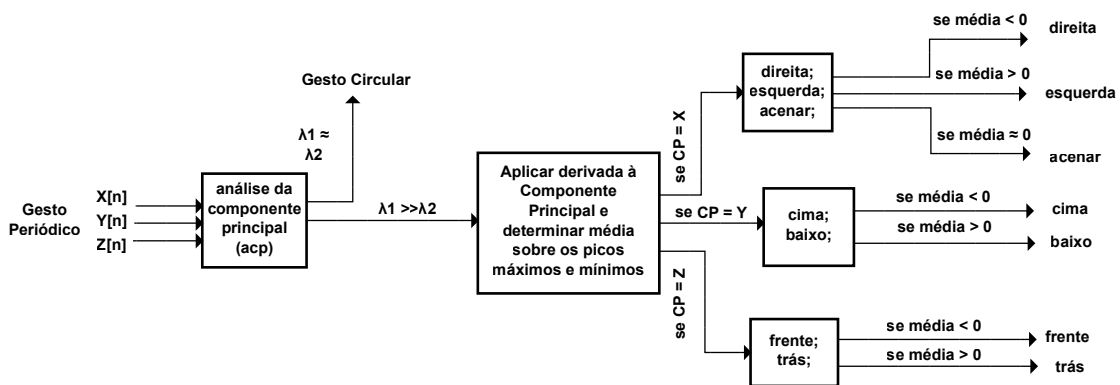
Este processamento é aplicado a uma sequência periódica de 120 amostras,  $x[n]$ , proveniente da determinação da componente principal. Como primeiro passo, esta mesma sequência é recortada pelo nível  $c$ , com objectivo de eliminar o ruído causado por deformações na trajectória a analisar, este torna o método menos propício a falhas. Para o seu cálculo é determinado o máximo absoluto de toda a sequência, e este utilizado da seguinte forma:  $c = k \times \max(|x[n]|)$ , onde  $k = 0.3$ . Usando este nível de corte, uma sequência de valores centrados em 0 é gerada a partir do sinal de entrada como se segue (ver Fig. 3.8):

$$X[n] = \begin{cases} x[n] & , \text{ se } x[n] \geq c \\ x[n] & , \text{ se } x[n] \leq -c, \\ 0 & , \text{ outros casos.} \end{cases} \quad n = 1, \dots, 120 \quad (3.9)$$

De seguida, é criada uma janela deslizante de 7 amostras e a cada 7 amostras é procurado o seu máximo, caso estejamos numa janela onde as amostras são apenas positivas, ou é encontrado o mínimo caso estejamos numa janela onde existem apenas amostras negativas. Um mínimo ou máximo são, de facto, encontrados quando existe uma transição entre os valores das amostras, ou seja, quando passamos de uma sequência de 7 amostras apenas positivas ou negativas para uma onde se encontram valores tanto negativos como positivos. No final é feita a média destes mesmos máximos e mínimos encontrados. É importante referir que esta média é feita com pares de máximos e mínimos, ou seja, apenas é feita a média com o mesmo número de máximos e mínimos. Através deste método desenvolvido, agora é possível determinar qual a direcção do gesto e o seu sentido. Todo o processamento é sucintamente ilustrado na Fig. 3.9.



**Figura 3.8** – Representação do passo de recorte do algoritmo de contagem de picos negativos e positivos da derivada. (a) - Representação gráfica da velocidade antes do recorte; (b) - Representação gráfica da velocidade depois do recorte.



**Figura 3.9** – Esquema que ilustra todo o processo de reconhecimento de gestos periódicos.

## 3.2 Confirmação do método proposto recorrendo a uma experiência.

Todo o processamento de reconhecimento de gestos referido acima foi feito com base no pressuposto que, ao se executar os gestos com sentidos diferentes, a velocidade se iria alterar de acordo com o mesmo, e desta forma, é de esperar que seja possível identificar a semântica associada a cada gesto. Para testar a robustez do método implementado, foi então preparada uma experiência para o testar, de maneira a justificar a implementação da mesma.

Esta experiência teve como propósito avaliar a intenção humana quando esta era submetida a um conjunto de acções. Acções essas que eram reflectidas sob a forma de gestos com intuito indicativo. Desta forma, foi montado um cenário (ver Fig. 3.10 ) de modo a que um sujeito iria dar indicações a outro, com objectivo de o orientar num caminho, sendo que este último não teria qualquer conhecimento da trajectória a percorrer.

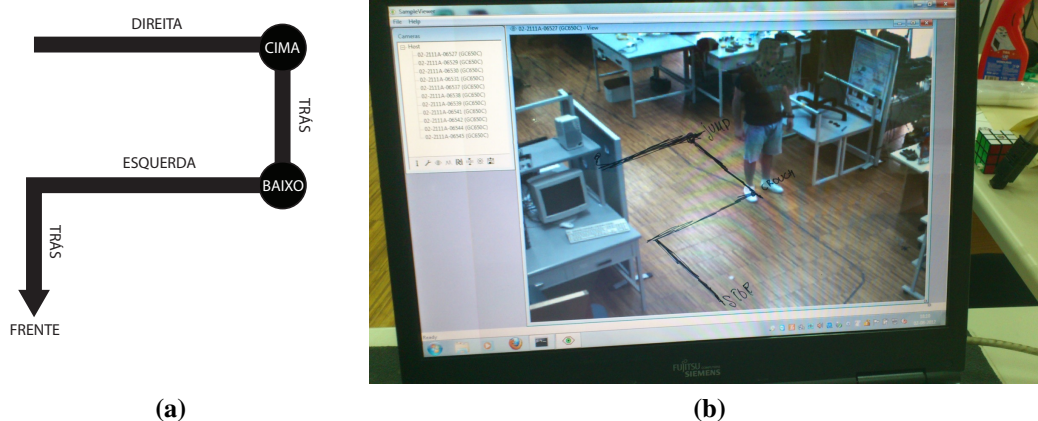


(a)

**Figura 3.10** – Esquema montado de maneira a realizar a experiência. (1) - Sujeito que indica direcções; (2) - Sujeito que recebe as direcções e percorre caminho; (3) - Ilustração do mapa; (4) - Sensor Xbox Kinect; (5) - Sensor Optotrak.

Para a recolha de dados foram usadas duas tecnologias diferentes, o Xbox Kinect e o Optotrak. Este último recorre a marcadores para extrair os dados da trajectória dos gestos e, o primeiro, ao seguimento das juntas das mãos através da biblioteca utilizada [2]. É importante referir que foram utilizadas duas frequências de amostragem diferentes, 30 Hz e 60 Hz, respectivamente, com a expectativa de obter melhores resultados no segundo caso.

Como é possível ver pela Fig. 3.11 foi traçado um caminho de modo a analisar os seguintes gestos: movimento para a direita, esquerda, cima, baixo, frente e trás. Todos eles executados em troços específicos do trajecto.



**Figura 3.11** – Ilustração do caminho efectuado para testar experiência. (a) - Ilustração do caminho; (b) - Forma como o caminho era mostrado.

Depois da recolha dos dados de 11 sujeitos, foi feita a análise dos mesmos de modo a testar a robustez do método implementado. Os resultados obtidos são apresentados sobre forma de uma tabela (Tab. 3.1) onde estão representados os 6 gestos reconhecidos correspondentes às três direcções possíveis ( $X$ ,  $Y$  e  $Z$ ) executados pelos 11 sujeitos e a indicação se o método funcionou ou não.

Sujeito	Coordenada X		Coordenada Y		Coordenada Z	
	Esquerda	Direita	Baixo	Cima	Frente	Trás
1	Sim	Sim	Sim	Não	Sim	Não
2	Sim	Sim	Sim	Não	Sim	Sim
3	Não	Não	Não	Não	Não	Sim
4	Sim	Sim	Não	Não	Não	Sim
5	Não	Não	Sim	Sim	Sim	Sim
6	Sim	Não	Sim	Sim	Sim	Sim
7	Sim	Não	Sim	Sim	Sim	Sim
8	Sim	Não	Sim	Sim	Não	Sim
9	Sim	Sim	Não	Sim	Sim	Não
10	Sim	Não	Sim	Sim	Sim	Sim
11	Sim	Sim	Sim	Sim	Sim	Não

**Tabela 3.1** – Resultados obtidos pela experiência através dos dados do Xbox Kinect, sendo "Sim" a indicação que o método funcionou e "Não" caso contrário.

Verifica-se que o método teve uma taxa de sucesso 68,18% dos gestos executados. Este facto significa que quando existe a intenção humana de executar o gesto, este vai

evidenciar uma velocidade maior para o sentido que queremos, demonstrando que a implementação apresenta resultados válidos. No entanto, existe 31,82% dos quais o método não verificou o seu processamento. A razão para este número está relacionado com os dados estarem corrompidos com ruído e as amostras que representam o gesto não serem as melhores para o seu processamento. O tempo de execução gestual também é outra razão para o número de insucesso apresentado, isto é, apesar do sujeito que recebeu a indicação ter-se apercebido da ordem, o tempo de execução gestual não foi suficiente para a detecção deste como periódico pelo algoritmo implementado.

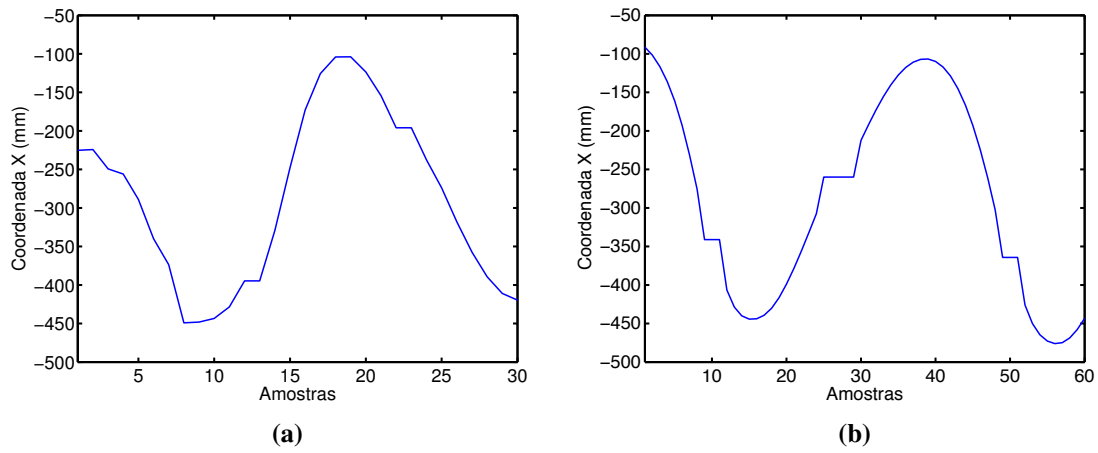
Na Tab. 3.2 seguinte são comparados se os mesmos gestos usando Optotrak, este traria melhores resultados ao termos aproximadamente o dobro das amostras. Não foi possível testar todos os sujeitos, pois apenas em 4 é que foram adquiridos os dados através do Optotrak.

Sujeito	Coordenada X		Coordenada Y		Coordenada Z	
	Esquerda	Direita	Baixo	Cima	Frente	Trás
	Kinect					
7	Sim	Não	Sim	Sim	Sim	Sim
8	Não	Não	Sim	Sim	Não	Sim
10	Sim	Não	Sim	Sim	Sim	Sim
11	Sim	Sim	Sim	Sim	Sim	Não
	Optotrak					
7	Sim	Sim	Sim	Sim	Sim	Sim
8	Sim	Sim	Sim	Sim	Sim	Sim
10	Sim	Sim	Sim	Sim	Não	Sim
11	Não	Sim	Sim	Sim	Sim	Sim

**Tabela 3.2** – Comparação entre os resultados óbtidos pelo Optotrak e pelo Xbox Kinect. Sim - indica que o método funcionou; Não - indica que o método não funcionou.

No entanto, é possível ver que existem casos onde, com os dados adquiridos por este último, o método funcionou, enquanto com o Xbox Kinect não. A explicação para este fenómeno deve-se ao facto de obtermos mais amostras para a mesma representação da trajectória no mesmo espaço de tempo, o que influencia directamente o cálculo da derivada. Desta forma, o calculo da derivada vai apresentar resultados de melhor qualidade, ou seja, com níveis de ruído reduzidos. Na Fig. 3.12 está ilustrado a comparação entre dados obtidos, na mesma sequência de amostras do gesto, pelo Xbox Kinect com uma frequência de amostras de 30 Hz e os dados obtidos pelo Optotrak com uma frequência de amostras de 60 Hz.

Em suma, os resultados obtidos com esta experiência são suficientes para perceber que o método implementado vai ter bons resultados se o gesto for executado de forma



**Figura 3.12** – Trajetórias obtidas pelo Xbox Kinect e pelo Optotrak. (a) - Trajetória obtida pelo Xbox Kinect; (b) - Trajetória obtida pelo Optotrak.

natural e intuitiva, e se não houver falha por parte do sistema de seguimento utilizado, havendo pouca margem para falhas.

# Capítulo 4

## Gestos demonstrativos

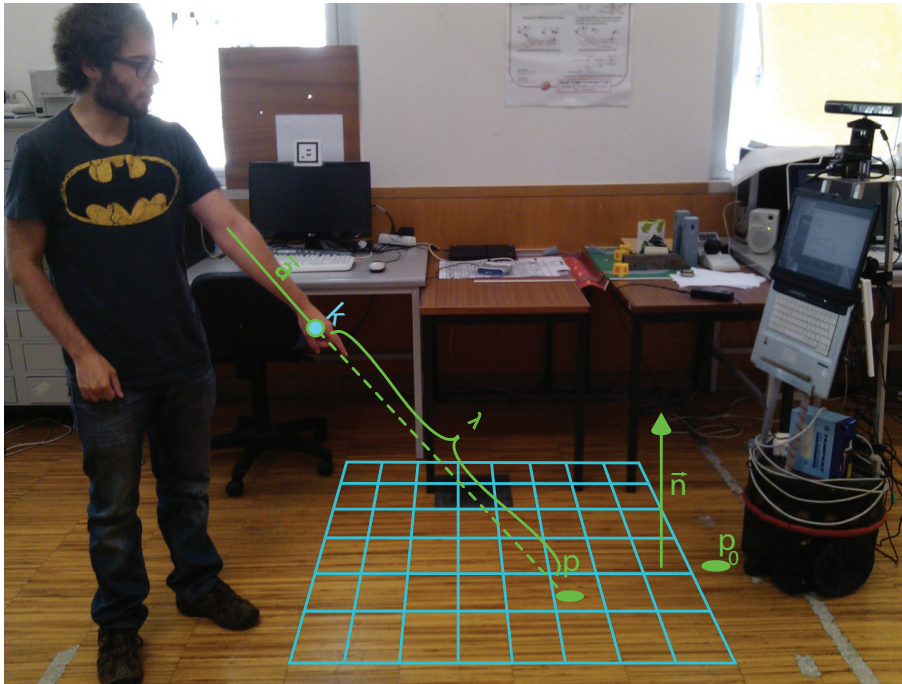
Numa interacção intuitiva e natural, os gestos deícticos ou demonstrativos têm bastante relevância, uma vez que são utilizados de uma maneira não forçada quando queremos indicar a posição de um determinado objecto ou lugar. Como Jovic [34] refere, de entre os gestos naturais humanos, o de apontar é o que é mais fácil de se interpretar e neste trabalho é usado um mecanismo para recriar esse gesto. Mesmo para para utilizadores tecnicamente inexperientes, o gesto de apontar é especialmente útil uma vez que esta é uma operação utilizada diariamente e não necessita de um conhecimento inicial sobre a sua execução.

Neste capítulo é apresentado um método que recria o gesto de apontar, e também apresentado um método para suprimir o ruído que existe nos dados utilizados, de forma a tornar o sistema mais robusto.

### 4.1 Gesto apontar através da intersecção com plano do chão

A abordagem aqui utilizada não pretende reconhecer o gesto de apontar propriamente dito mas, usando primeiro um gesto de atenção (acenar, por exemplo) para inicializar a interacção, activando assim o método, podemos de seguida apontar para um determinado lugar ou objecto. Na Fig. 4.1 é mostrado o esquema no qual foi baseado o processamento do gesto de apontar.

Através da interface de programação de aplicações OpenNI [2, 35], é possível, com o procedimento de seguimento, obter as posições das juntas, cotovelo e mão. Este par de pontos 3D define o vector de apontar  $\mathbf{a}$ .



**Figura 4.1** – Ilustração do método implementado para o gesto apontar.

Para determinar a posição que o utilizador pretende que o robô assuma, faz-se uso das informações do plano do chão dado pelo OpenNI. O plano do chão é então definido por um vector normal  $\mathbf{n}$  e um ponto 3D  $P_0$  do chão, como:

$$\mathbf{n} \cdot (P - P_0) = 0, \quad (4.1)$$

sabendo que  $P$  é um ponto genérico do plano.

Com o vector  $\mathbf{a}$  e o plano definidos em (4.1), podemos saber a intersecção do ponto, que deve corresponder à posição indicada pelo utilizador. Usando a equação vectorial de uma linha e a equação do plano (4.1),

$$\begin{cases} \mathbf{n} \cdot (P - P_0) = 0 \\ P = K + \lambda \cdot \mathbf{a}, \end{cases} \quad (4.2)$$

definindo  $K$  como sendo um ponto da linha, com a mesma direcção que  $\mathbf{a}$ ,  $\lambda$  relaciona  $\|\mathbf{a}\|$  e a distancia entre  $K$  e o ponto no plano,  $P$ . Resolvendo o sistema de equações (4.2) podemos definir o  $\lambda$  como,

$$\lambda = \frac{(P_0 - K) \cdot \mathbf{n}}{\mathbf{a} \cdot \mathbf{n}}, \quad (4.3)$$

e desta forma determinar as coordenadas 3D ( $X, Y$  e  $Z$ ) do ponto em relação a origem do referencial do Xbox Kinect.



Contudo, o sistema não pode estar dependente apenas de um ponto de intersecção, pois não garante a posição exacta que queremos indicar, foi necessário arranjar um mecanismo de forma a confirmar a posição final. Desta forma, é determinada a média de 60 de amostras contínuas, obtendo assim o ponto mais próximo para o qual queremos indicar. Ao utilizar este número de amostras o método tem um tempo de confirmação da posição de 2 segundos, conseqüente da frequência de amostragem de 30 Hz.

## 4.2 Método para redução de ruído

É sabido que todos estes sistemas de seguimento geram ruídos sucessivos que se reflectem nos resultados finais. Existem várias razões que afectam a viabilidade dos dados que podem ser, a iluminação do ambiente; tamanho corporal de uma pessoa; a distância a que a pessoa está do sensor; a pose da pessoa (por exemplo, para os dados da mão, quando ela está aberta e fechada), efeitos de arredondamentos introduzidos pelo processamento, entre outros. Um importante passo antes de utilizar os pontos 3D do seguimento do esqueleto humano, é a redução do ruído o mais possível para obter um seguimento mais robusto e com resultados finais de maior precisão. Já que estamos a falar de um gesto de apontar é ainda mais fulcral reduzir estes erros, dado que necessitamos que exista uma grande precisão nos resultados.

Desta forma, foi implementado um algoritmo baseado no método *Double Exponential Smoothing-Based Prediction* (DESP) [36], este vai suavizar as pequenas variações que estão sempre inerentes aos pontos retornados pelo seguimento (ver algoritmo 1).

---

**Algorithm 1** filtragem de coordenadas

---

```
1: Input: currentPosition
2: Output: newPosition
3: newFilterOutput = ( $\alpha \times currentPosition$ ) + (1 -  $\alpha$ )  $\times$  (lastFilterOutput + lastTrend)
4: newTrend =  $\gamma \times$  (newFilterOutput - lastFilterOutput) + (1 -  $\gamma$ )  $\times$  lastTrend
5: newPosition = newFilterOutput + newTrend
6: lastFilterOutput = newFilterOutput
7: lastTrend = newTrend
8: return newPosition
```

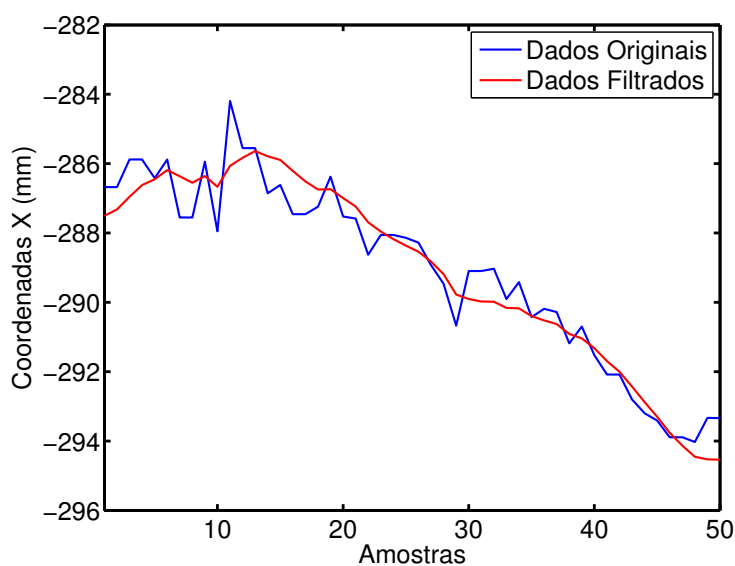
---

Este filtro prevê a tendência que os dados apresentam tendo em conta as posições anteriores. Uma soma entre a tendência anterior (*lastTrend*) e a saída do filtro anterior (*lastFilterOutput*) é determinada, e é usada para determinar a próxima saída do filtro (*newFilterOutput*). Assim, somando a nova saída do filtro (*newFilterOutput*) e a nova tendência (*newTrend*) é obtida a nova posição (*newPosition*) a utilizar pelo sistema. A inclusão do cálculo da tendência ajuda a reduzir o atraso que o filtro adiciona ao resultado.

O  $\gamma$  controla o peso dos dados utilizados para o cálculo da mesma, isto é, controla o quanto sensível é a tendência a mudanças recentes dos mesmos. Um elevado  $\gamma$  resulta numa baixa latência na tendência, isto é, a posição de saída, a mudanças recentes na entrada, segue uma variação mais rápida, enquanto que um valor baixo de  $\gamma$  a posição de saída segue uma variação mais lenta. É importante referir que a tendência determinada é a ligeira diferença entre as duas últimas saídas do filtro (*lastFilterOutput* e *newFilterOutput*), que correspondem à última posição utilizada e à nova posição a ser utilizada, respectivamente, assim, pode-se dizer que a tendência é a velocidade estimada de uma junta.

Um filtro ideal deveria remover as pequenas variações dos dados obtendo uma experiência de utilização onde não houvesse qualquer tipo de atrasos na resposta ao sistema. No entanto, o facto de neste processamento se dar preferência à suavização dos dados, pois apenas se vai utilizar esta filtragem quando queremos adquirir a posição na qual estamos a apontar, ou seja, o nosso membro vai ficar estático. Isto irá gerar um atraso devido à filtragem, aumentando assim a latência. Esta pode ser definida como o tempo que leva a partir de quando uma pessoa faz um movimento, até ao momento em que a pessoa vê a resposta do seu movimento do corpo no ecrã, o que é indesejável para a utilização normal do sistema. Mas como já foi referido é um problema que vai ser contornado, pois apenas se quer suavizar as pequenas variações presentes nos dados e não utilizar este método como preditor de posições quando estas são obstruídas, onde neste caso o aumento de latência traria problemas de fluidez do sistema.

Na figura seguinte estão ilustrados os resultados depois de submeter os dados originais ao filtro, para os obter são aplicados valores de  $\alpha$  e  $\gamma$  iguais a 0.2.



**Figura 4.2** – Comparação entre os dados originais e os filtrados após a intervenção do algoritmo.

Pela análise visual da figura é possível concluir uma notória redução do ruído inerente aos dados do seguimento.

### 4.3 Resultados

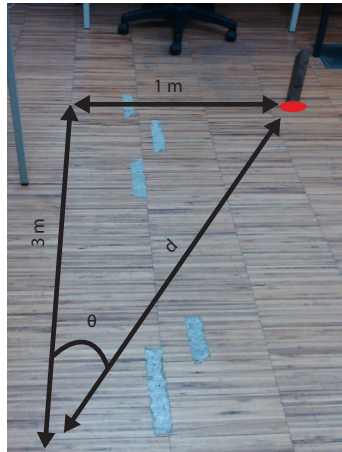
O método implementado para a utilização do gesto apontar, como ilustrado na Fig. 4.3, é capaz de indicar uma posição para onde um robô se deve deslocar, que corresponde a identificar as posições num ponto no chão.



**Figura 4.3** – Resultado do gesto apontar. O ponto branco representa a localização para onde o gesto está direccionado.

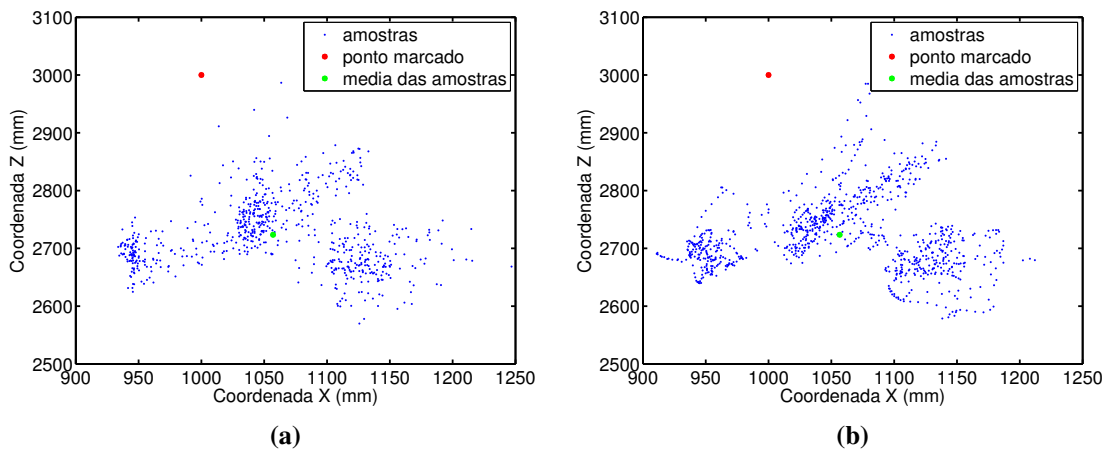
Foi também possível medir a precisão do sistema, comparando a distância entre um ponto conhecido e a média das amostras obtidas pelo sistema, tendo o Xbox Kinect como ponto de referência. Para isso, foi especificado um ponto que fica a 3 m de distancia e a 1 m à direita do sensor (ver Fig. 4.4), e foram retiradas 900 amostras de forma a mostrar a robustez do método.

Nas Fig. 4.5 (a) e (b) estão ilustrados as posições adquiridas das 900 amostras retiradas directamente do sistema sem filtragem e com filtragem, respectivamente, a representação da média das mesmas e o ponto marcado. Como se pode verificar, ao aplicar o método de filtragem não são notórias melhorias na exactidão dos pontos adquiridos, sendo o erro para ambos os casos  $\approx 9\%$  para a coordenada  $Z$  e  $\approx 2\%$  para a coordenada  $X$ . Este erro poderá estar relacionado pelo facto de o vector utilizado para intersectar o chão não ser o utilizado por nós para apontar, isto é, a não utilização da direcção do olhar pode estar a afectar os resultados quanto à exactidão. No entanto, com os resultados apresentados é possível indicar uma área bastante próxima do ponto que queremos apontar. Apesar de



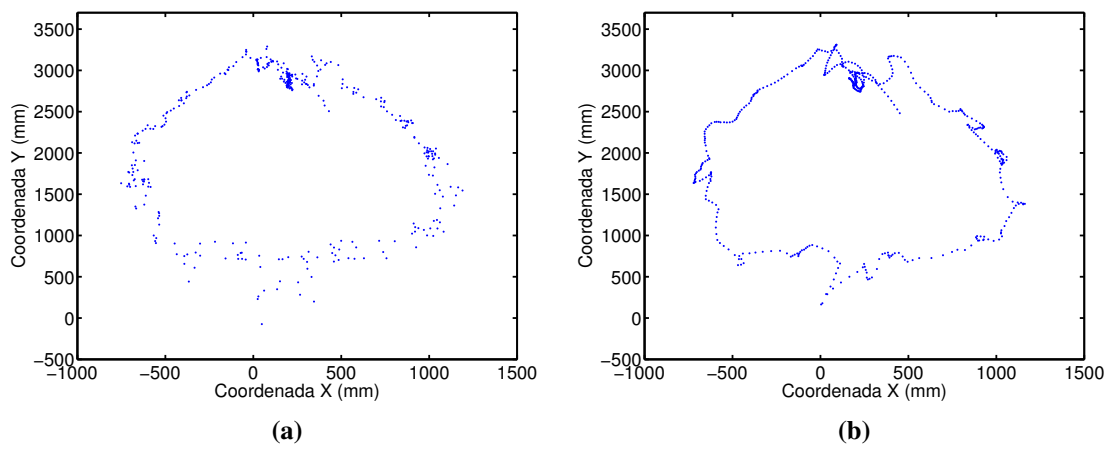
**Figura 4.4** – Ilustração do ponto marcado para testar a precisão do sistema.

não se obter um ponto preciso de localização, se o objetivo for deslocar o robô para essa zona os resultados são válidos e satisfatórios.



**Figura 4.5** – Resultados obtidos para a medição da precisão. Comparação do entre ponto médio de 30 amostras e o ponto marcado a 3 m de profundidade e a 1 m para a direita do sensor, tendo este como referencia.

Apesar do método de redução de ruído não apresentar melhorias quanto à precisão, foi possível com os resultados apresentados nas Fig. 4.6 demonstrar que a aplicação da filtragem pode melhorar a precisão dos dados. Desta forma o que se fez foi traçar uma forma circular no chão e, assim, comparar as trajetórias obtidas com e sem filtragem. Os resultados demonstram que o método de filtragem vai melhor, consideravelmente, a trajetória do círculo. Assim, para além de ser possível apontar uma posição para o robô se deslocar, é também possível traçar uma trajetória para este a percorrer.



**Figura 4.6** – Resultados obtidos depois de traçada a forma circular. (a) - Trajectória sem recorrer à filtra-  
gem; (b) - Trajectória com filtragem.



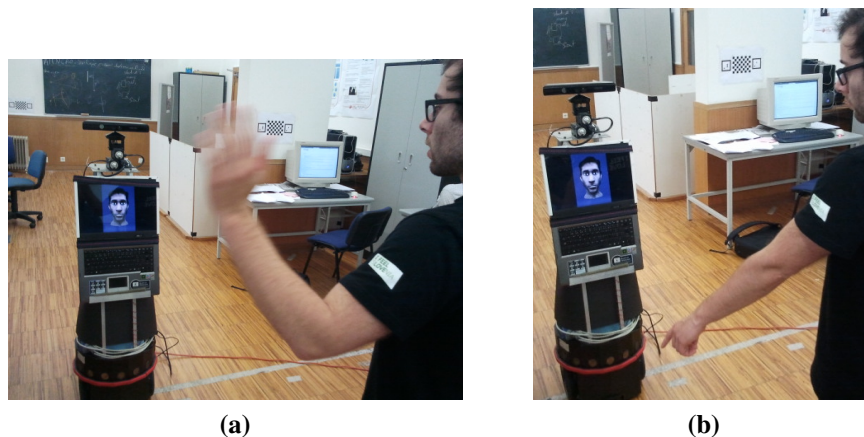
# Capítulo 5

## Conclusão

Pretende-se, com o trabalho desenvolvido nesta dissertação, utilizar comportamentos gestuais com os quais estamos familiarizados, de forma a efectuar interacção com a Máquina. Estes, não são mais que os usados natural e intuitivamente na rotina diária por nós, para demonstrar posições de objectos ou pessoas, saudar, indicar orientações, gestos que não necessitam de aprendizagem prévia para os usar. É espontânea a forma com que se demonstra a posição de um objecto em vez de descreve-lo verbalmente. Por exemplo, quando um arrumador de carros indica as orientações a efectuar, os gestos que ele executa são aqueles que surgem intuitivamente.

Desta forma, o que se propôs nesta tese de mestrado foi o desenvolvimento de mecanismos capazes de reconhecer gestos periódicos e déicticos ou demonstrativos, cujos seus comportamentos vão de acordo com o que se objectivou.

Na Fig. 5.1 (a) está ilustrado o gesto acenar, que sintetiza a definição de um gesto periódico natural. Este é utilizado intuitivamente para saudar e o seu movimento comporta-se de uma forma periódica. Em relação aos gestos demonstrativos ou déicticos, como ilustra a Fig. 5.1 (b), pode-se dizer que, são também úteis, e contextualizam um pedido que se faça a um robô, indicando-lhe um objecto, uma pessoa ou um lugar, servindo maioritariamente para este tipo de acções.



**Figura 5.1** – Representação de dois gestos naturais e intuitivos. (a) - Gesto acenar; (b) - Gesto apontar.

O facto da estimação da posição das mãos se basear numa interface inovadora e de baixo custo, como a OpenNI [2], tornou possível suprimir problemas que outros sistemas de interacção apresentam, tais como métodos de seguimento computacionalmente pesados, utilização de marcadores ou qualquer outro tipo de acessório, o processamento destes ser limitado a um referencial 2D, entre outros. Para além disto, com esta interface utilizada, é possível obter uma terceira coordenada  $Z$  através de mapas de profundidade adquiridos por um sensor de baixo custo Xbox Kinect, tornando a implementação dos mecanismos mais natural.

Os resultados obtidos da detecção dos gestos periódicos mostram ser capazes de distinguir as coordenadas que conferem ao gesto a periodicidade. Foi também possível, pelos resultados, analisar que estes funcionam de forma robusta, pois quando confrontados com movimentos aleatórios sem qualquer característica periódica, não apresentaram comportamentos indesejáveis, proporcionando consistência ao sistema. No entanto, dos métodos desenvolvidos, o que teve como base de funcionamento a auto-correlação foi o que se destacou nos melhores resultados apresentados, dando preferência a este para sua utilização. É importante referir que este método está dependente do tempo de execução do gesto, tendo de existir um certo número de repetições para que este seja considerado periódico, resultando em tempos de resposta  $\approx 4$  segundos. Os resultados alcançados mostram a viabilidade da utilização deste tipo de gestos, tendo os mesmos sido publicados nas actas da conferência Ro-MAN 2012 [3].

Com o reconhecimento dos gestos periódicos, foi evidenciada a forma natural de como estes são executados, mostrando que a intensidade do sentido que queremos indicar tem um peso importante para o reconhecimento dos mesmos. Com a intensidade entende-se que existe uma maior velocidade para o sentido que intuitivamente queremos demonstrar,



o contrário verifica-se quando se dá o abrandamento dos mesmos. Desta forma, foi possível determinar a semântica associada aos gestos e reconhecer acções como: para a direita, esquerda, baixo, cima, frente e trás. É ainda possível reconhecer o gesto de acenar, sendo este um gesto de cumprimento poderá ser utilizado como uma abordagem inicial natural para a interacção.

Em relação aos gestos deícticos ou demonstrativos, foi desenvolvido um mecanismo capaz de determinar o gesto de apontar, sendo este o de mais fácil interpretação numa interacção [34]. É mais fácil e mais preciso apontar para um objecto do que descrevê-lo verbalmente ou indicar a sua posição. Contudo, a abordagem aqui utilizada não consistiu em reconhecer este gesto propriamente dito, sendo necessário um gesto inicial para introduzir o método (por exemplo, o gesto acenar, ver Fig. 5.1). Posteriormente, é feita uma análise vectorial para o recriar. Através da intersecção de entre um vector, que dá a direcção do gesto definido por 2 pontos 3D (cotovelo e mão), e o plano do chão, é determinada uma posição que indica o ponto que queremos apontar. Pela análise dos resultados foi possível concluir que o método apresenta valores válidos, sendo possível indicar com precisão uma área para o qual queremos que o robô se mova. Foi demonstrado que, usando o método de filtragem implementado, é possível reduzir o ruído proveniente dos dados do seguimento, concluindo que é possível definir trajectórias claras com exactidão.

A grande vantagem do reconhecimento deste grupo de gestos consiste na não necessidade de uma aprendizagem prévia para os utilizar, são gestos intuitivos e simples. Para além disso, os mecanismos utilizados são computacionalmente leves e não utilizam nenhuma tecnologia de custo elevado, facilitando assim a integração destes. Por oposição, nos vários métodos referenciados no Estado de Arte, houve sempre o intuito de detectar poses da mão de forma a criar linguagens gestuais para efectuar a interacção, tornando a sua utilização penosa para o operador.

Os mecanismos aqui implementados são capazes de reconhecer os gestos periódicos e deícticos ou demonstrativos em que se objectivou esta dissertação, obtendo resultados válidos e conclusivos. Conclui-se que estes gestos têm uma grande versatilidade na sua utilização, devido ao facto de que, dadas as suas características de execução naturais e intuitivas, podem ser implementados em diversos sistemas de IHM. Em carácter social, estes sistemas podem ser inseridos nos designados *care robots*, proporcionando aos utilizadores uma menor preocupação em se adaptar ao sistema de interacção. Estes gestos proporcionam o efeito de *telepresence*<sup>1</sup>, permitindo uma ideia de proximidade por parte dos participantes numa video conferência, apesar de estarem em localizações diferentes. Este efeito pode demarcar-se quando um sistema de video conferência está incorporado

---

<sup>1</sup>*telepresence*: termo que define a tecnologia que dá a sensação real de presença numa comunicação virtual

num robô e este acompanha o utilizador, podendo este indicar com um gesto demonstrativo algum objecto que queira descrever. Em termos industriais, estes gestos podem ser usados de uma maneira natural e intuitiva, e sem qualquer uso de periféricos para ordenar máquinas de grande produção standardizada, dando-lhes indicações de funcionamento.

Como se verifica a IHM tem já uma vasta utilidade em todo o tipo de comunicação, sendo possível utilizar os gestos acima referidos para simplificar este processo de interacção.

## 5.1 Trabalho futuro

O uso benéfico destes gestos em interacções simples e naturais como forma de comunicação, possibilita a sua utilização noutros tipos de aplicações. Numa vertente mais lúdica, podemos inseri-los em jogos virtuais, onde as acções inerentes a eles podem ser reflectidas em comandos dos mesmos, relegando por completo o uso de qualquer tipo de hardware para comandar acções. Este aspecto, apenas com algumas modificações, pode servir para efeitos fisioterapêuticos, usando jogos virtuais controlados gestualmente para fazer movimentar os pacientes. Ainda no campo da saúde, a forma repetida como os gestos periódicos são executados pode proporcionar ajuda a pessoas com dificuldades motoras, sendo possível criar um sistema que possa ter proveito aos mesmos. Falando no caso de pessoas que necessitam de uma cadeira de rodas no seu quotidiano, o controlo desta poderá ser feito com a ajuda da ferramenta aqui desenvolvida, abandonando o controlo dos movimentos da cadeira por via de joysticks. O gesto deíctico ou demonstrativo pode ainda ser utilizado como forma de ferramenta de apoio no ensino ou em palestras, substituindo o tradicional apontador quando é feita uma apresentação. Devido ao facto deste tipo de gestos não necessitar de uma aprendizagem prévia, a adaptação ao uso destas melhorias seria feita de maneira fácil e intuitiva.

Para possibilitar a diversificação nas aplicações dos mecanismos aqui desenvolvidos, é necessário ter em conta algumas melhorias. Tal como foi aplicada no método de apontar, a filtragem pode ser utilizada em todo o sistema. Como é usado essencialmente o seguimento do esqueleto através do OpenNI [2], é importante reduzir todos os ruídos inerentes a estes sistemas. Outro problema que estes acarretam é a perda de informação de um determinado ponto, devido ao estado inferido do seu seguimento. As causas deste fenómeno podem ser: a oclusão do sistema de seguimento devido a objectos, a outras pessoas, a outras partes de corpo da própria pessoa e a mover o ponto (junta) fora do alcance do sensor. Este tipo de acontecimentos geram picos temporários causados pela

imprecisão, podendo ser minimizados através de técnicas de boa implementação e com o uso de filtros [37].

Em relação ao gesto apontar propriamente dito, o que se poderá fazer futuramente é utilizar a mesma mecânica aqui utilizada, mas em vez do vector apontar intersectar apenas o plano do chão intersectar, também, planos do ambiente que nos rodeiam. Assim, é possível aumentar as superfícies que para as quais podemos apontar. Por exemplo, se a direcção do vector apontar, intersectar uma estrutura alta, provavelmente, este está-se a referir a um objecto que o robô deverá analisar (por exemplo uma mesa), caso contrário quererá dizer uma localização para onde o robô terá que ir. Desta forma, um método de segmentação de planos baseado na Transformada Hough 3D é sugerida aqui [38], estes trabalham essencialmente com pontos no plano  $\mathbb{R}^3$  com objectivo de suavizar o processamento envolvido na transformada de Hough.



## Referências Bibliográficas

- [1] A. Libin and E. Libin, “Robots who care:robotic psychology and robototherapy approach,” *American Association for Artificial Intelligence*, 2005.
- [2] OpenNI organization, *OpenNI User Guide*, November 2010.
- [3] J. Quintas, L. Almeida, M. Brito, G. Quintela, P. Menezes, and J. Dias, “Context-based understanding of interaction intentions,” in *RO-Man*, 2012.
- [4] Ô. Birk, T. B. Moeslund, and C. B. Madsen, “Real-time recognition of hand alphabet gestures using principal component analysis.,” (Lappeenranta, Finland), Proc. Scandinavian Conference on Image Analysis, June 1997.
- [5] Y. Fang, J. Cheng, J. Wang, K. Wang, J. Liu, and H. Lu1, “Hand posture recognition with co-training,” (Beijing), Chinese Academy of Sciences, IEEE, 2008.
- [6] L. Bretzner, I. Laptev, and T. Lindeberg, “Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering,” in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, (KTH, 100 44 Stockholm, Sweden), IEEE, 2002.
- [7] R. J, K. M. S. Mahishi, D. R, S. S, and N. V. Pujari, “Finger detection for sign language recognition,” in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, (Hong Kong), IMECS, March 2009.
- [8] S. Calinon and A. Billard, “Recognition and reproduction of gestures using a probabilistic framework combining pca, ica and hmm,” in *Proceedings of the 22nd International Conference on Machine Learning*, (Bonn, Germany), 2005.
- [9] R. J.L., S. R, . K. U., and P. P.B, “Real-time robotic hand control using hand gesture,” in *2nd international conference on Machine Learning and Computing*, (Bangalore, India,), pp. 12–16, Feb 2010.
- [10] A. Heap and F. Samaria, “Real-time hand tracking and gesture recognition using smart snakes.,” in *In Interface to Real and Virtual Worlds*, (Montpellier), 1995.
- [11] E. Sánchez-Nielsen, L. Antón-Canalís, and M. Hernández-Tejera, “Hand gesture recognition for human-machine interaction,” *WSCG*, vol. 12, no. 1-3, 2003.
- [12] A. Lanitis, T. Taylor, C. Cootes, and T. Ahmed, “Automatic interpretation of human faces and hand gestures using flexible models,” in *Proc. International Conference on Automatic Face and Gesture Recognition*, (Zurich), pp. 98–103, 1995.

- [13] A. Al-Hamadi, M. Elmezain, and B. Michaelis, “Hand gesture recognition based on combined hand gesture recognition based on combined features extraction,” *International Journal of Information and Mathematical Sciences*, vol. 6, 2010.
- [14] R. Kelley, M. Nicolescu, A. Tavakkoli, M. Nicolescu, C. King, and G. Bebis, “Understanding human intentions via hidden markov models in autonomous mobile robots,” 2008.
- [15] C. Zhu, Q. Cheng, and W. Sheng, “Human intention recognition in smart assisted living systems using a hierarchical hidden markov model,” in *Automation Science and Engineering*, 2008.
- [16] D.-I. Ko and G. Agarwal, *Gesture recognition: Enabling natural interactions with electronics*. Texas Instruments, April 2012.
- [17] J. Guo, “Hand gesture recognition and interaction with 3d stereo camera,” Master’s thesis, Department of Computer Science Australian National University, November 2011.
- [18] D.-H. Lee and K.-S. Hong, “Game interface using hand gesture recognition,” in *5th International Conference on Computer Science and Convergence Information Technology*, pp. 1092 – 1097, 2010.
- [19] X. Li, J. ho An, J. hong Min, and K.-S. Hong, “Hand gesture recognition by stereo camera using the thinning method,” in *Multimedia Technology (ICMT), 2011 International Conference*, pp. 3077– 3080, 2011.
- [20] Y. Wang, T. Yu, L. Shi, and Z. Li, “Using human body gestures as inputs for gaming via depth analysis,” in *Multimedia and Expo, 2008 IEEE International*, pp. 993–996, 2008.
- [21] M. Holte, T. Moeslund, and P. Fihl., “Fusion of range and intensity information for view invariant gesture recognition.,” in *Conference on Computer Vision and Pattern Recognition Workshops*, 2008.
- [22] A. Prusak, O. Melnychuk, H. Roth, I. Schiller, and R. Koch., “Pose estimation and map building with a time-of-flight-camera for robot navigation,” in *Dynamic 3D imaging workshop*, (Heidelberg, Germany), 2007.
- [23] Z. Li and R. Jarvis, “Real time hand gesture recognition using a range camera,” in *Australasian Conference on Robotics and Automation (ACRA)*, (Sydney, Australia), 2009.
- [24] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images,” *Microsoft Research Cambridge and Xbox Incubation*, 2011.
- [25] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, “Robust hand gesture recognition with kinect sensor,” *ACM Multimedia*, 2011.

- [26] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun, “Real time hand pose estimation using depth sensors,” *IEEE International Conference on Computer Vision Workshops*, 2011.
- [27] P. Menezes, F. Lerasle, and J. Dias, “Towards human motion capture from a camera mounted on a mobile robot,” *Image and Vision Computing*, vol. 29, no. 6, pp. 382–393, 2011.
- [28] D. Droschel, J. Stückler, and S. Behnke, “Learning to interpret pointing gestures with a time-of-flight camera,” in *Proceeding of 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (Bonn, Germany).
- [29] J. Y. Oh, C. W. Lee, and B. J. You, “Gesture recognition by attention control method for intelligent humanoid robot,” in *Knowledge-Based Intelligent Information and Engineering Systems*, vol. Part I, pp. 1139–1145, 2005.
- [30] M. Takahashi, K. Irie, K. Terabayashi, and K. Umeda, “Gesture recognition based on the detection of periodic motion,” IEEE, 2010.
- [31] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex fourier series.,” *Mathematics of Computation*, pp. 297–301, 1965.
- [32] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, “A comparative performance study of several pitch detection algorithms,” pp. 399 – 401, Revised 1976.
- [33] H. Hotelling, “Analysis of a complex of statistical variables into principal components.,” *Journal of Educational Psychology*, no. 24, pp. 417–441, 1933.
- [34] N. Jovic, B. Brumitt, B. Meyers, S. Harris, and T. Huang, “Detection and estimation of pointing gestures in dense disparity maps,” in *International Conference on Automatic Face and Gesture Recognition*, (France), 2000.
- [35] PrimeSense Inc., 2010.
- [36] J. J. L. V. Jr., “Double exponential smoothing: An alternative to kalman filter-based predictive tracking,” in *Eurographics Workshop on Virtual Environments*, 2003.
- [37] M. Azimi, *Skeletal Joint Smoothing White Paper*, 2011.
- [38] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter, “The 3d hough transform for plane detection in point clouds: A review and a new accumulator design,” in *3D Research Center and Springer 2011*, 2011.

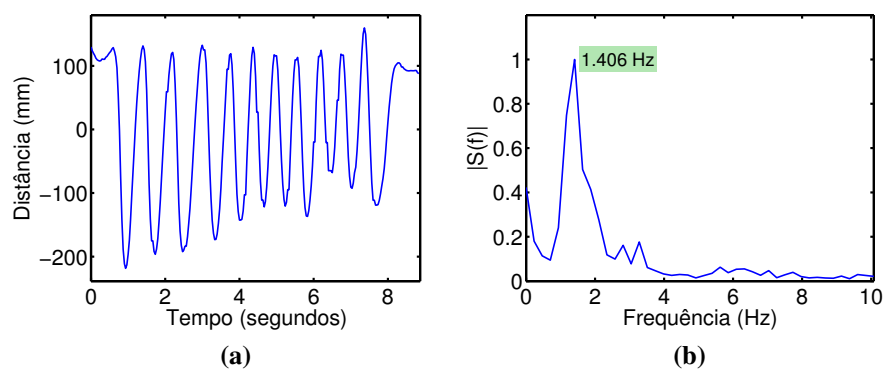




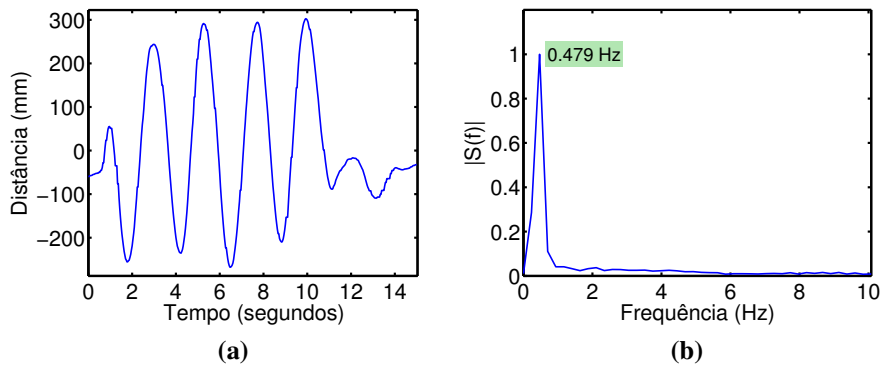
# Anexos A

## Resultados Gestos Periódicos

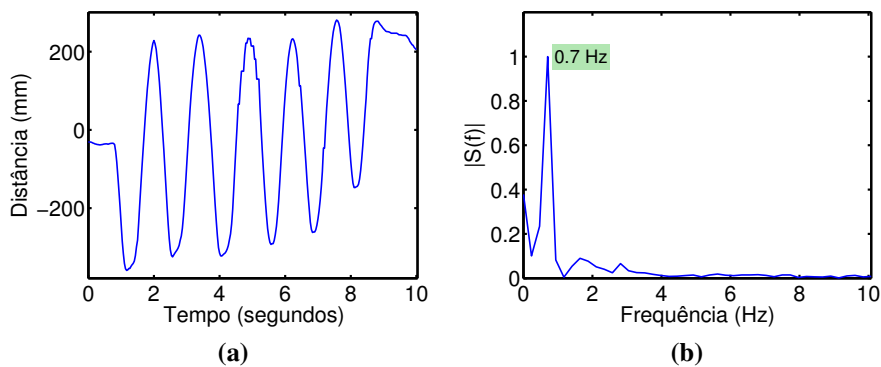
### A.1 Gestos Periódicos



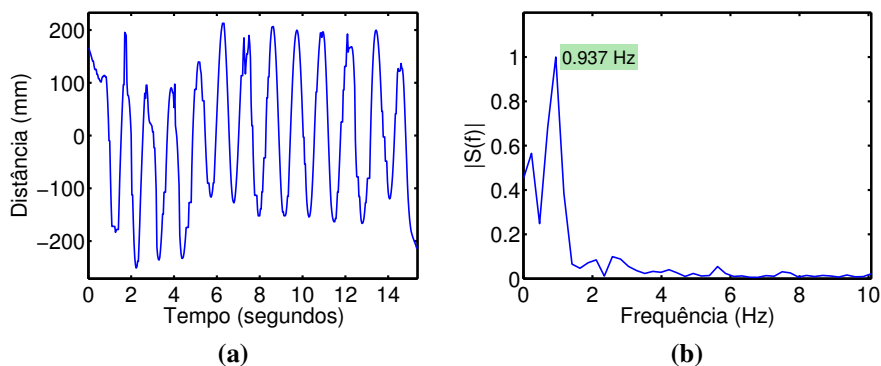
**Figura A.1** – Progressão no tempo da coordenada Y de um gesto com movimento vertical e a respectiva análise espectral com a frequência corresponde de 1.406Hz. (a) - Coordenada Y de um gesto com direcção vertical; (b) - Análise espectral.



**Figura A.2** – Progressão no tempo da coordenada X de um gesto com movimento circular a respectiva análise espectral com a frequência corresponde de 0.479Hz. (a) - Coordenada X de um gesto com movimento circular; (b) - Análise espectral.

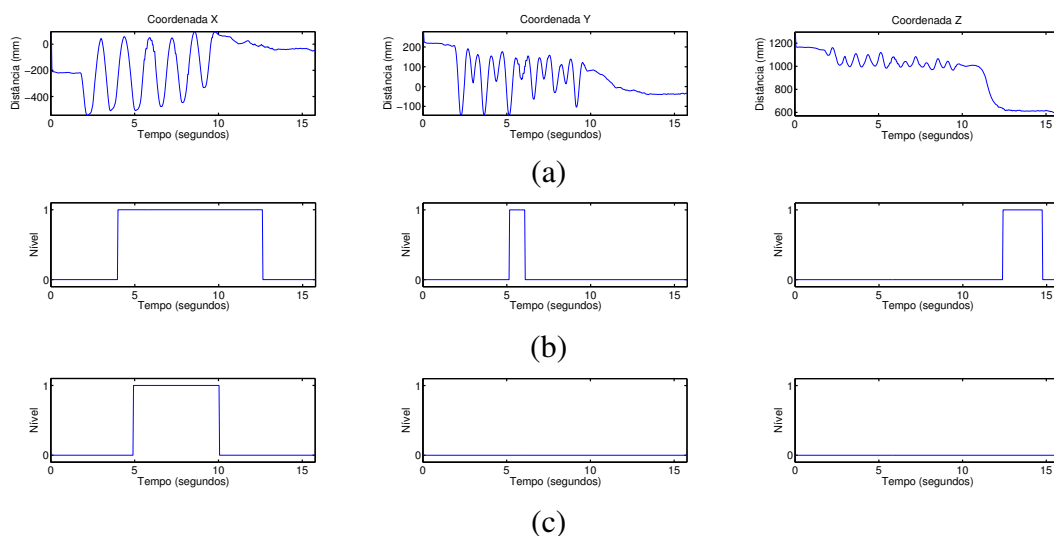


**Figura A.3** – Progressão no tempo da coordenada X do gesto acenar a respectiva análise espectral com a frequência corresponde de 0.7Hz. (a) - Coordenada X do gesto acenar; (b) - Análise espectral.

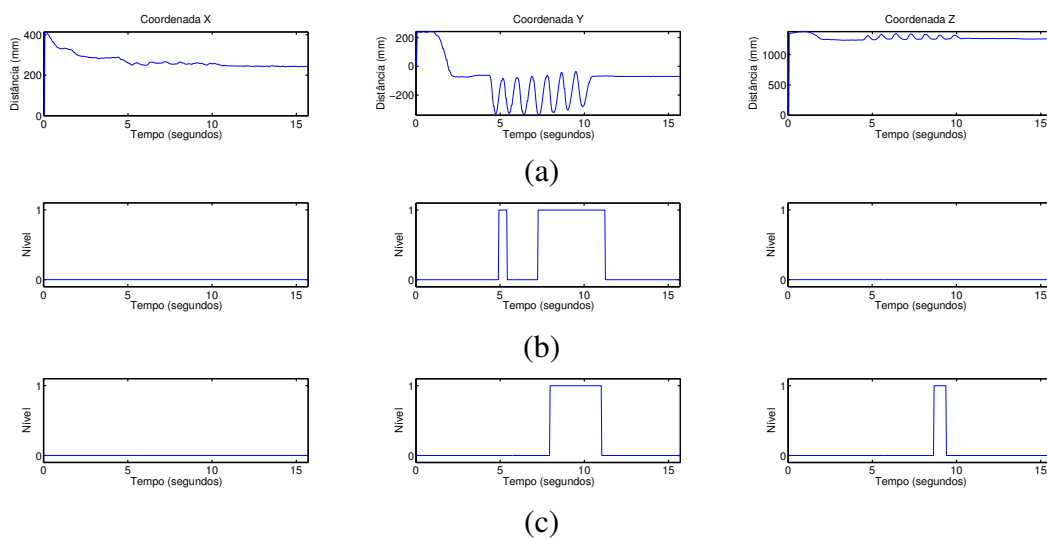


**Figura A.4** – Progressão no tempo da coordenada Z de um gesto com movimento horizontal e a respectiva análise espectral com a frequência corresponde de 0.937Hz. (a) - Coordenada Z de um gesto com movimento horizontal; (b) - Análise espectral.

## A.2 Detecção dos Gestos Periódicos



**Figura A.5** – Ilustração das componentes  $X$ ,  $Y$  e  $Z$  de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas  $X$ ; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação.



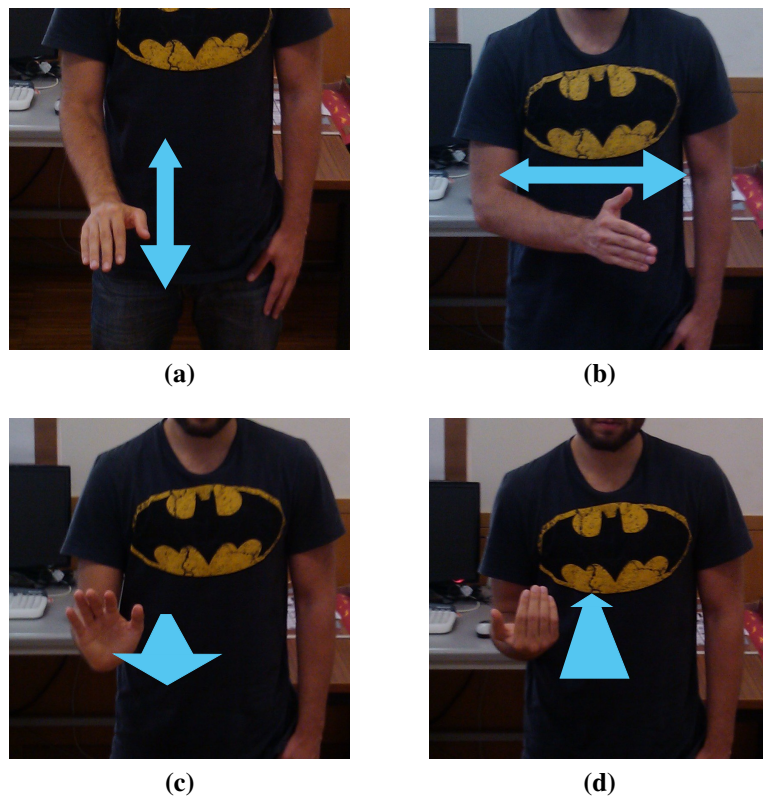
**Figura A.6** – Ilustração das componentes  $X$ ,  $Y$  e  $Z$  de um gesto periódico e resultados obtidos pelos métodos implementados. (a) - Gesto horizontal sobre o eixo de coordenadas  $Y$ ; (b) - Resultados obtidos com recurso à FFT; (c) - Resultados obtidos com recurso à auto-correlação.



## Anexos B

# Reconhecimento de Gesto Periódicos

### B.1 Gestos periódicos a reconhecer



**Figura B.1** – Alguns gestos periódicos. (a) - Gesto que dá a indicação de movimento para baixo; (b) - Gesto que dá a indicação de movimento para a esquerda; (c) - Gesto que dá a indicação de movimento para a frente; (d) - Gesto que dá a indicação de movimento para a trás.