

UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

**Reconhecimento de expressões faciais como
forma de interação entre o computador e
pessoas com deficiência**

Dissertação de Mestrado em Engenharia de Reabilitação e Acessibilidade
Humanas

Andreia Joana Venâncio de Matos

Sob orientação do Professor Doutor Pedro Alexandre Mogadouro do
Couto e Professor Doutor Vítor Manuel de Jesus Filipe



Vila Real, 2016

UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

**Reconhecimento de expressões faciais como
forma de interação entre o computador e
pessoas com deficiência**

Dissertação de Mestrado em Engenharia de Reabilitação e Acessibilidade
Humanas

Andreia Joana Venâncio de Matos

Orientador: Professor Doutor Pedro Alexandre Mogadouro do Couto
Coorientador: Professor Doutor Vítor Manuel de Jesus Filipe

Composição do Júri:

Doutor Pedro Alexandre Mogadouro do Couto, Professor Auxiliar da UTAD.

Doutor Carlos Manuel Gregório Santos Lima, Professor Auxiliar da Universidade
do Minho.

Doutor Luís José Calçada Torres Pereira, Professor Auxiliar da UTAD

Vila real, Julho de 2016

Dissertação exposta à Universidade de Trás-os-Montes e Alto Douro para obtenção do grau de Mestre em Engenharia de Reabilitação e Acessibilidade Humanas, elaborada sob a orientação do Professor Doutor Pedro Alexandre Mogadouro Couto da Universidade de Trás-os-Montes e Alto Douro e do Professor Doutor Vítor Manuel de Jesus Filipe da Universidade de Trás-os-Montes e Alto Douro.

“Somos o que fazemos, mas somos principalmente o que fazemos para mudar o que somos.”

Eduardo Galeano

Resumo

A tecnologia evolui a cada dia que passa e como esta pode ser utilizada em atividades da vida diária convém que a interação seja fácil e intuitiva. No entanto, nem sempre a tecnologia está ao alcance de todos devido a fatores como a falta de acessibilidade, dependência de terceiros e custos monetários. Existem diferentes tipos de interações homem-computador com o objetivo de tornar a tecnologia acessível, mesmo para as pessoas com algum tipo de limitação física. Existem já alguns trabalhos feitos nesta área, mas poucos abordam a questão da deficiência e o quão complicado é para as pessoas com algum tipo de deficiência interagirem com o computador.

Nesta dissertação é proposto um sistema automático de reconhecimento de expressões faciais que possibilita a interação com o computador, a pessoas sem qualquer tipo de mobilidade. O reconhecimento é efetuado através do processamento das imagens adquiridas da face do utilizador, usando um algoritmo composto por três etapas: detecção da face, extração de características (boca e olhos) e classificação da expressão. Com base no algoritmo foi construído um protótipo, que depois de treinado é capaz de reconhecer as expressões faciais de uma pessoa com uma doença neuromuscular degenerativa. O sistema de reconhecimento foi testado por esta pessoa e por cinco utilizadores sem qualquer tipo de deficiência para obter uma avaliação geral. Em ambos os casos obtiveram-se taxas de reconhecimento próximas dos 100%.

Os resultados obtidos comprovam que o reconhecimento automático de expressões faciais pode ser um tipo de interação fácil e intuitivo e pode quebrar barreiras para pessoas com mobilidade reduzida, permitindo-lhes interagir com o computador.

Palavras-chave: Interação homem-máquina, reconhecimento de expressões faciais, pessoas com deficiência física.

Abstract

The technology is every day more present in our lives, therefore, it needs to be easy and intuitive to use. This interaction not always is available to all people for various reasons. These reasons depend on several factors such as: lack of accessibility, dependence on third parties and monetary costs. There are different types of human-computer interactions and in this particular case was used facial expression recognition. There are already some studies in this area, but few of them deal with the issue of disability and how complicated it is for people with disability to interact with the computer. It was constructed a prototype, trained for make recognition of two facial expressions of a person with a degenerative neuromuscular disease. The system of recognition was tested by this person and five users without any kind of disability to obtain a universal evaluation. In both of cases obtained recognition rates nearby of 100%.

In this dissertation presents a facial expression recognition system. This will have as a future goal to allow that people without mobility can to interact with computer.

Through the results obtained demonstrated that the facial expression recognition can be an easy and intuitive interaction type and can break down barriers for people with reduced mobility can interact with computer.

Keywords: Human-computer interaction, facial expression recognition, people with physical disability.

Agradecimentos

A realização desta dissertação foi uma importante etapa na minha vida tanto a nível pessoal como académico. Isto só foi possível com a presença e apoio prestados por algumas pessoas. A estas pessoas gostaria de deixar aqui os meus sinceros agradecimentos:

À Universidade de Trás os Montes e Alto Douro por me ter dado a oportunidade de ter tirado este curso e por todo o apoio durante o mesmo.

Aos meus orientadores, Professor Doutor Pedro Alexandre Mogadouro Couto e Professor Doutor Vítor Manuel de Jesus Filipe por me orientarem neste projeto, por toda a disponibilidade, apoio e dedicação;

Aos meus pais e irmão que me deram a oportunidade de puder realizar este percurso académico e por estarem sempre do meu lado nestes projetos;

À Bárbara Ferreira pelo seu contributo neste estudo e por ter sido tão paciente durante os testes;

Ao Jorge Freitas por todo o seu apoio, atenção e compreensão durante esta etapa;

Aos meus amigos que direta ou indiretamente contribuíram para que o meu percurso académico fosse o melhor possível.

A todos, o meu mais sincero obrigado!

Índice

1. Introdução	1
1.1 Objetivos.....	1
1.2 Metodologia.....	2
1.3 Organização da dissertação	3
2. Estado da arte	4
2.1 Deficiência física	4
2.2 Interação homem-computador	4
2.2.1 Dispositivos de interação	5
2.2.2 Tipos de interação para as diferentes deficiências	6
2.2.3 Reconhecimento de gestos	7
2.2.4 Reconhecimento de voz	8
2.2.5 Reconhecimento de escrita	8
2.2.6 Reconhecimento do tato.....	9
2.2.7 Reconhecimento do olhar	9
2.3 Detecção de faces	9
2.3.1 Técnicas de detecção facial	10
2.3.1.1 Técnicas baseadas em modelos	10
2.3.1.2 Técnicas baseadas na Cor.....	11
2.3.1.3 Técnicas baseadas na geometria da face	12
2.3.1.4 Fisherfaces.....	14
2.3.2 Dificuldades nas técnicas de detecção facial	14
2.4 Reconhecimento de expressões faciais.....	15
2.4.1 <i>Frame</i>	17
2.4.1 Sequência de imagens	19
2.5 Conclusão	23

3. Sistema de interação para computador através de expressões faciais	25
3.1 Detecção da face.....	25
3.2 Extração das características faciais.....	28
3.3 Normalização de escala	28
3.4 Reconhecimento	28
3.5 Conclusão	30
4. Implementação	31
4.1 Estudo de caso	31
4.2 Implementação do algoritmo de Reconhecimento	32
4.3 Implementação da aplicação para interação com o computador	38
4.4 Conclusão	38
5. Avaliação do algoritmo de reconhecimento expressões faciais	39
5.1 Procedimento	39
5.2 Resultados do estudo de caso	40
5.3 Resultados dos testes com pessoas sem deficiência	45
5.4 Comparação de resultados	47
5.5 Conclusão	50
6. Considerações finais e trabalho futuro	51
6.1 Trabalho futuro	51
Bibliografia.....	53
Anexos.....	58
Anexo A – Tabela dos resultados da análise das expressões faciais na boca em pessoas sem deficiência.....	58

Índice de figuras

Figura 1- <i>Trackball</i> da Microsoft.	6
Figura 2- Sistema de Eye Tracking (Majaranta & Bulling, 2014).	6
Figura 3 - Diagrama dos tipos de interação para as diferentes deficiências.	7
Figura 4 - Resultados obtidos por Yow e Cipolla (1997).	13
Figura 5 - Estrutura básica da análise do reconhecimento de expressões faciais (adaptado de (Sumathi, Santhanam, & Mahadevi, 2012).	16
Figura 6 - Modelo facial humano.	17
Figura 7 - Resultados obtidos por Thai et al. (2011).	19
Figura 8 - Pontos característicos do método de Tong et al. (2007).	20
Figura 9 - Funcionamento do método de Varona et al. (2008).	21
Figura 10 - Diagrama do sistema de reconhecimento de expressões faciais.	25
Figura 11 - Exemplos dos recursos de Haar. A imagem A é uma divisão horizontal, a B é uma divisão vertical, a C tem 2 divisões horizontais e a D é criada por divisões horizontais e verticais (Viola & Jones, 2001).	26
Figura 12 - As duas melhores características aplicadas durante a fase de treino (Viola & Jones, 2001).	27
Figura 13 - Resultados do método desenvolvido por Viola e Jones (Viola & Jones, 2001). ...	27
Figura 14 - Exemplo de média das imagens de treino. As imagens acima são imagens de treino, sendo que a debaixo é a média das imagens acima.	29
Figura 15 - Funcionamento de um exemplo baseado em análise de componentes principais.	30
Figura 16 - Detecção da face.	32
Figura 17 - Detecção da boca.	33
Figura 18 - Erro na detecção da boca.	33
Figura 19 - Detecção correta da boca a sorrir e a mandar beijinho.	33
Figura 20 - Detecção dos olhos a mandar beijinho e a sorrir.	34
Figura 21 - Detecção do olho esquerdo.	34
Figura 22 - Menu do algoritmo.	35
Figura 23 - Reconhecimento da expressão facial correto.	36
Figura 24 - Quando a expressão não é idêntica com nenhuma das expressões da base de dados.	36
Figura 25 - Reconhecimento da expressão facial através do olho esquerdo.	37

Figura 26 - Diagrama do funcionamento do algoritmo desenvolvido.	37
Figura 27- Erro na classificação da expressão facial.	41
Figura 28 - Erro na classificação da expressão facial: excluiu a expressão quando ela existe.41	
Figura 29 - Erro na detecção da boca.....	41
Figura 30 - Olho esquerdo de uma pessoa sem deficiência. Na imagem da esquerda a pessoa está a sorrir e na imagem da direita a pessoa está a fazer beijinho. A semelhança entre as duas é muito grande.....	47

Índice de gráficos

Gráfico 1- Gráfico referente ao modelo de classificação do estudo de caso.....	44
Gráfico 2 - Gráfico do modelo de avaliação de pessoas sem deficiência.	46
Gráfico 3- Taxas de acerto em ambos os casos.....	47
Gráfico 4 - Taxas de erro na detecção da boca.	48
Gráfico 5 - Resultados da precisão, sensibilidade e especificidade do estudo de caso.....	49
Gráfico 6 - Resultados da precisão, sensibilidade e especificidade dos testes das pessoas sem deficiência.	49

Índice de tabelas

Tabela 1 - Resultados da análise das expressões faciais na boca do estudo de caso.	40
Tabela 2 - Matriz de confusão para o estudo de caso.....	43
Tabela 3 - Avaliação dos resultados no estudo de caso.	43
Tabela 4- Matriz de confusão dos resultados de pessoas sem deficiência.	45
Tabela 5 - Avaliação dos resultados das pessoas sem deficiência.	46
Tabela 6 - Resultados da análise das expressões faciais na boca em pessoas sem deficiência.....	59

Siglas e acrónimos

CIF	Classificação Internacional de funcionalidade
RGB	Escala de cores: Red, Green, Blue
BCI	Brain Computer Interface
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
HMM	Hidden Markov Model
FAP	Facial Action Parameters
FACS	Facial Action Coding System
RNA	Redes Neurais Artificiais
SVM	Suport Vector Machines
FAV	Unidades de ação facial

1. Introdução

Com o notório desenvolvimento tecnológico ao longo dos últimos anos, a tecnologia está cada vez mais presente no nosso dia-a-dia. Existe a necessidade de garantir a todas as pessoas o igual acesso à tecnologia, contudo, existem pessoas com determinadas deficiências a quem é impossibilitada essa mesma interação. Amputações, tetraplegia ou paraplegia, doenças neuromusculares degenerativas ou paralisia cerebral são deficiências que não permitem o movimento ou a força necessária para, por exemplo, o controlo do computador.

De forma a combater a exclusão social das pessoas com deficiência na interação com a tecnologia, surge a ideia de utilizar o reconhecimento de expressões faciais como forma de interação. Existem diversas aplicações que utilizam o reconhecimento das expressões faciais como a biometria, segurança, vigilância, verificação da identidade, sistema de justiça, investigações em bases de dados e interação com o computador (Bakshi & Singhal, 2014). É nesta última aplicação que, utilizando métodos de visão por computador e processamento digital de imagem, é realizado o reconhecimento da expressão facial como forma de interação, servindo como intermediário entre o utilizador e o computador. As expressões faciais são pré-programadas, sendo atribuída uma função específica no computador por cada expressão ou conjunto de expressões feitas pelo utilizador.

1.1 Objetivos

Com a realização da presente dissertação pretende-se criar uma interação com o computador para pessoas que não o consigam pelos meios já disponíveis. Pretende-se utilizar a visão por computador para o reconhecimento de expressões faciais e para tal será desenvolvido um sistema de auxílio à interação com o *hardware*. Este deve ser capaz de identificar a expressão facial do utilizador para lhe ser associada a uma ação programada, a realizar pelo computador. É necessário ter uma base de dados com as diferentes expressões faciais possíveis, de forma a que este sistema possa ser aplicado a toda a população com diferentes características. O sistema poderá constituir um instrumento de apoio à aprendizagem em pessoas com mobilidade reduzida, possibilitando ao utilizador o acesso ao computador. Para esse efeito utilizou-se um estudo de caso para testes e que no futuro poderá tirar partido da utilização do sistema. São definidos os seguintes objetivos específicos para o trabalho:

- Desenho da arquitetura de um sistema capaz de reconhecer expressões faciais através duma base de dados e associar a estas funções relativas à interação com o computador.
- Construção da aplicação.
- Avaliação da aplicação desenvolvida com respetivos testes.

1.2 Metodologia

Para a realização desta dissertação será necessário efetuar uma breve abordagem às técnicas de interação homem-computador bem como às técnicas de reconhecimento de expressões faciais. Será também feita uma pesquisa sobre os diferentes trabalhos já realizados no âmbito do tema do projeto. Após este estudo será apresentada uma proposta de arquitetura global do sistema com a sua especificação, implementação e realização de testes para a avaliação do protótipo desenvolvido.

Assim numa primeira fase será efetuada uma abordagem às diferentes técnicas de interação homem-máquina, aos dispositivos usados nesta interação e relacionando os diferentes tipos de interação com as diferentes deficiências. De seguida serão apresentadas técnicas de reconhecimento facial e mais especificamente de reconhecimento de expressões faciais. Será também feita uma breve pesquisa de diversos projetos que foram desenvolvidos na área do reconhecimento de expressões.

Numa segunda fase será feita uma descrição de todo o sistema pretendido. Será apresentada uma arquitetura global do sistema seguido de uma especificação completa e uma pequena apresentação do estudo de caso. Por último será descrita a implementação de um protótipo sendo justificado cada opção tomada.

A terceira e última fase consiste na apresentação dos testes realizados ao utilizador do estudo de caso e a outros utilizadores sem deficiência. Será explicado todo procedimento na realização dos testes bem como os resultados dos mesmos.

1.3 Organização da dissertação

No primeiro capítulo foi feita uma introdução ao trabalho realizado e uma apresentação da motivação e objetivos do projeto. Ainda neste capítulo foi descrita a metodologia e a organização da dissertação.

No segundo capítulo é apresentada uma revisão do estado da arte sobre os temas abordados neste trabalho e é apresentado também um conjunto de trabalhos já realizados por outros autores.

No terceiro capítulo é realizada uma descrição do sistema de reconhecimento de expressões faciais desenvolvido. É apresentada a descrição de cada etapa do sistema de reconhecimento.

No quarto capítulo é dado a conhecer o estudo de caso e a implementação do protótipo com a explicação das várias partes do algoritmo. É também explicado o funcionamento da aplicação que faz a interação com o computador.

No quinto capítulo são apresentados os resultados, e posterior análise, dos testes efetuados ao protótipo desenvolvido para o reconhecimento de expressões faciais num estudo de caso específico e em pessoas sem deficiência.

No sexto e último capítulo são apresentadas as conclusões retiradas deste trabalho, assim como propostas de trabalho futuro de forma a melhorar e potenciar todo o trabalho desenvolvido nesta dissertação.

2. Estado da arte

Para a realização deste trabalho foi importante estudar e compreender alguns conceitos referentes ao tema da acessibilidade para pessoas com deficiência. Como o público-alvo desta dissertação é pessoas com deficiência física então inicialmente é feita uma abordagem aos tipos de deficiência física e doenças que condicionam o movimento dos membros. De seguida é feita uma abordagem à interação homem máquina, com uma descrição dos tipos existentes. Apresenta-se alguns conceitos sobre reconhecimento facial e técnicas mais usadas. Por fim, é apresentado o reconhecimento de expressões faciais, relatando também alguns estudos que foram feitos na área.

2.1 Deficiência física

Segundo a CIF (2004), deficiências são “problemas nas funções ou nas estruturas do corpo, tais como, um desvio importante ou uma perda.” (Saúde, 2004). Existem vários tipos de deficiências, mas a deficiência física é a que demonstra mais dificuldade em usar o *hardware* de acesso ao computador. A deficiência física é considerada como “alteração completa ou parcial de um ou mais segmentos do corpo humano, acarretando o comprometimento da função física, apresentando-se sob a forma de paraplegia, paraparesia, monoplegia, monoparesia, tetraplegia, tetraparesia, triplegia, triparesia, hemiplegia, hemiparesia, amputação ou ausência de membro, paralisia cerebral, membros com deformidade congênita ou adquirida, exceto as deformidades estéticas e as que não produzam dificuldades para o desempenho de funções” (Bersch & Machado, 2007).

Algumas doenças que afetam a mobilidade total são: doenças neuromusculares degenerativas (como a esclerose múltipla e a esclerose lateral amiotrófica por exemplo), doença de Parkinson, paralisia cerebral, entre outras (Pedro & Ribeiro, 2010) (Rito, 2006). Também em casos de amputação de membros, o recurso a interfaces alternativas facilita a interação com o computador.

2.2 Interação homem-computador

A tecnologia evolui a cada dia que passa e como esta pode ser utilizada em atividades da vida diária convém que a interação seja fácil e intuitiva. Existem diferentes tipos de interações entre o homem e a máquina como por exemplo reconhecimento de fala, de gestos e de palavras manuscritas. Perfeito seria usar os diferentes tipos de interações em simultâneo

para evitar os problemas que estas têm individualmente (Valverde, Pereira, & Vassallo, 2011).

A interação homem-máquina é desenvolvida com vários objetivos em diferentes áreas como: realidade virtual, saúde, jogos de computador interativos, acessibilidade, entre outras (Ren, Meng, & Yuan, 2011). A interação humano-computador é uma área interdisciplinar (abrangendo áreas como engenharia, psicologia, ergonomia e design) que trabalha o *design*, implementação e avaliação das formas com que os seres humanos utilizam e interagem com dispositivos tecnológicos. O objetivo principal da interação é melhorar ao máximo a usabilidade. Normalmente com uma melhor interação, maior a produtividade (Kim G. J., 2015).

Hansen (1971) referia que para o desenvolvimento de uma boa interação é necessário “*Know thy user*”, isto é “conhece o teu utilizador”. É importante saber quais os tipos de utilizadores e quais as suas capacidades para desenvolver a melhor interação. Características como idade, escolaridade, experiência com tecnologia e cultura são das mais importantes para aprimorar o contato com a tecnologia (Kim G. J., 2015).

A visão, o tato e a audição são os sentidos mais presentes na interação homem-máquina. Os dedos, voz, olhos, cabeça e posição do corpo são os intervenientes nesta interação. Os dispositivos mais básicos de interação com o computador são o teclado, o rato e ecrãs táteis (Dix, Finlay, Abowd, & Beale, 2005).

2.2.1 Dispositivos de interação

Existem diversos dispositivos de interação entre o homem e a tecnologia. Os mais usuais são o rato e *touchpad* (teclado tátil). Existem outros, menos utilizados, como *trackball* (figura 1), *thumbwheel* (botão rotativo, figura 2) e *joystick* (Dix, Finlay, Abowd, & Beale, 2005).



Figura 1- *Trackball* da Microsoft.

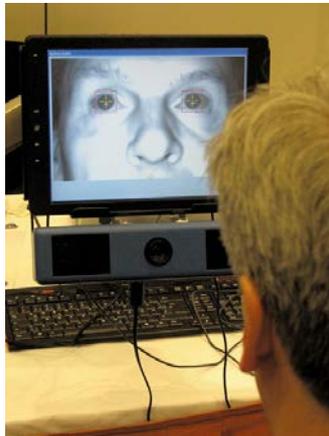


Figura 2- Sistema de Eye Tracking (Majaranta & Bulling, 2014).

Existem vários dispositivos de interação que foram pensados especialmente para pessoas com deficiência como o *Eye-Tracking* (controle através da retina do olho), BCIs (“*Brain Computer Interfaces*”, como exemplo o Cyberlink), *switch*, entre outros (Guerreiro & Jorge, 2006).

2.2.2 Tipos de interação para as diferentes deficiências

Existem diferentes tipos de interação para cada deficiência (Figura 3). A escolha do dispositivo ou interface mais apropriado para a interação com o computador depende fortemente da deficiência do utilizador.

- Deficiência visual: através do toque ou áudio;
- Deficiência auditiva: não impõe grandes dificuldades nas diferentes interações, excetuando nas que utilizem som;
- Deficiência motora: interações através do olhar, voz, BCIs;

- Deficiência na fala: não existe dificuldades de interação, a não ser que seja necessário o reconhecimento da voz;
- Dislexia: para facilitar a leitura e a escrita, sistemas de voz podem ser muito úteis, sendo que *software* de revisão de escrita também pode ser vantajoso;
- Deficiência cognitiva: a nível físico não há impedimentos, apenas será necessário *software* e *hardware* intuitivo e fácil de usar (Dix, Finlay, Abowd, & Beale, 2005).

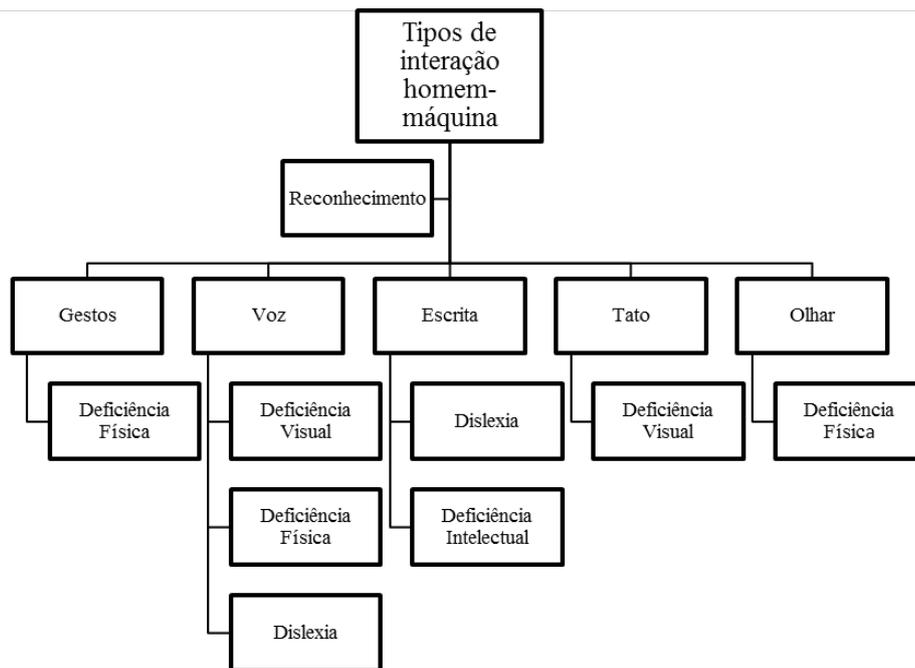


Figura 3 - Diagrama dos tipos de interação para as diferentes deficiências.

O trabalho desta dissertação incide sobre o tipo de interação com o computador através do reconhecimento de expressões faciais de maneira a responder às necessidades de utilizadores que não possuem qualquer mobilidade nos membros.

2.2.3 Reconhecimento de gestos

A interação homem-computador feita por reconhecimento de gestos é das interações mais naturais e simples pois é idêntica à interação entre humanos. Este tipo de interação facilita muito o controlo do computador para pessoas com deficiência auditiva (Ren, Meng, & Yuan, 2011).

Uma das formas de fazer o reconhecimento de gestos é através de luvas de dados que funcionam com sensores eletromecânicos ou magnéticos. É o dispositivo mais complexo em

medições em tempo real. No entanto tem algumas desvantagens: é dispendioso, dificulta o gesto natural da mão e necessita de muita calibração para ter uma medição exata. Outro dispositivo bastante usado para a detecção de gestos é o Microsoft Kinect, que no entanto apresenta limitações no desempenho quando se faz um gesto apenas com a mão ou dedos (Ren, Meng, & Yuan, 2011).

Existem outras maneiras de detetar gestos, utilizando processamento digital de imagem, por exemplo. A primeira etapa (talvez a mais importante) para o reconhecimento de gestos é a detecção da cor da pele. Esta etapa é considerada uma etapa de pré-processamento e é onde é feita a segmentação da cor da pele, fazendo com que apenas as regiões com a cor da pele sejam as detetadas. Uma das maneiras de fazer esta segmentação é utilizando o modelo de cores RGB juntamente com uma combinação de 16 Gaussianas para se determinar a probabilidade de cada *pixel* ser da cor pele. Os pixéis mais claros têm uma maior probabilidade de serem cor de pele (Valverde, Pereira, & Vassallo, 2011).

Existem vários dispositivos que reconhecem os gestos: Kinect, 3GEAR, Softkinetic, Leap Motion e a Myo (Silva R. K., 2013).

2.2.4 Reconhecimento de voz

O reconhecimento de voz ainda é um sistema bastante problemático pois existem diversas questões que ainda precisam de ser respondidas e desenvolvidas. Os maiores problemas são: diferentes vozes, sotaques, emoções, vocabulário específico de algum local, entre outros. Este tipo de interação é usado essencialmente como um dispositivo de texto alternativo e em situações em que não é apropriado o uso de teclado (Dix, Finlay, Abowd, & Beale, 2005). Atualmente o Microsoft Windows 8 e 10 têm reconhecimento de voz, no entanto está limitado aos seguintes idiomas: inglês (Estados Unidos e Reino Unido), francês, alemão, japonês, mandarim (chinês simplificado e chinês tradicional) e espanhol, limitando assim utilizadores de alguns países (como Portugal) da utilização deste tipo de interação.

2.2.5 Reconhecimento de escrita

O reconhecimento de escrita, apesar de ser fácil de usar, ainda tem algumas problemáticas como as diferenças nos tipos de letra. Por isso que é mais vantajoso usar letra escrita a computador por ser mais fácil de reconhecer. Nestes casos utiliza-se os *scanners* como dispositivo de interação. Este tipo de interação é usado sobretudo nos telemóveis

porque ser fácil implementar teclados pequenos e precisos (Dix, Finlay, Abowd, & Beale, 2005).

2.2.6 Reconhecimento do tato

O uso do tato numa interface é conhecido por interação háptica e cinestésica. É das diferentes interações a mais usada pelas pessoas com deficiência visual. O dispositivo mais utilizado para este tipo de deficiência é a linha braille (Dix, Finlay, Abowd, & Beale, 2005).

2.2.7 Reconhecimento do olhar

O contato com os olhos e a direção do olhar são pistas importantes na comunicação humana. Este contato visual é utilizado na regulação da interação estabelecendo uma ligação emocional e indica também qual o alvo do nosso interesse visual. O *Eye tracking* refere-se ao processo de acompanhamento dos movimentos dos olhos, focando-se no ponto do olhar do utilizador que está focado na cena visual. Existem várias técnicas de reconhecimento do olhar tais como: “*videoculography*”, vídeo infravermelho, reflexão da pupila corneal e Electro Oculografia. Existe um grande potencial no uso do olhar nas interfaces homem-computador seja como um método de entrada ou como uma fonte de informação para interfaces proactivas (Majaranta & Bulling, 2014).

2.3 Detecção de faces

Na visão por computador (que simula o comportamento visual humano), a deteção facial de um indivíduo constitui um grande desafio. Para esta deteção, pode ser necessário um processo de treino cuidadoso (para que o programa seja o mais correto possível), sendo efetuado em diversas etapas. É necessário que haja a deteção das faces e a sua segmentação, de maneira que este processo seja o mais rápido e concreto possível (Filipe, 2008).

Existem certas aplicações em tempo real que necessitam de uma resposta rápida e uma das técnicas que podem usar é a deteção de faces e expressões faciais humanas. Estas aplicações são: sistemas de vigilância, verificação de identidade, sistemas de justiça penal, investigações de bases de dados de imagens, aplicações “Smart Card”, entre outras (Bakshi & Singhal, 2014). Estas usam técnicas que requisitam o conhecimento sobre a localização das faces dentro da imagem. Esta localização deve ser detetada automaticamente (Filipe, 2008).

A deteção de faces é uma técnica bastante complexa pois existem fatores, tais como, cabelo, maquilhagem, barba, bigode, óculos ou chapéus, que dificultam a localização das diferentes características faciais. Além deste problema, existe ainda outro que é relativo à

escala e orientação da face humana na imagem, o que leva à não utilização de modelos fixos para localizar as características. A presença de outros objetos e ruídos na imagem torna-se também uma dificuldade no processamento desta técnica (Filipe, 2008).

As técnicas reconhecimento e detecção facial podem ser distribuídas por quatro categorias:

- Métodos baseados no conhecimento: baseiam-se nas características da face. Geralmente são utilizados para localização.
- Abordagens de características invariantes: utilizam características estruturais que não suportam influência das mudanças de posição, cor, iluminação, etc. esta categoria é utilizada também para efeitos de localização.
- Métodos de casamento de padrões (“*template matching*”): são utilizados diversos modelos de padrões de faces. A detecção é feita entre a comparação dos modelos com as imagens. Nesta categoria são usados para a localização e reconhecimento.
- Métodos baseados na aparência: são usados modelos, mas estes são instruídos a partir de treino. O algoritmo aprende a reconhecer a face humana. São utilizados essencialmente para a detecção de faces (Almeida, 2006).

2.3.1 Técnicas de detecção facial

Existem diferentes métodos para detetar faces humanas. Alguns métodos utilizam a informação baseada na aparência ou em modelos, já outros usam a informação da cor (Filipe, 2008).

2.3.1.1 Técnicas baseadas em modelos

Esta técnica extrai características faciais com base em modelos de treino (Dhawan & Dogra, 2012).

Yuille *et al.* (1989) desenvolveram um método baseado em modelos deformáveis. Esses modelos parametrizados permitem um conhecimento *a priori* sobre a forma esperada das características faciais para guiar o processo de detecção. (Dhawan & Dogra, 2012)

Sung e Poggio (1998) apresentaram um sistema para detetar faces frontais em situações complexas usando uma aprendizagem fundamentada em modelos. Em cada posição da imagem, um vetor de características é processado entre o modelo local da imagem e o modelo de treino. O sistema tem um classificador treinado que se baseia no vetor de características e

avalia se existe ou não uma face humana na orientação atual da imagem. Este método foi testado em duas bases de dados diferentes. A primeira base de dados é composta por 301 imagens frontais e quase frontais de 71 pessoas. A outra base de dados tem 23 imagens, contendo um total de 149 modelos de faces. Para a primeira base de dados, o método determinou 96.3% de todos os modelos de faces e devolveu três falsas detecções (Filipe, 2008).

2.3.1.2 Técnicas baseadas na Cor

Alguns modelos de cor (RGB por exemplo) ajudam a detetar a região da pele. A imagem obtida depois da detecção da zona da pele é binarizada. Isto é feito para eliminar a cor e os valores de saturação e considerar apenas a parte de luminância. De seguida elimina-se o ruído. No final, as diferentes características faciais (olhos, nariz e boca) podem ser extraídas (Bakshi & Singhal, 2014).

Daí e Nakano (1996) decidiram separar a região próxima do laranja no espaço de cor YIQ como região idêntica à pele e eliminaram as restantes regiões. Depois aplicaram as características de textura em imagens com nível de cinza para reconhecer faces nas regiões da pele. No primeiro teste foram escolhidas ao acaso 10 pessoas numa base de dados de faces e fizeram cinco ou seis testes por pessoa. Nestes testes, estavam abrangidas as faces com rotação, inclinação e com diferentes expressões. As taxas de verdadeiras detecções averiguadas foram de 98%. A grande desvantagem deste método é que não deteta faces com oclusão parcial, pessoas com óculos e faces de perfil (Dai & Nakano, 1996).

Cai e Goshtasby (1999) propuseram um método que faz a detecção das faces através de processamentos realizados no espaço de cor da Commission Internationale d'Éclairage Lab (CIE Lab). Cada cor é modificada no nível de cinza correspondente e usam essa informação com uma função de distribuição de probabilidades para determinar as regiões que podem ser faces. De seguida colocam um modelo de face para o reconhecimento final. Nos testes, quando foi usado um limiar de 0.5, foram perdidas 13% das faces e 8.7% das faces reconhecidas de forma errada não eram faces (Cai & Goshtasby, 1999).

O algoritmo proposto por Yachida et al. (1999) deteta faces em imagens coloridas, fundamentando-se na teoria *fuzzy*. Trabalha com dois modelos *fuzzy*: um para dizer a cor da pele e outro para descrever a cor do cabelo, usando um espaço de cor percentual para ampliar a precisão do sistema. Criaram também um algoritmo para obter as regiões da cor da pele e

outro para obter as regiões da cor do cabelo. Nos testes foi utilizada uma base de dados com 233 faces, 186 eram faces asiáticas e as outras caucasianas. O tamanho das faces varia entre 20 x 24 a 200 x 240 pixels. A taxa de verdadeiras detecções foi de 97% em imagens com tamanho de faces superiores a 50 x 60 pixels. Os erros deste algoritmo devem-se a causas, tais como: a variação da iluminação, a oclusão facial, as faces adjacentes (se as faces estiverem muito próximas, os algoritmos que descrevem a cor da pele e do cabelo podem ser unidos, originando numa forma bem diferente de uma única cabeça) e o estilo do cabelo (Chen, Wu, & Yachida, 1995).

Kim et al. (2000) apresentaram um método de detecção de faces fundamentado num objeto. Este algoritmo tem dois passos: segmentação e detecção da região facial. No primeiro passo, a imagem de entrada é segmentada por um algoritmo genético dentro de algumas regiões iniciais. De seguida, as regiões são unificadas de acordo com uma semelhança espacial, pois as regiões de formação de um objeto partilham de algumas características espaciais análogas. No segundo passo, as regiões faciais são reconhecidas a partir dos resultados do primeiro passo, usando um modelo da cor da pele. A taxa de detecções corretas é de 82% e a taxa de detecções erradas é de 17%. Este valor é devido a erros que acontecem na segmentação da imagem (Kim, Kim, Hwang, & Kim, 2000).

2.3.1.3 Técnicas baseadas na geometria da face

Neste tipo de técnicas é extraído o tamanho e a posição relativa de componentes importantes da imagem. Inicialmente é detetada a direção e bordas das componentes importantes e de seguida são feitos os vetores dessas bordas (Bakshi & Singhal, 2014).

Mark Nixon (1985) apresenta um algoritmo baseado na geometria da face. Para a detecção do olho apresenta uma medição geométrica do espaçamento dos olhos com a transformada de Hough de maneira a detetar uma forma circular e elipsoidal. O resultado deste algoritmo revela que é possível obter a medida do espaçamento entre os olhos através da detecção das duas íris (Dhawan & Dogra, 2012).

Yow e Cipolla (1997) apresentaram um algoritmo que usa filtros Gaussianos para localizar formas das características faciais nas imagens (Figura 4). Comparando estas características com as de um modelo de face, as possíveis faces são localizadas. Este algoritmo pode reconhecer faces de diferentes tamanhos numa imagem se se mudar o tamanho do filtro Gaussiano. Nos testes, a detecção correta teve uma taxa de 85% numa base

de dados com 110 imagens de faces com diferentes escalas, orientações e ângulos (Yow & Cipolla, 1997).



Figura 4 - Resultados obtidos por Yow e Cipolla (1997).

Jeng et al. (1998) detetam as faces fundamentando-se no modelo da geometria da face. Nesse algoritmo, estabelecem-se as possíveis localizações dos olhos em imagens binarizadas. Para cada provável par de olhos, o sistema faz uma pesquisa à procura de um nariz, uma boca e sobrancelhas. Cada característica facial tem uma avaliação agregada, que é usada para determinar a face candidata. O método mostrou uma taxa de 86% de detecção numa base de dados com 114 imagens. Quando numa imagem existem várias faces ou nenhuma, o sistema não faz uma detecção correta (Jeng, Liao, Han, Chern, & Liu, 1998).

Lin e Fan (2001) criaram um método onde a detecção de faces usa as relações geométricas do triângulo. Este algoritmo tem duas partes principais: a primeira consiste na pesquisa de regiões que possam possuir faces e a segunda efetua a verificação da face.

Quatro passos da primeira parte do algoritmo:

1. Ler a imagem e transformá-la numa imagem binária;
2. Classificar os quatro elementos ligados na imagem para formarem vários blocos e encontrar o centro de cada um deles;
3. Detetar-se quaisquer três centros, de três diferentes blocos, para formarem um triângulo isósceles no caso de uma imagem frontal ou um triângulo retângulo se for uma imagem de perfil;
4. Juntar os blocos que satisfazem o critério de triângulo como possível face.

A segunda parte do algoritmo é constituída por três fases:

1. Normalizar o tamanho de todas as possíveis regiões;
2. Prover a cada região de possível face normalizada um peso na função máscara;

3. Efetuar a verificação por *thresholding* do peso obtido na fase anterior.

Este algoritmo suporta diferentes tamanhos de faces nas imagens, diferentes condições de iluminação, ruído, problema de desfocagem, variação de pose e de expressão. O método pode também detetar faces de perfil, faces com problema de oclusão parcial da boca e com óculos de sol. Para os testes foram utilizadas 500 imagens, de 450 pessoas diferentes, sendo no total 600 faces. O método teve uma taxa de sucesso de 98% (Lin & Fan, 2001).

2.3.1.4 Fisherfaces

Esta técnica de reconhecimento facial foi desenvolvida por Belhumeur, Hespanha e Kriegman (1997) e é feita uma análise discriminante de Fisher que encontra subespaços que dividem de maneira mais eficiente as classes do sistema. Esta técnica é supervisionada, isto é sabe-se previamente as classes do sistema e isso pode aperfeiçoar a projeção dos dados (Silva V. A., 2008).

2.3.2 Dificuldades nas técnicas de deteção facial

Na deteção facial existem alguns fatores que contribuem para algumas falhas dos algoritmos. Estes podem dividir-se em dificuldades no espectro visível, detetar as diferentes características faciais e a localização da face perante a câmara de filmar.

A deteção facial baseia-se em análise de imagens de entrada, no entanto tem algumas limitações no espectro visível, nomeadamente: iluminação, poses, mudanças de expressão facial e disfarces. Para minimizar o problema da iluminação utilizam-se filtros de processamento de imagem e modelos faciais. Uma das soluções para as mudanças de poses é a utilização de câmaras de infravermelhos. Estas detetam as características em diferentes ambientes de iluminação, informação anatómica e expressões. A visão térmica dos infravermelhos consegue detetar os disfarces faciais, no espectro visível isso não era possível. No entanto, o uso de óculos pode dificultar o reconhecimento mesmo utilizando infravermelhos devido ao vidro das lentes diminuir a precisão do método.

Existem também diversas características que definem uma face humana tais como: olhos, nariz, boca, sobrancelhas, cabelo, entre outras. As informações retiradas destas características são fundamentais para a deteção facial. No entanto, existem vários fatores que dificultam a extração de características, tais como:

- Pose: existem duas poses principais (frontal e perfil) que têm diferentes graus de variação, podendo ocultar totalmente ou parcialmente algumas características.
- Outras estruturas: algumas pessoas possuem barba, óculos ou bigode, que pode impedir a extração de algumas características.
- Expressão facial: modificam as características da face (Almeida, 2006).

Diferentes investigadores têm usado a extração de características faciais em imagens de infravermelhos. Estas investigações basearam-se em padrões binários locais, Transformada de Wavelet, Transformada de Curvelet, rede vascular e perfusão sanguínea. Também foram utilizadas redes neurais para estudos e a taxa de sucesso foi de 92% (Arya, Neeraj, & Bhatia, 2015).

Relativamente à localização da face existem alguns fatores que conduzem a diversas dificuldades. São estes fatores que fazem com que haja muita pesquisa nesta área. Os fatores são:

- Oclusão: as imagens podem ter oclusão parcial ou total de certas características devido a objetos presentes na cena.
- Orientação da imagem: o posicionamento da face pode variar.
- Condições da imagem: pode haver variações de luminosidade ou presença de ruídos (Almeida, 2006).

2.4 Reconhecimento de expressões faciais

A expressão facial dum pessoa é a demonstração do seu estado afetivo, da atividade cognitiva, do pensamento. Transmite uma comunicação não-verbal com outra pessoa, sendo que contribui 55% numa conversa. Antes de 1977 era complicado identificar as expressões faciais pois essa identificação era feita através de observadores, no entanto este método não era viável pois os observadores podiam ser influenciados e as expressões variam nas diferentes culturas. Para resolver este problema Ekman e Friesen (1978) desenvolveram um sistema FACS (Facial Action Coding System) que descreve as diferentes expressões faciais, tornando-se num sistema padrão. Este método baseia-se no movimento dos músculos: identifica os músculos faciais que individualmente ou em grupo provocam mudanças nos comportamentos da face (Sumathi, Santhanam, & Mahadevi, 2012).

Para a análise de expressões faciais é necessário identificar e classificá-la segundo o manual de FACS. Existem três etapas nesta análise: aquisição da face, extração da expressão facial e o reconhecimento da expressão. Na primeira etapa, o sistema deve reconhecer faces (2D ou 3D) nas imagens de entrada. Depois deve ser detetada a expressão facial segundo diversas características. Estas podem ser de dois tipos: características geométricas ou permanentes e características de aparência ou transitórias (Sumathi, Santhanam, & Mahadevi, 2012).

Características geométricas ou permanentes: estão sempre presentes no rosto, mas que podem ser deformadas quando é feita uma expressão. Referem-se aos olhos, sobrancelhas, boca e nariz. Os componentes da cara são extraídos de modo a criar um vetor para representar a geometria da face.

Características de aparência ou transitórias: referem-se a características que aparecem na face durante a expressão facial. Por exemplo: rugas e protuberâncias.

No final da análise é classificada a expressão facial. O classificador deve identificar não só emoções básicas, mas também expressões espontâneas e a intensidade da expressão. Pode ver-se pela Figura 5 a sequência de um sistema de reconhecimento de expressões faciais (Sumathi, Santhanam, & Mahadevi, 2012).

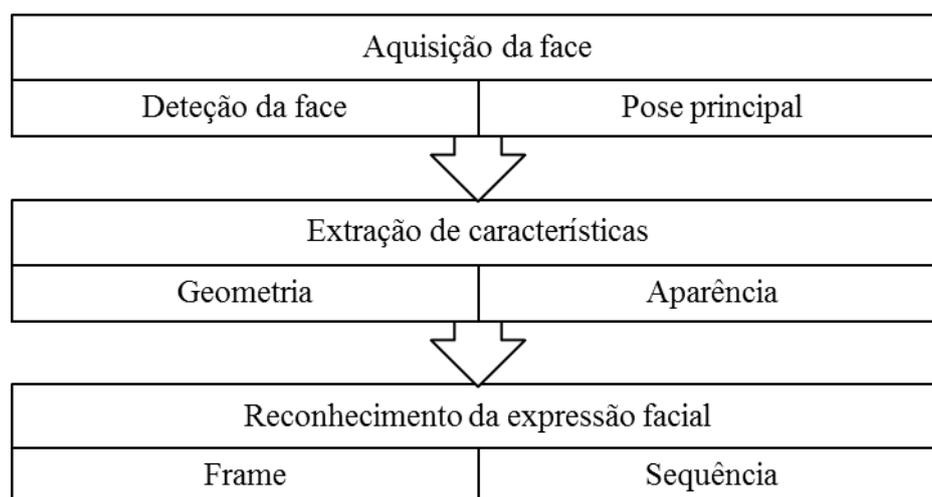


Figura 5 - Estrutura básica da análise do reconhecimento de expressões faciais (adaptado de (Sumathi, Santhanam, & Mahadevi, 2012).

O reconhecimento da expressão facial pode ser feito por duas maneiras: por *frames* e por *sequência*.

2.4.1 Frame

Utilizando *frames*, a imagem de entrada é uma imagem estática e é tratada de forma independente. Existem diversos métodos para o reconhecimento de expressão facial por *frame*: redes neurais, vetores de suporte, análise discriminante linear e regra baseada em classificadores (Sumathi, Santhanam, & Mahadevi, 2012).

Deng et al. (2005) desenvolveram um sistema de reconhecimento de expressões faciais usando filtros locais de Gabor. Foram feitos dois estágios de compressão com recurso ao método de PCA (“principal component analysis”) e LDA (“linear discriminant analysis”) para seleccionar e comprimir o recurso de Gabor. Numa fase inicial é feito o pré-processamento das imagens, onde estas ficam com uma intensidade normalizada, tamanho e forma uniforme. De seguida é feita a extração de características: deteção dos olhos, boca e nariz, rotação para alinhar as coordenadas do olho e localizar e cortar a região do rosto de acordo com o modelo de face (Figura 6). Utilizaram os filtros de Gabor pois estes têm propriedades como localização ideal e análise no domínio de frequência que contribuem muito no reconhecimento de expressões faciais. Para reduzir a dimensão das imagens foram utilizados os métodos PCA (projeção que melhor representa os dados originais em mínimos quadrados) e LDA (projeção que melhor separa os dados em mínimos quadrados). Os resultados foram positivos e melhores comparados com o método Gabor tradicional (Deng, Jin, Zhen, & Huang, 2005).

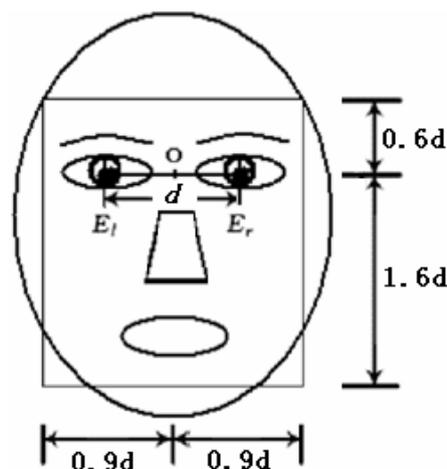


Figura 6 - Modelo facial humano.

Aleksic e Katsaggelos (2006) criaram um método de reconhecimento de expressões faciais que utiliza HMM (“Hidden Markov Model”). Este sistema utiliza também Parâmetros faciais de animação (FAPs) para controlar o movimento das sobrancelhas e dos lábios. Este

método recebe o vídeo de entrada, utiliza as FAPs para dar informação acerca do movimento das sobrancelhas e dos lábios e de seguida utiliza o método HMM para fazer o reconhecimento automático da expressão facial. Como resultados, este método reduziu os erros de reconhecimento de expressões faciais em 44% relativamente ao método HMM normal (Aleksic & Katsaggelos, 2006).

Bartlett et al. (2006) utilizaram as FACs (reconhecimento de ações faciais automatizado) para criar um sistema de deteção de expressões faciais espontâneas independentemente do utilizador. O sistema deteta automaticamente os rostos frontais no vídeo. Com diferentes iluminações e fundos, o sistema tem uma taxa de sucesso de 90%, quando a iluminação é controlada e não tem fundo, a taxa de sucesso sobe. O método opera a 24 *frames* por segundo e as imagens são passadas pelos filtros de Gabor. Os classificadores determinam qual a expressão facial que foi feita (Bartlett, et al., 2006).

Koutlas e Fotiadis (2008) desenvolveram um método que faz o reconhecimento da expressão facial em 3 etapas: utilizando o banco de filtros de Gabor, extração do vetor de características e classificação. Para a classificação usam redes neurais artificiais (RNAs). Os autores usaram também PCA. A desvantagem do método é que certos passos (como marcar pontos fulcrais na face) são feitos manualmente (Koutlas & Fotiadis, 2008).

Thai et al. (2011) desenvolveram um método utilizando Canny (método de deteção de bordas), PCA e redes neurais artificiais para a classificação de expressões faciais. Na primeira etapa é utilizado o Canny para deteção da face. Na segunda etapa, as características faciais são apresentadas com base nas PCA (Figura 7). Para a classificação da expressão facial usam redes neurais artificiais. Este sistema obteve uma taxa de sucesso de 85.7%, enquanto que o método que só utiliza redes neurais artificiais teve 73.3% e o método que usa multi-redes neurais artificiais teve 83%. A maior desvantagem deste método é a fraca deteção do rosto (Thai, Nguyen, & Hai, 2011).

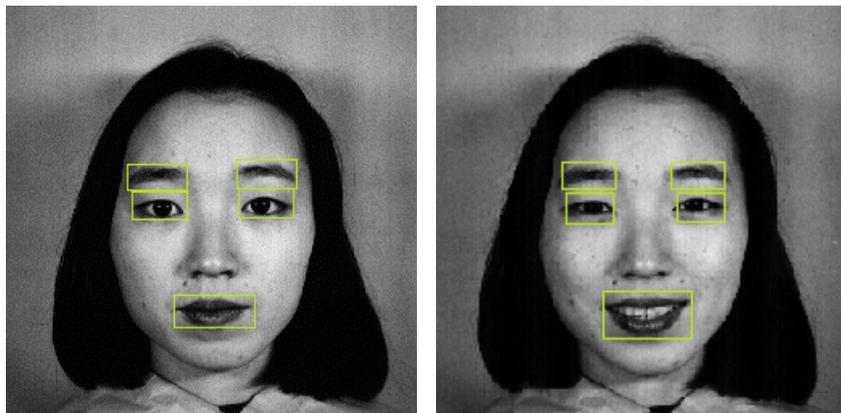


Figura 7 - Resultados obtidos por Thai et al. (2011).

2.4.1 Sequência de imagens

No método com sequências, este utiliza a informação temporal das sequências para reconhecer as expressões faciais (utilizando um ou mais *frames*). Para utilizar a informação temporal podem ser utilizadas diferentes técnicas como redes neurais recorrentes, classificadores de análise de expressões faciais ou HMM (“Hidden Markov Model”) (Sumathi, Santhanam, & Mahadevi, 2012).

Yeasin et al. (2006) desenvolveram um sistema de reconhecimento de expressões faciais que se baseia em vetores do movimento facial obtidos através de sequências de vídeo. Numa primeira etapa é aplicado um banco de classificação linear em vetores de fluxo ótico para que seja criada uma “assinatura” característica de cada expressão facial. Estas “assinaturas” foram usadas para treinar os modelos de Markov (HMM). O sistema utiliza a base de dados de expressões faciais “Cohn-Kanade” e consiste em 488 sequências de vídeo. A taxa de sucesso foi de 90,9%. Os autores pretendem que o próximo passo seja reconhecer também os movimentos labiais enquanto uma pessoa fala pois também são determinantes para a classificação das expressões faciais (Yeasin, Bulot, & Sharma, 2006).

Bartlett et al. (2005) testaram diferentes técnicas para determinar qual a melhor para o reconhecimento automático de expressões faciais. Experimentaram técnicas de seleção de recursos, como por exemplo o AdaBoost para a seleção antes da classificação por SVM (Support Vector Machines) ou LDA. Os melhores resultados foram obtidos utilizando filtros de Gabor, AdaBoost e classificação por SVM. Este sistema faz o reconhecimento em tempo real e teve uma taxa de sucesso de 93%. As saídas dos classificadores alteram suavemente como uma função no tempo e assim é possível medir a dinâmica da expressão facial. Estes autores adicionaram as FACS ao sistema desenvolvido e assim a taxa de sucesso subiu para 94,8% (Bartlett, et al., 2006).

Littlewort et al. (2006) desenvolveram um sistema experimentando várias ferramentas, no entanto o que teve melhores resultados foi aquele que usou filtros de Gabor, AdaBoost e por fim o SVM (93% de taxa de sucesso). Este sistema fornece códigos de expressão facial em 24 *frames* por segundo. Em tempo real opera com resoluções até 16 pixels no olho. O problema da classificação das expressões faciais pode ser resolvido com sistemas lineares, estes têm uma perda de desempenho, mas são mais rápidos em aplicações que atuam em tempo real (Littlewort, Bartlett, Fasel, Susskind, & Movellan, 2006).

Tong et al. (2007) desenvolveram um método que utiliza as *wavelets* de Gabor e níveis de cinza para uma representação eficaz e eficiente dos pontos característicos (Figura 8). Neste método é utilizada uma hierarquia de rosto para melhorar a precisão e robustez da técnica. Permite caracterizar simultaneamente as limitações do rosto e os detalhes das características locais. Como faz o reconhecimento em vídeo, as posições dinâmicas dos pontos são localizados por métodos multimodais (PCA). Neste sistema é ignorada a relação entre as diferentes componentes do rosto, sendo tratadas como independentes para evitar erros que frequentemente acontecem quando é feito o reconhecimento em vídeo. No entanto, estes autores consideram esta “falha” como trabalho futuro (Tong, Wang, Zhu, & Ji, 2007).



Figura 8 - Pontos característicos do método de Tong et al. (2007).

Kotsia e Pitas (2007) criaram dois novos métodos para reconhecimento de expressões faciais em sequência de imagens. O utilizador tem de colocar alguns marcos de Candide para que o método veja o deslocamento entre estes no evoluir da expressão. No final desta evolução, a expressão vai corresponder à sua maior intensidade e consecutivamente é classificada. Num dos métodos, a diferença entre as posições dos pontos de Candide vão ser dados de entrada para o classificador SVM. No outro método são usadas unidades de ação facial (FAUs) para a classificação da expressão. A base de dados usada para ambos os

métodos foi a Cohn-Kanade. Para o sistema que usa o SVM a taxa de sucesso é de 99,7%, sendo este método com a taxa mais alta de sucesso dos métodos de reconhecimento de expressões faciais (com a base de dados de Cohn-Kanade). No outro método é de 95,1% (Kotsia & Pitas, 2007).

Varona et al. (2008) desenvolveram um projeto que permitia que pessoas com deficiência motora pudessem aceder ao computador através das expressões faciais (Figura 9). Este sistema não necessita de calibração (para não ser incómodo para os utilizadores) e deteta automaticamente a face através do algoritmo Viola e Jones. De seguida, o método divide o rosto em três regiões: olhos e sobrancelhas, boca e nariz. Nestas regiões são gerados pontos que as localizam. Estes pontos nem sempre são detetados corretamente devido à variância da iluminação. Para a deteção do nariz foi utilizada a base de dados BioID devido à resolução das imagens e da aquisição destas serem idênticas a este sistema. É utilizado um método Gaussiano 3D para detetar a região da cor da pele. Para determinar a região de olhos e sobrancelhas é feita a binarização da imagem. Esta parte pode falhar para as pessoas que usam óculos. De seguida é feita a localização da retina para depois usar no controlo do rato pelo olhar. Neste caso, o único gesto facial a ter em conta é o piscar do olho. O movimento do rato é feito pela posição do nariz e a piscadela dos olhos pode ter diversas funções. Assim pessoas sem movimentos nos braços ou sem qualquer mobilidade no tronco conseguem controlar o computador (Varona, Manresa-Yee, & Perales, 2008).

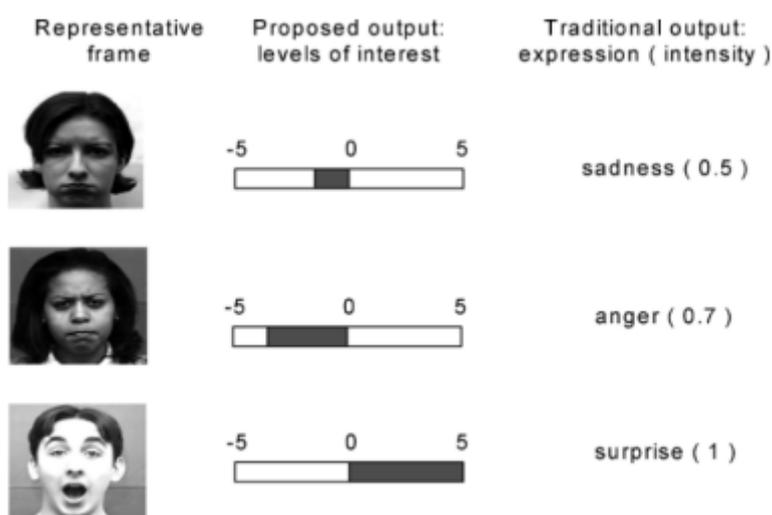


Figura 9 - Funcionamento do método de Varona et al. (2008).

Yang et al. (2009) criaram um método que faz o reconhecimento de expressões faciais através de características dinâmicas codificadas. Este método baseia-se em três etapas: extração de características dinâmicas, codificação destas características e a aprendizagem com o AdaBoost. Para a primeira etapa, os autores utilizaram as características de Haar para captar as variações temporais das expressões. Para a codificação foi utilizado o padrão binário de codificação e os autores analisaram a distribuição de cada característica dinâmica e criaram um código para cada uma delas. Estas ainda são mapeadas para os recursos do padrão binário. No final, o AdaBoost é utilizado para aprender um conjunto de recursos codificados discriminantes para o reconhecimento da expressão facial. Os autores consideram que o sistema tem um desempenho promissor (Yang, Liu, & Metaxas, 2009).

Lau (2010) desenvolveu um método de detecção de expressões faciais para pessoas com deficiência. É utilizado o algoritmo Viola e Jones para detetar o rosto frontal. As características Haar são usadas para detetar as características do rosto e a base de dados utilizada foi a AdaBoost. Como classificador é utilizado o LDA e este usa *Eigenfaces*. Este sistema funciona tanto de dia como à noite. Inicialmente é feito um processo de treino. O sistema permite a detecção de expressões faciais de forma automática ou manual. Cada utilizador pode criar o seu próprio perfil. Foram feitos testes com 10 pessoas e a taxa de sucesso do sistema foi de 89% (Lau, 2010).

Nagarajan et al. (2011) fizeram um método de reconhecimento de expressões faciais para pessoas com deficiência motora. Inicialmente é feita a aquisição de dados através de vídeo. São colocados autocolantes refletores em pontos específicos no rosto do utilizador. De seguida é pedido à pessoa para realizar 8 expressões faciais específicas (cada expressão tem de ser feita 10 vezes). O *software* extrai as coordenadas dos autocolantes em cada *frame* (são cerca de 100 *frames*). Este método utiliza Transformada de Wavelet discreta (devido à variação temporal da informação) para extrair as características. É calculado o desvio padrão para cada wavelet, criando assim vetores característicos para a classificação. São usados três classificadores diferentes: Redes Neurais Artificiais, “k-Nearest Neighborhood” e LDA. São classificadas 8 expressões faciais. O classificador que obteve o melhor resultado foi o LDA com uma taxa de sucesso de 99,43% (Nagarajan, Hariharan, & Satiyan, 2011).

Jamshidnezhad e Nordin (2011) criaram um método que é baseado em regras *fuzzy* do tipo *Mamdani*. A base do sistema tem duas componentes: regras *fuzzy* e algoritmo genético. O sistema classifica os vetores de características de entrada numa das seis expressões básicas.

Para mapear os vetores de entrada é utilizado a função *Gaussian* (pertence à função *fuzzy*). O algoritmo genético é dividido em diferentes etapas: definir o comprimento do cromossoma, função de seleção, cruzamento, mutação, reprodução e condições de encerramento. Os autores decidiram utilizar estas duas ferramentas para reduzir o grau de complexidade e melhorar o desempenho do sistema (Jamshidnezhad & Nordin, 2011).

Vasanthan et al. (2012) criaram um sistema para controlar o rato do computador com expressões faciais. Inicialmente são colados quatro autocolantes: um entre as sobrancelhas e os outros no contorno inferior do lábio. Faz o reconhecimento de cinco expressões faciais: movimento da bochecha para esquerda e direita, movimento para cima e para baixo da sobrancelha e abrir a boca. O movimento da bochecha equivale ao movimento horizontal do cursor, o movimento da sobrancelha ao movimento vertical e o clique é abrir a boca. A posição inicial marca as coordenadas iniciais, depois de realizada a expressão facial, são marcadas as coordenadas finais e a diferença entre as iniciais e as finais dão origem à expressão facial e conseqüentemente à ação no rato. Este sistema tem algumas desvantagens como a sensibilidade à iluminação, ter de usar autocolantes na cara é intrusivo para o utilizador e a resposta lenta à ação feita (Vasanthan, Nagarajan, Murugappan, & Ilias, 2012).

Dongre e Patil (2015) criaram um *software* para auxiliar o controlo do rato para pessoas com deficiência. O método utiliza a *webcam* do computador. O algoritmo utilizado é o Clustering. Depois de detetada a face e todas as componentes faciais, é binarizada a imagem. Por fim cada expressão facial é transformada numa ação do rato. O cursor pode ser movido através do olhar (Dongre & Patil, 2015).

2.5 Conclusão

Durante a nossa pesquisa e com o nosso conhecimento não foram encontrados estudos de caso com pessoas com deficiência. Processamento digital de imagem e visão por computador são áreas muito vastas, sendo o reconhecimento de expressões faciais uma área com alguns estudos e em desenvolvimento constante. Em alguns estudos foi dito que os métodos utilizados poderiam usados por pessoas com deficiência, no entanto não apresentam testes. Questiona-se assim se esses programas são realmente possíveis de usar por pessoas que apresentem algumas limitações.

Para dar credibilidade ao nosso trabalho desenvolvido na presente dissertação, além do desenvolvimento do sistema, foram feitos testes para saber se o sistema é eficaz na sua utilização com pessoas com deficiência.

3. Algoritmo proposto para reconhecimento de expressões faciais

O objetivo desta dissertação é o desenvolvimento de uma ferramenta que permita a interação com o computador através de expressões faciais. Este tipo de interação é fundamental para quem não possui mobilidade ou membros, tornando a interação com o computador possível. O reconhecimento automático das expressões faciais é efetuado com base num algoritmo que processa as imagens adquiridas por uma câmara apontada à face do utilizador. A Figura 10 apresenta o diagrama do algoritmo de processamento digital de imagem proposto nesta dissertação.

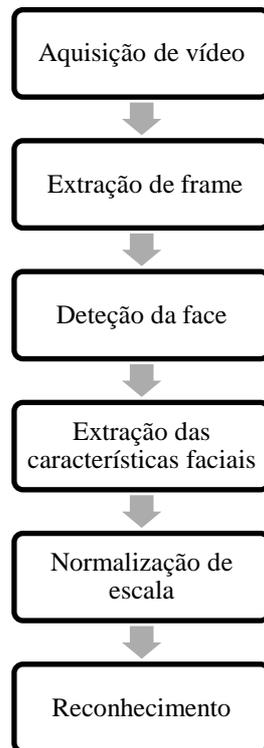


Figura 10 - Diagrama do sistema de reconhecimento de expressões faciais.

Inicialmente é feita uma aquisição de um vídeo e de seguida a extração de cada *frame* do vídeo. Cada *frame*, representado em RGB é analisado de forma a extrair as características faciais usadas para o reconhecimento da expressão facial.

3.1 Detecção da face

Para a deteção da face no *frame* utilizou-se o método de Viola e Jones. O método de Viola e Jones (2001) constitui um processo de deteção, apto para o processamento de imagens de uma forma rápida e eficaz. Sendo um dos métodos mais utilizados para detetar as faces humanas (figura 13), inclui uma nova representação da imagem, designada imagem integral. Esta imagem permite que as características utilizadas pelo detetor sejam processadas uma única vez na imagem. Tem um classificador simples e eficiente. De um grupo muito grande

de potenciais características, este método escolhe as características visuais críticas. Esta técnica baseia-se nas características de Haar (Figura 11). O valor de uma função de dois retângulos é a diferença entre a soma dos pixels dentro de duas regiões retangulares. As regiões têm o mesmo tamanho e forma e são horizontalmente ou verticalmente adjacentes. Uma característica com três retângulos é calculada através da soma dentro de dois retângulos sendo subtraída a soma do retângulo do centro. Num recurso com quatro retângulos é calculada a diferença entre os pares diagonais dos retângulos. Então, os resultados dos recursos acima descritos irão representar o valor encontrado pela característica para determinada região (Viola & Jones, 2001).

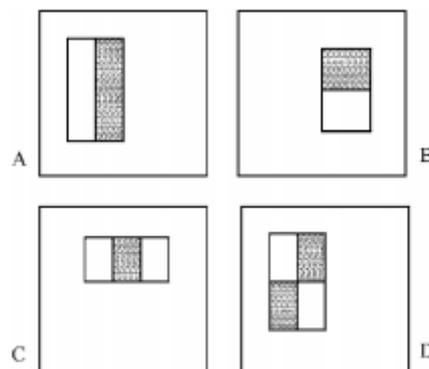


Figura 11 - Exemplos dos recursos de Haar. A imagem A é uma divisão horizontal, a B é uma divisão vertical, a C tem 2 divisões horizontais e a D é criada por divisões horizontais e verticais (Viola & Jones, 2001).

A quantidade de possíveis combinações das características de Haar é imensa, então é necessário obter apenas um conjunto de características mais importantes. Então Viola e Jones escolheram o método de AdaBoost para facilitar o processo de classificação. O AdaBoost é um método que usa a combinação de vários classificadores fracos para conseguir uma classificação forte. É utilizado para selecionar um conjunto de características bem como para treinar o classificador. Durante a fase de treino, os retângulos são localizados e analisados para ver se são úteis ou não ao classificador (Figura 12) (Santanas, Gomes, & Santos, 2015).

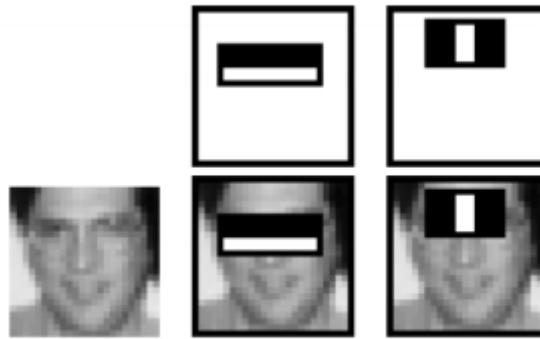


Figura 12 - As duas melhores características aplicadas durante a fase de treino (Viola & Jones, 2001).

Uma outra contribuição está presente num método para ajustar os classificadores numa cascata, que possibilita descartar o fundo da imagem rapidamente, tendo como foco apenas as regiões que realmente interessam. Esta contribuição aumenta muito a velocidade do detetor. Os autores deste método traçam dois tipos de classificadores. O primeiro classificador é de uma única camada. Este não rejeita nenhuma subjanela durante toda a classificação, ou seja, todas as subjanelas são processadas por todas as características escolhidas pelo algoritmo. A cascata completa tem 32 classificadores (Viola & Jones, 2001).



Figura 13 - Resultados do método desenvolvido por Viola e Jones (Viola & Jones, 2001).

3.2 Extração das características faciais

Após detecção da face, o algoritmo proposto extrai, utilizando também o algoritmo Viola e Jones, características faciais (detecção da boca e olhos). Para a detecção da boca e dos olhos, é parametrizado o modelo de classificação correspondente no algoritmo. Relativamente aos olhos, apenas é feita a detecção do olho esquerdo para melhorar a performance do sistema.

3.3 Normalização de escala

Depois da detecção e extração das características faciais tem de ser feita uma normalização da escala. A dimensão das imagens correspondentes às características faciais é diferente das imagens de treino inseridas na base de dados. Para corrigir isso é necessário normalizar a dimensão das imagens. A dimensão do olho esquerdo é de 92x112 pixéis e a da boca é de 145x92 pixéis. Estas dimensões foram definidas após vários testes com diferentes escalas e determinou-se que estas seriam as dimensões ideais para o sistema funcionar o melhor possível.

3.4 Reconhecimento

Para a identificação da expressão facial foi usado o método de análise de componentes principais (PCA). Esta técnica é uma técnica baseada na aparência. Neste tipo de técnica é mantida toda a informação da característica facial na sua extração, rejeitando assim informação redundante (Bakshi & Singhal, 2014).

Este método foi apresentado por Turk e Pentland (1997) e faz uma análise das características principais da imagem para fazer a extração dos atributos. O objetivo deste algoritmo é descobrir as características principais da distribuição das faces ou os auto-vetores da matriz de covariância dos conjuntos das faces (uma imagem é um ponto ou um vetor). As componentes principais são determinadas em ordem decrescente de importância. A primeira componente possui mais informação do que a segunda e assim sucessivamente. O PCA tenta construir um pequeno conjunto de componentes que resumem os dados originais, reduzindo a dimensão dos mesmos, preservando os componentes mais significantes (Silva V. A., 2008).

Para haver este reconhecimento da expressão facial, o programa segue os seguintes passos:

1. Calcula a média das imagens de entrada (imagens detetadas);
2. Subtrai a média das imagens de treino (exemplo da Figura 14) a média anteriormente calculada;

3. Calcula os *eigenvectors* e os *eigenvalues*;
4. Retém apenas os 10 *eigenectores* com o maior número de *eigenvalues* (isto é, as componentes principais);
5. Calcula a distância Euclidiana entre os vetores característicos e os das imagens de treino. (Silva V. A., 2008).



Figura 14 - Exemplo de média das imagens de treino. As imagens acima são imagens de treino, sendo que a debaixo é a média das imagens acima. É um exemplo como funciona o PCA.

Consegue-se determinar se existe expressão ou não devido à distância Euclidiana dos vetores característicos e os de recurso correspondentes. Quanto menor for esta distância, mais semelhantes são (como se pode ver na Figura 15). Ou seja, não havendo a expressão facial (igual à detetada) na base de dados, os valores da distância vão ser muito grandes. Isto funciona tanto na boca como nos olhos.

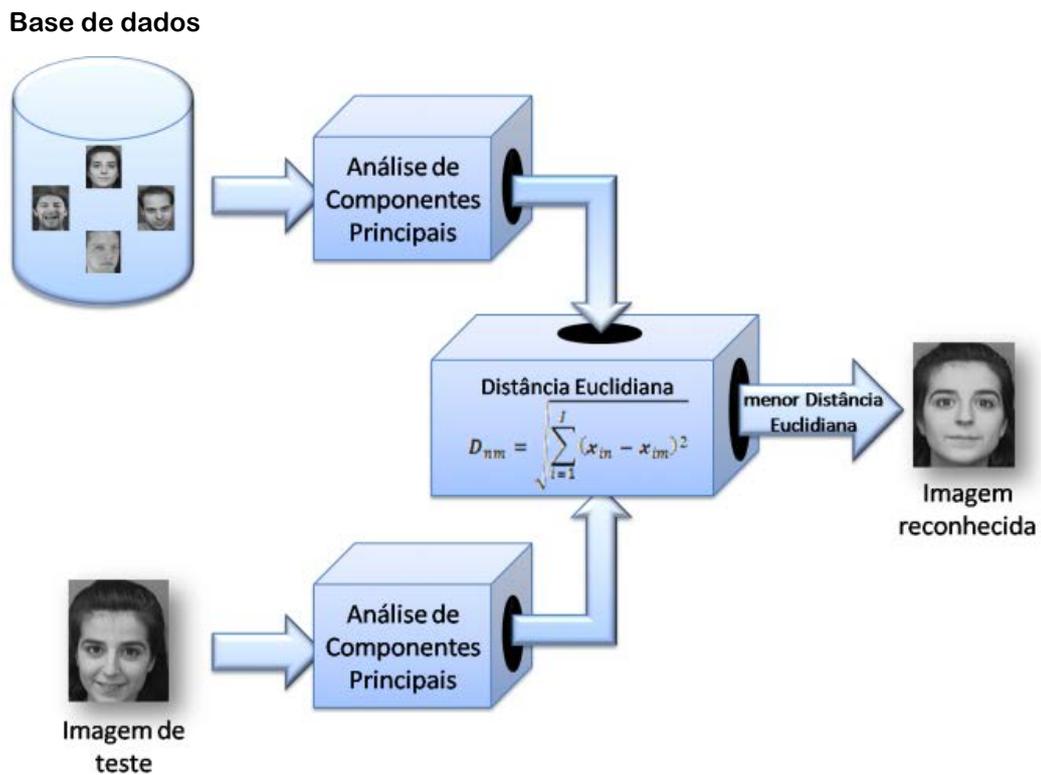


Figura 15 - Funcionamento de um exemplo baseado em análise de componentes principais.

3.5 Conclusão

Neste capítulo foi apresentado o algoritmo relativo ao sistema de reconhecimento de expressões faciais. É exposto o diagrama geral do funcionamento do algoritmo seguido da explicação de cada etapa. É também explicado com algum detalhe o método do Viola e Jones na fase da detecção da face e o método PCA na etapa do reconhecimento da expressão facial.

4. Implementação

Com este trabalho pretende-se criar e desenvolver de uma ferramenta que permita detetar expressões faciais para que futuramente possa ter alguma ação no computador de acordo com a expressão feita. Para isso o sistema deverá detetar as diferentes partes da cara (face, olhos e boca) e posteriormente classificar a expressão. Para a criação deste projeto foi utilizada a *framework* Matlab (versão 2015) e os testes foram feitos em imagens extraídas de vídeos.

4.1 Estudo de caso

O sistema de reconhecimento de expressões faciais foi especificamente treinado e adaptado para ser testado num estudo de caso, de uma pessoa com deficiência. A pessoa do estudo é do sexo feminino, com 14 anos e tem uma doença neuromuscular degenerativa que lhe afeta os músculos, ficando sem qualquer tipo de mobilidade nos membros. Os únicos músculos com mobilidade são os da face. Possui um ventilador para respirar. No dia-a-dia, frequenta aulas de ensino especial, terapia da fala, terapia ocupacional, fisioterapia e psicomotricidade. Nas aulas e nas terapias, numa parte do tempo tem acesso ao computador. Nas aulas são feitos jogos didáticos sobre diversos temas de maneira a que ela possa aprender mais conceitos. Na terapia da fala é treinada a parte da comunicação. É utilizado o computador para ela usar quadros de comunicação com o olhar, no entanto esta interação não é viável pois a terapeuta supõe para onde ela olha e nem sempre é o correto. Na terapia ocupacional são feitos alguns jogos que trabalhem a interação da criança. No entanto como ela não consegue interagir com o computador todas estas atividades tornam-se monótonas e difíceis de realizar, não conseguindo cumprir com o plano de estudos. Por esta razão, utilizou-se este estudo de caso pois a criança não tem qualquer solução para interagir com o computador. Esta rejeita o Eye Tracker por saber que é algo que está a controlar. Então achou-se que o reconhecimento de expressões faciais podia funcionar trabalhando em segundo plano sem ela se aperceber. A professora ou terapeuta que estivesse com ela fazia-lhe perguntas e ela respondia através da expressão facial, causando no computador uma ação de causa-efeito. As expressões faciais que são executadas de uma forma simples e que se distinguem bastante bem são o sorriso e a mandar um beijo. Serão estas expressões que vão ser utilizadas neste trabalho. As funções que posteriormente irão ser executadas por cada expressão facial serão duas funções do rato, sendo que a principal será o clique esquerdo do rato. Esta decisão foi tomada de acordo com as atividades que a pessoa faz durante as aulas.

4.2 Implementação do algoritmo de Reconhecimento

Logo que é iniciada a aplicação, o sistema começa a analisar um vídeo *frame a frame*. Assim que começa a análise de um *frame* é detetada a cara da pessoa. Utilizando o algoritmo Viola-Jones é necessário usar o detetor *vision.CascadeObjectDetector*. A propriedade “modelo de classificação” controla o tipo de objeto a detetar. Por omissão, o detetor é configurado para detetar faces (MathWorks, 2016).

Para a face da pessoa utilizou-se o modelo de classificação: *FrontalFaceCART*. Este deteta os rostos que estão na posição vertical e frontal (Figura 16). Este modelo baseia-se na análise da árvore de classificação e regressão (CART). Esses classificadores utilizam recursos Haar para codificar características faciais. Os classificadores baseados em CART têm a capacidade de modelar dependências de ordem mais elevada entre as características faciais (MathWorks, 2016).

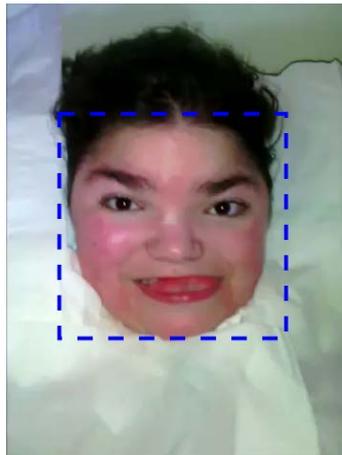


Figura 16 - Detecção da face.

De seguida é detetada a boca. Para isso utilizou-se o modelo de classificação parametrizando como “Mouth” que deteta a boca (Figura 17). São utilizados os recursos de Haar para detalhar a forma da boca (MathWorks, 2016).



Figura 17 - Detecção da boca.

Neste estudo de caso, a pessoa tem sempre a língua de fora e por vezes o algoritmo confunde-se (Figura 18).



Figura 18 - Erro na detecção da boca.

Para contornar esta dificuldade e melhorar a performance do algoritmo, antes da deteção da boca divide-se a imagem da cara em duas partes e utiliza-se a parte inferior (Figura 19).



Figura 19 - Detecção correta da boca a sorrir e a mandar beijinho.

A imagem é normalizada (como foi referido anteriormente).

Para a deteção dos olhos, inicialmente, o modelo de classificação foi o “EyePairSmall” que deteta os olhos numa imagem pequena (Figura 20).

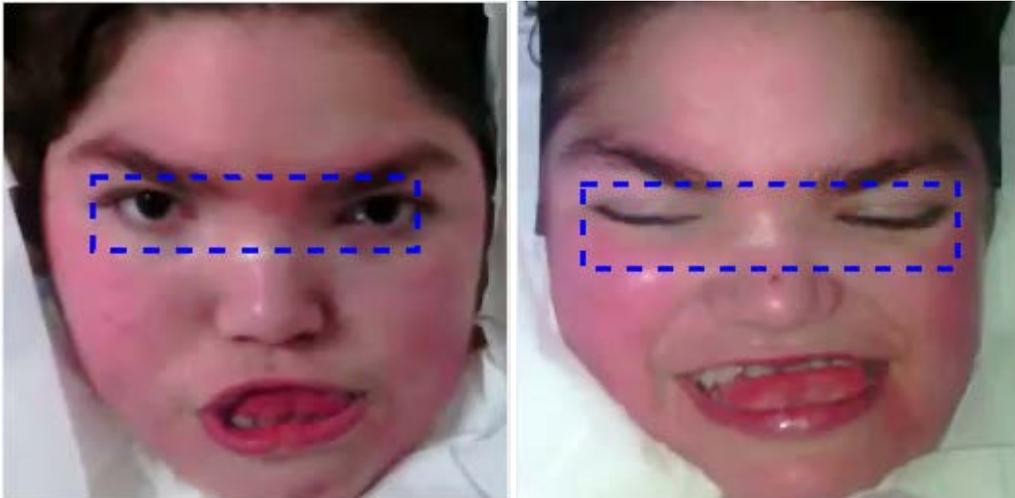


Figura 20 - Deteção dos olhos a mandar beijinho e a sorrir.

Para uma melhor performance do algoritmo optou-se por detetar apenas um olho. Foi escolhido o olho esquerdo e para este ser detetado corretamente, divide-se a imagem verticalmente em duas, analisando só a parte esquerda. Para a deteção do olho utiliza-se como parâmetro no modelo de classificação “LeftEyeCART”. (Figura 21)

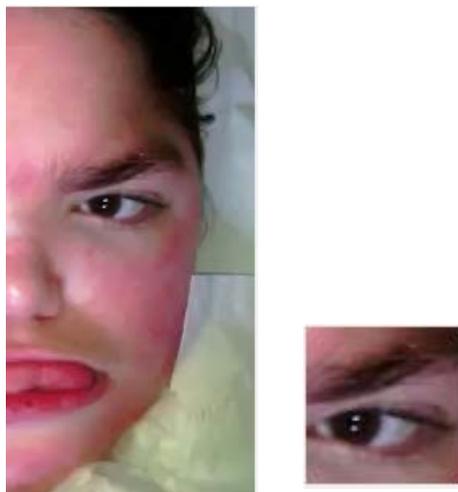


Figura 21 - Deteção do olho esquerdo.

O restante procedimento é igual ao da boca.

Tal como foi referido acima para a identificação da expressão facial foi usado o método baseado em PCA. Foi necessário criar uma base de dados com imagens da boca e dos olhos nas diferentes expressões (*sorrir* e *beijinho*). Foram escolhidas estas duas expressões pois são as expressões mais distintas que a pessoa do estudo de caso faz e as mais intuitivas para ela. Existem duas bases de dados, uma para a boca e outra para o olho esquerdo. Cada base de dados contém seis imagens de treino para cada expressão. Nestas seis imagens algumas delas são repetidas entre si, havendo outras diferentes. Isto porque apesar da expressão ser a mesma, a forma da boca ou dos olhos nem sempre são iguais. Estas imagens de treino servem para indicar ao programa quais as imagens que ele deve comparar com a imagem que detetar. Estas imagens são em tons de cinza, em formato “.pgm” e cada conjunto de cada base de dados tem uma dimensão diferente (devido à diferença entre as formas das diferentes partes da face). Estas imagens são extraídas de *frames* de vídeos diferentes. É feita a deteção da boca/olho com o algoritmo Viola e Jones e é mostrado todos os *frames*. Depois é feita uma seleção (se possível com a boca/olho de diferentes formas na mesma expressão) e com a “Ferramenta de Recorte” do Windows é recortada e gravada. Depois é colocada no tamanho e formato correto, colocando-se depois seis destas imagens na base de dados referente a cada expressão facial.

Depois de detetada a boca ou os olhos aparece um menu como se pode ver na Figura 22. Em primeiro lugar é necessário carregar a base de dados (1ª opção). É necessário fazer sempre este passo porque pode ter sido alterada alguma imagem na base de dados e assim garante que esta está sempre atualizada. De seguida é feito o segundo passo que é para detetar a expressão facial, começando aqui o método baseado em PCA.

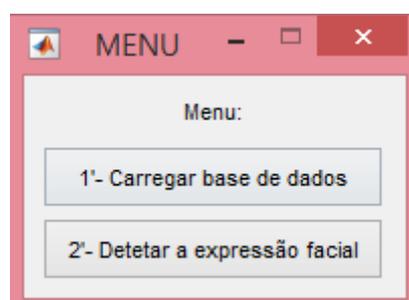


Figura 22 - Menu do algoritmo.

No que toca à boca pode haver duas situações: quando a expressão é encontrada (Figura 23) e quando esta não é semelhante a nenhuma das expressões da base de dados (Figura 24).

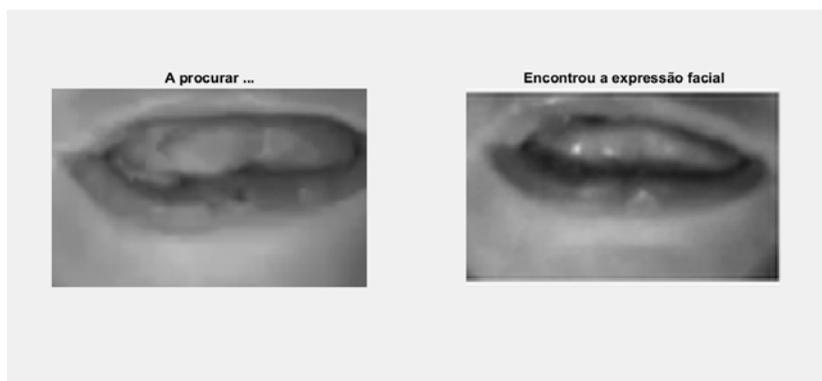


Figura 23 - Reconhecimento da expressão facial correto.

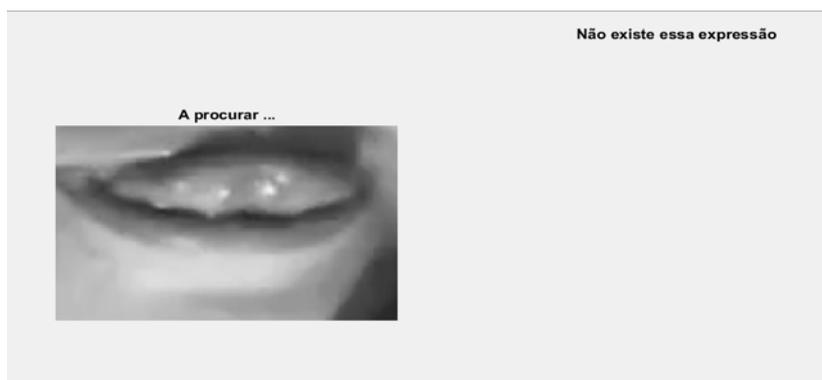


Figura 24 - Quando a expressão não é idêntica com nenhuma das expressões da base de dados.

Na figura acima a boca está na sua forma neutra (ou seja sem nenhuma expressão). Não havendo expressão facial sorrir ou beijinho, os valores da distância Euclidiana vão ser muito grandes. Colocou-se então um limite para que acima desse valor tudo o que for detetado não é expressão facial.

Para a deteção da expressão facial através dos olhos funciona de maneira análoga (Figura 25). No entanto não foi colocado o limite acima referido pois os olhos da pessoa com uma expressão neutra são iguais à forma dos olhos quando esta faz beijinho. Os olhos servem apenas de complemento à boca, pois esta é a característica mais distintiva na classificação da expressão facial.



Figura 25 - Reconhecimento da expressão facial através do olho esquerdo.

Apresenta-se por fim a Figura 26 que exemplifica como o protótipo funciona numa situação de teste.

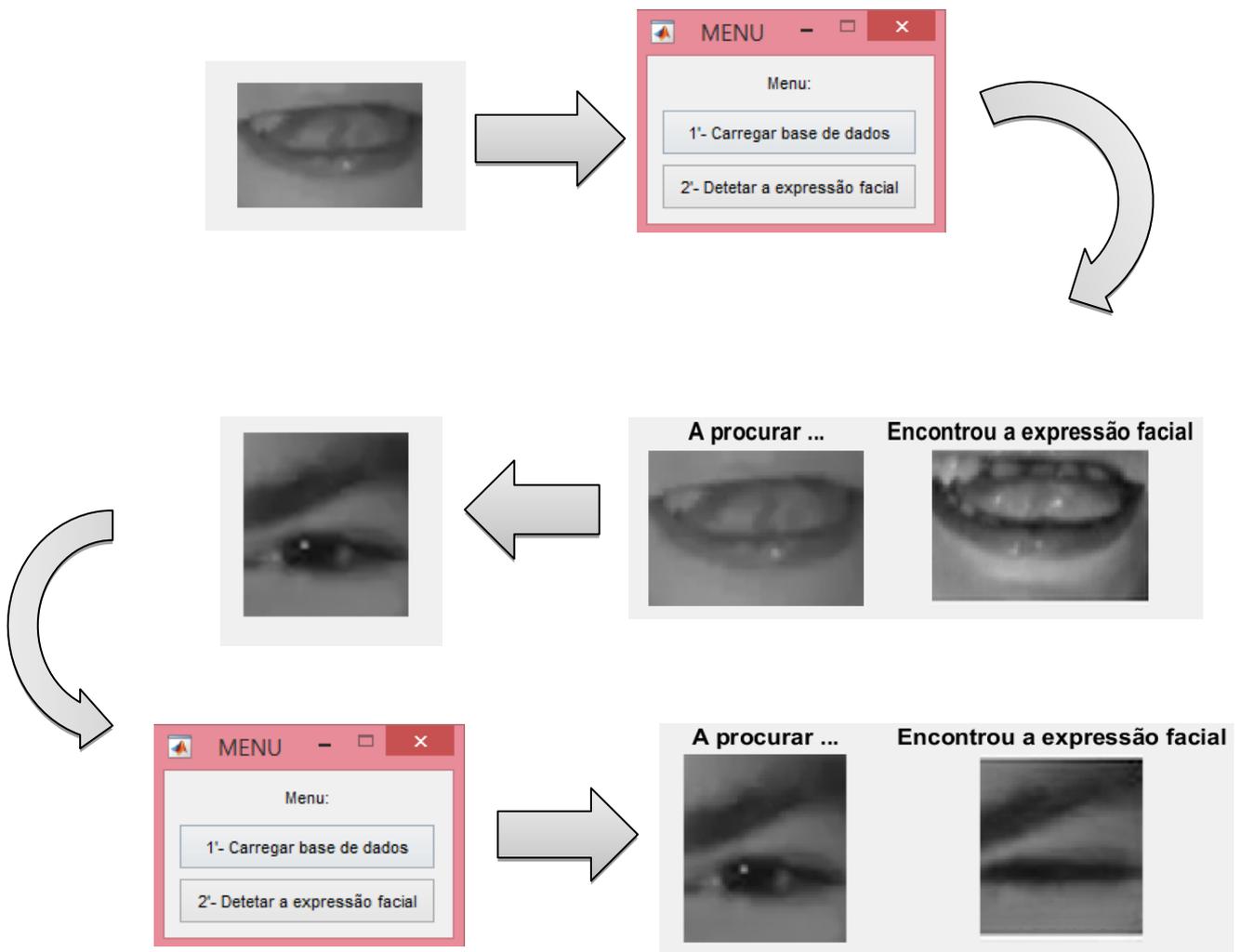


Figura 26 - Diagrama do funcionamento do algoritmo desenvolvido.

4.3 Implementação da aplicação para interação com o computador

Depois de desenvolvido o algoritmo de reconhecimento de expressões faciais foi necessário criar a aplicação que fizesse a interação com o computador. Esta aplicação funciona em tempo real, isto é, os vídeos foram substituídos pela imagem em tempo real captada pela *webcam* do computador. Inicialmente é feita uma fase de treino para construir as bases de dados com imagens das diferentes classes de expressões. A pessoa tem de fazer 3 vezes a mesma expressão para garantir que a base de dados fica completa e com imagens variadas (da mesma expressão). Segue-se depois a fase de funcionamento da aplicação. Esta vai analisando os *frames* que recolhe da *webcam* com o algoritmo acima explicado. No caso de não detetar corretamente a face ou a boca, o sistema ignora e passa para o *frame* seguinte. No caso de a pessoa fazer a expressão correspondente à expressão da base de dados é feito o clique esquerdo do rato. Escolheu-se esta função por ser a mais utilizada no dia-a-dia. Neste momento a aplicação ainda só está preparada para uma expressão facial. Para fechar a aplicação basta carregar em qualquer tecla do teclado.

4.4 Conclusão

Neste capítulo foi apresentado o algoritmo desenvolvido nesta dissertação bem como a aplicação que faz a interação com o computador. O algoritmo baseia-se nos métodos de Viola e Jones e do PCA. A aplicação permite que seja feita a interação do utilizador com o computador através do algoritmo desenvolvido.

5. Avaliação do algoritmo de reconhecimento expressões faciais

Neste capítulo é apresentada a avaliação dos resultados do sistema de reconhecimento de expressões faciais, nomeadamente a avaliação do detetor das regiões faciais e o classificador de expressões faciais.

Na avaliação do algoritmo apresentam-se resultados obtidos no estudo de caso, já descrito no capítulo 4, e os resultados obtidos em cinco pessoas sem deficiência. Por fim haverá uma comparação entre os resultados do estudo de caso e das pessoas sem deficiência.

5.1 Procedimento

Para cada teste foi necessário haver algumas filmagens das pessoas a fazerem as diferentes expressões faciais. Nestas filmagens utilizou-se o telemóvel “Nos Zilo”. Em cada filmagem tentou-se com que a iluminação fosse constante e a ideal. Pediu-se às pessoas que fizessem três expressões: sorrir, mandar beijo e expressão neutra. Devido ao convívio com a pessoa do estudo de caso, concluiu-se que o beijinho e o sorrir eram as expressões mais diferenciais que pessoa faz no dia-a-dia. A expressão neutra representa todas as outras expressões que não sejam sorrir e beijinho. Foi utilizada esta expressão como exemplo, para demonstrar que o sistema exclui qualquer expressão que não seja as referidas anteriormente. Estas expressões teriam que ser repetidas pelo menos cinco vezes (pois iriam ser utilizados cinco vídeos de cada expressão na avaliação do protótipo). De seguida os vídeos eram cortados por expressão pelo programa “Movie Maker” do Windows. Cada vídeo ficou com a duração de aproximadamente um segundo, correspondendo a 30 *frames*. Foi acordada esta duração pois neste tempo é possível definir uma expressão facial. De seguida eram construídas as bases de dados para cada pessoa. Decidiu-se fazer uma base de dados para cada pessoa pois a forma da boca nas expressões variam de pessoa para pessoa e assim garante que o programa funcione melhor. Na pessoa do estudo de caso foi avaliada a boca e olhos. Nas pessoas sem deficiência apenas foi avaliada a boca, pois estas não mostravam diferenças significativas nos olhos nas diferentes expressões. Como a boca é elemento principal numa expressão facial então vai ser esse o elemento de comparação entre o estudo de caso e os outros utilizadores.

5.2 Resultados do estudo de caso

Para testar o protótipo no estudo de caso decidiu-se filmar a pessoa várias vezes em diferentes dias. Foi feito isto para garantir que as condições fossem diferentes de vídeo para vídeo, testando assim este fator. Foi pedido que a pessoa fizesse as 3 expressões: sorrir, mandar beijo e neutra. Analisaram-se sete vídeos de cada expressão. No total analisaram-se 630 *frames*. Depois de o protótipo analisar cada vídeo, registavam-se os resultados. Na Tabela 1 pode-se ver os resultados obtidos relativamente à boca.

Tabela 1 - Resultados da análise das expressões faciais na boca do estudo de caso.

Vídeos	Sorrir	Beijinho	Neutro	Não Identificado	Duração (frames)	Expressão facial
Vídeo 1	27	2	0	1	30	Sorrir
Vídeo 2	28	1	0	1	30	
Vídeo 3	30	0	0	0	30	
Vídeo 4	30	0	0	0	30	
Vídeo 5	29	0	1	0	30	
Vídeo 6	28	2	0	0	30	
Vídeo 7	22	6	0	2	30	
Vídeo 8	0	30	0	0	30	Beijinho
Vídeo 9	0	30	0	0	30	
Vídeo 10	0	29	1	0	30	
Vídeo 11	3	27	0	0	30	
Vídeo 12	2	28	0	0	30	
Vídeo 13	15	15	0	0	30	
Vídeo 14	3	26	0	1	30	Normal
Vídeo 15	0	0	30	0	30	
Vídeo 16	0	0	30	0	30	
Vídeo 17	0	0	30	0	30	
Vídeo 18	0	0	30	0	30	
Vídeo 19	0	2	28	0	30	
Vídeo 20	0	0	21	9	30	
Vídeo 21	0	11	12	7	30	

Pela Figura 27 pode ver-se que existem ocasiões em que o protótipo erra, significa que detetou a expressão errada.

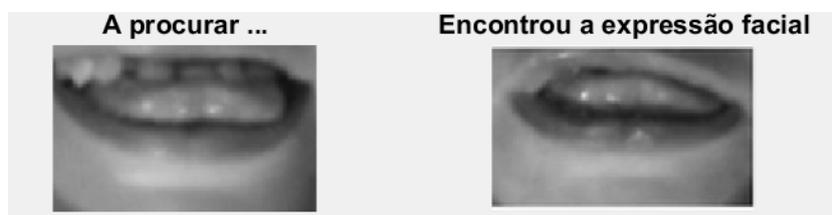


Figura 27- Erro na classificação da expressão facial.

Quando é feita a expressão neutra o que deve acontecer é uma mensagem a dizer que não existe essa expressão. Quando o sistema erra, acontece é que a pessoa pode estar a sorrir ou mandar beijo (Figura 28) e aparecer que não encontra a expressão facial. Isto acontece quando as semelhanças entre o *frame* e as imagens da base de dados são poucas e o valor da distância euclidiana é elevado, ultrapassando o limite dado para excluir todas as expressões que não sejam sorriso e beijinho.



Figura 28 - Erro na classificação da expressão facial: excluiu a expressão quando ela existe.

Como se pode ver pela Figura 29, por vezes que a boca é detetada noutra sítio da face. Então nesses casos é avaliado como “Não Identificado”.



Figura 29 - Erro na deteção da boca.

Estes exemplos anteriormente referidos são os erros do protótipo. Estes erros podem surgir devido a diversos fatores: fraca iluminação, as mesmas expressões mas com formas diferentes (diferentes sorrisos por exemplo) e a cara não estar totalmente frontal. Acontece

também haver erros do algoritmo que não é 100% correto, havendo algumas falhas sem justificção aparente.

Para uma caracterização mais detalhada do desempenho do sistema foi calculada a matriz de confusão (ou classificação) e as métricas precisção, sensibilidade e especificidade.

Numa matriz de confusão, são classificados todos os casos do sistema de reconhecimento em categorias, determinando se o valor previsto correspondeu ao valor real. Serve para analisar os resultados na sua globalidade. A categoria real corresponde a cada linha da matriz sendo que a prevista corresponde a cada coluna. Os elementos diagonais indicam as classificações corretas (Parker, 2001).

A precisção é a proporção de casos positivos que são corretamente identificados. Esta é calculada pela seguinte fórmula:

$$P = \frac{VP}{VP + FP}$$

A sensibilidade é a capacidade do sistema identificar os verdadeiros positivos. A fórmula de cálculo deste parâmetro é:

$$S = \frac{VP}{VP + FN}$$

A especificidade do sistema representa a capacidade que o protótipo tem em identificar os verdadeiros negativos.

$$E = \frac{VN}{VN + FP}$$

A tabela 2 apresenta a matriz de confusão do sistema de reconhecimento para as 3 categorias de expressões faciais, no estudo de caso.

Tabela 2 - Matriz de confusão para o estudo de caso.

Expressões previstas Expressões reais	Sorrir	Beijinho	Neutro	Não Identificado	Total
	(%)	(%)	(%)	(%)	(%)
Sorrir	92,4	5,2	0,5	1,9	100
Beijinho	11	88	0,5	0,5	100
Neutro	0	6,2	86,2	7,6	100

Pela Tabela 2 é possível verificar que a expressão que obteve melhores resultados foi o sorriso com uma percentagem de 92,4, de seguida o beijinho e por fim a expressão neutra. A expressão facial beijinho foi a que obteve maior número de erros na classificação da expressão, isto é, foram classificados 11% dos *frames* como sorriso. A nível de erros de deteção da boca, a expressão neutra foi a que apresentou mais erros (7,6% dos *frames*).

Na Tabela 3 é possível ver o número de *frames* e a percentagem correspondente, de verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN) em cada expressão. É também apresentada a precisão, sensibilidade e especificidade do sistema de reconhecimento.

Tabela 3 - Avaliação dos resultados no estudo de caso.

	VP	VN	FP	FN	Precisão	Sensibilidade	Especificidade
Sorrir	194 (30,8%)	397 (63%)	23 (3,7%)	16 (2,5%)	89,40%	92,40%	94,50%
Beijinho	185 (29,4%)	396 (62,9%)	24 (3,8%)	25 (4%)	88,50%	88,10%	94,30%
Neutro	181 (28,7%)	418 (66,3%)	2 (0,3%)	29 (4,6%)	98,90%	86,20%	99,50%

VP - Verdadeiros Positivos; VN - Verdadeiros Negativos; FP - Falsos Positivos; FN - Falsos Negativos.

A precisão foi muito idêntica para a expressão sorrir e beijinho, destacando-se na expressão neutra com uma percentagem de 98,90%. Esta tem um melhor resultado devido ao baixo valor de falsos positivos.

Segundo a tabela acima, nota-se que a expressão facial sorriso é a que apresenta melhor sensibilidade, com 92,40%, pois revela um menor número de falsos negativos. Pelo contrário, a expressão neutra é a que apresenta mais falsos negativos, fazendo com que a sensibilidade seja mais baixa (86,20%).

Os valores da especificidade na expressão sorriso e beijinho foram muito idênticos pois têm um número de verdadeiros negativos e falsos positivos muito idêntico. Destaca-se a expressão facial neutra que atinge um valor muito perto de 100% (99,50%). Isto acontece devido ao reduzido número de falsos positivos (apenas 2 *frames*).

Para resumir estes resultados e ter uma melhor visualização destes foi feito o Gráfico 1 onde é possível ver os números reduzidos de falsos positivos e falsos negativos. Os valores de verdadeiros negativos são idênticos para as três expressões faciais. O mesmo acontece com os verdadeiros positivos.

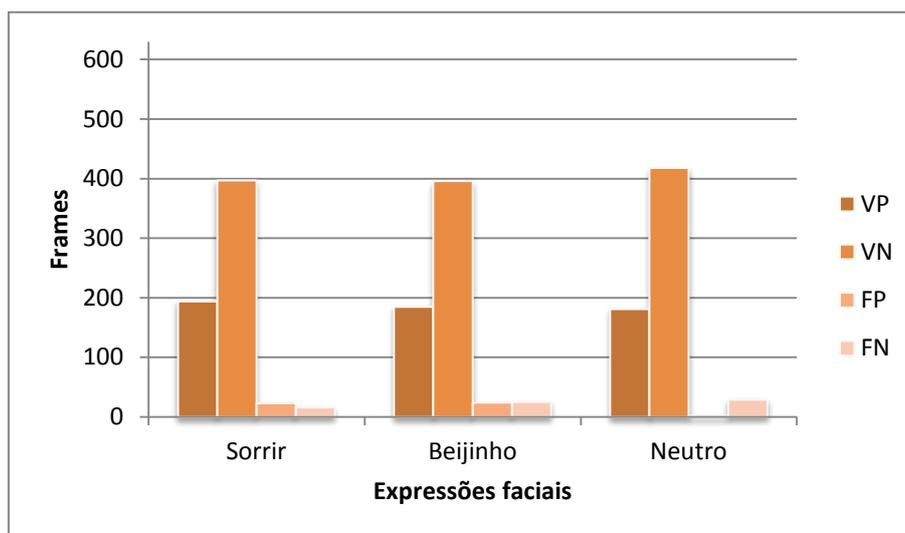


Gráfico 1- Gráfico referente ao modelo de classificação do estudo de caso.

Foram analisados os mesmos vídeos para o olho esquerdo. Como a característica mais distintiva de uma expressão facial é a boca, não se fez a mesma análise exaustiva para os olhos. Apenas observou-se se, quando a expressão facial falhava na boca, o olho conseguia colmatar essa falha e classificar a expressão corretamente. Realmente isso verificou-se. No entanto no olho surgem os mesmos problemas que a boca. Por vezes é mal detetado ou a expressão é confundida. Apesar disso o objetivo é concluído com sucesso pois cada vez que a boca falha, o olho colmata essa falha e faz a classificação corretamente.

5.3 Resultados dos testes com pessoas sem deficiência

A avaliação foi feita à pessoa do estudo de caso, avaliando a detecção da boca e dos olhos juntamente com a classificação da expressão facial. Para percebermos qual a diferença da performance do protótipo em pessoas com deficiência e sem deficiência então decidiu-se fazer testes a cinco pessoas sem qualquer tipo de deficiência. As pessoas que participaram nos testes têm idades compreendidas entre 23 a 53 anos. Das pessoas que participaram na avaliação, duas são sexo masculino, sendo as restantes (três) do sexo feminino. Foi pedido que as pessoas fizessem as 3 expressões: sorrir, mandar beijo e neutra. Analisaram-se cinco vídeos de cada expressão. No total analisaram-se 2250 *frames*. Na Tabela 6 do Anexo A pode ver-se os resultados obtidos relativamente à boca.

Relativamente aos erros do sistema, estes são os mesmos que foram detetados no estudo de caso.

A matriz de classificação (Tabela 4) mostra os resultados da Tabela 6 de forma resumida.

Tabela 4- Matriz de confusão dos resultados de pessoas sem deficiência.

Expressões previstas \ Expressões reais	Sorrir (%)	Beijinho (%)	Neutro (%)	Não Identificado (%)	Total (%)
Sorrir	97,46	0,53	0,27	1,74	100
Beijinho	2,8	90,94	0	6,26	100
Neutro	0	0	96,94	3,06	100

Como se pode ver pela Tabela 4 a expressão sorriso foi a que obteve melhor resultado (97,46%). A expressão beijinho obteve o valor mais baixo. A expressão beijinho foi a que teve um maior número de classificações incorretas e também da deteção incorreta da boca. A expressão neutra não apresentou erros de classificação.

O método de avaliação dos resultados foi o mesmo para as pessoas sem deficiência e assim foi construída uma tabela com verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Na tabela 5 pode-se ver também o resultado dos cálculos da precisão, sensibilidade e especificidade.

Tabela 5 - Avaliação dos resultados das pessoas sem deficiência.

	VP	VN	FP	FN	Precisão	Sensibilidade	Especificidade
Sorrir	731 (32,5%)	1479 (65,7%)	21 (0,9%)	19 (0,8%)	97,20%	97,50%	98,60%
Beijinho	682 (30,3%)	1496 (66,5%)	4 (0,2%)	68 (3%)	99,40%	90,90%	99,70%
Neutro	727 (32,3%)	1498 (66,6%)	2 (0,1%)	23 (1%)	99,70%	96,90%	99,90%

VP - Verdadeiros Positivos; VN - Verdadeiros Negativos; FP - Falsos Positivos; FN - Falsos Negativos.

A precisão é muito idêntica na expressão beijinho e neutra, sendo muito próximo de 100%. Isto deve-se ao baixo número de falsos positivos em cada expressão. Na expressão sorrir este número já é mais reduzido, levando a uma menor precisão do sistema.

No caso da sensibilidade, a expressão que obteve melhor resultado foi a expressão sorrir, tendo um número baixo de falsos negativos. Por sua vez, a expressão beijinho foi a que teve um valor mais baixo pois apresenta um número elevado de falsos negativos (68 *frames*).

Na especificidade, os valores das três expressões são muito parecidos, destacando-se a expressão beijinho e neutra que têm valores próximos de 100%.

No Gráfico 2 é possível ver os valores do modelo de avaliação do sistema para pessoas sem deficiência.

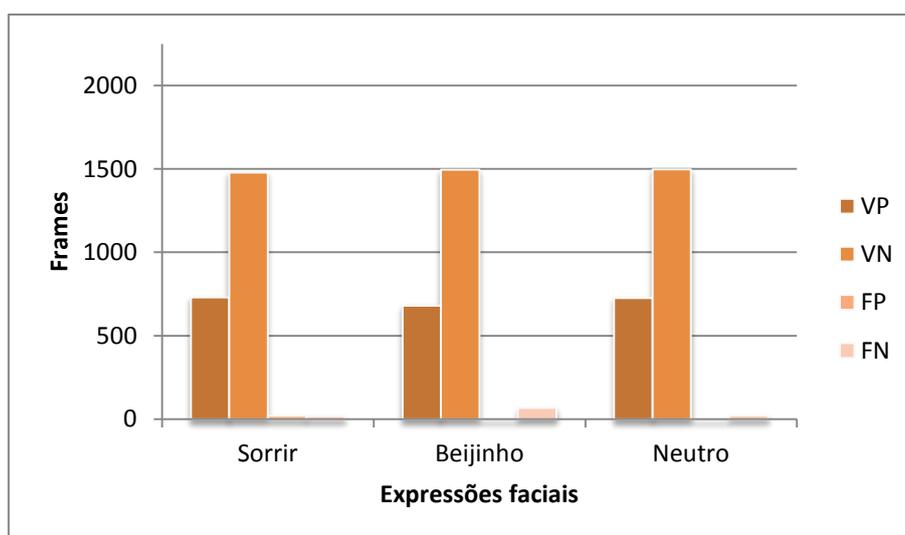


Gráfico 2 - Gráfico do modelo de avaliação de pessoas sem deficiência.

No caso das pessoas sem deficiência optou-se por não testar o olho esquerdo pois este é muito idêntico nas duas expressões (sorrir e beijinho) e por isso não auxilia na classificação

da expressão facial. Pode ver-se pela Figura 30 que não existe uma diferença significativa para que seja possível distinguir expressões faciais apenas pelo olho esquerdo.

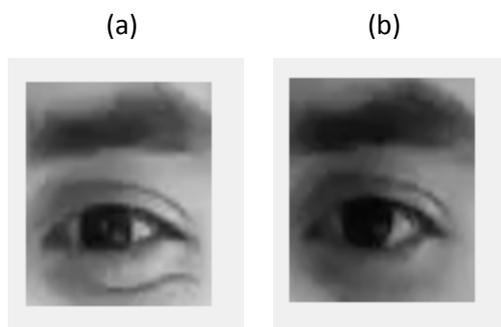


Figura 30 - Olho esquerdo de uma pessoa sem deficiência nas seguintes expressões: (a) – sorrir; (b) – beijinho.

5.4 Comparação de resultados

Para perceber a diferença entre os resultados do estudo de caso e das pessoas sem deficiência fez-se uma breve comparação.

Em comparação com a matriz de confusão obtida no estudo de caso, pode concluir-se que os resultados das pessoas sem deficiência foram melhores em todas as expressões. Como pode observar-se no Gráfico 3 as taxas de acerto foram sempre superiores em pessoas sem deficiência, não havendo diferenças significativas em relação às taxas do estudo de caso.

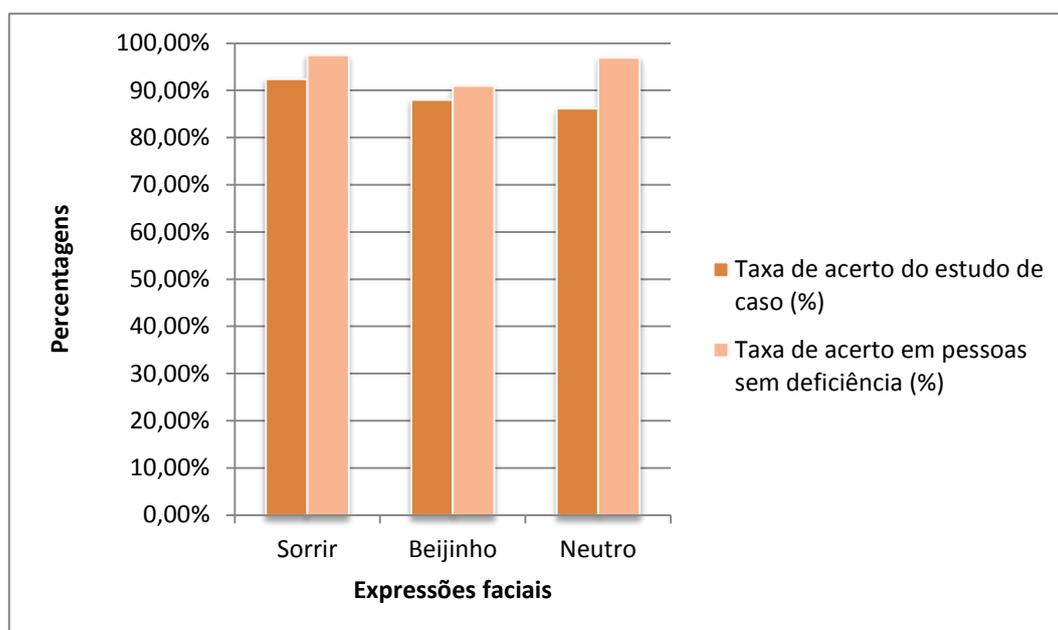


Gráfico 3- Taxas de acerto em ambos os casos.

No Gráfico 4 pode ver-se as taxas de erro da deteção da boca (correspondendo à coluna “Não identificado” nas matrizes de confusão). A expressão neutra foi a que teve mais erros no estudo de caso, sendo que nas pessoas sem deficiência foi a expressão beijinho. No estudo de caso, a expressão beijinho foi a que obteve menos erros de deteção e nas pessoas sem deficiência foi a expressão sorrir.

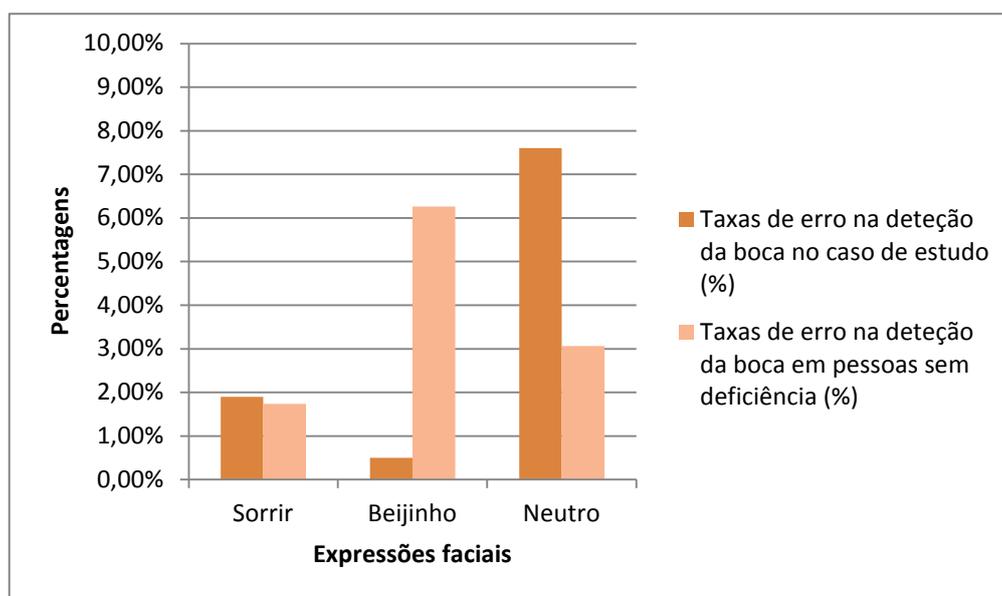


Gráfico 4 - Taxas de erro na deteção da boca.

Em comparação com o estudo de caso, os valores da precisão em pessoas sem deficiência são mais elevados sendo que o mais elevado corresponde à expressão neutra tal como no estudo de caso. Esta comparação revela que o sistema é mais preciso (identifica mais casos positivos que são corretamente identificados) em pessoas sem deficiência. Relativamente à sensibilidade, os valores são também elevados e o maior corresponde em ambos os casos à expressão facial sorrir. No entanto no estudo de caso o menor corresponde à expressão neutra e em pessoas sem deficiência corresponde à expressão beijinho. Com base nestes resultados e na nossa opinião, pode dizer-se que o protótipo tem mais capacidade em identificar verdadeiros positivos em pessoas sem deficiência na expressão sorrir e neutra, sendo que na expressão beijinho verificou-se o contrário. Por fim, na especificidade a expressão que teve resultados mais elevados em ambas as situações foi a expressão neutra com valores muito idênticos. No estudo de caso, as outras expressões obtiveram resultados mais baixos relativamente aos resultados das pessoas sem deficiência. Nestas expressões, o sistema tem mais capacidade em detetar verdadeiros negativos nas pessoas sem deficiência. Esta comparação de resultados pode ser ver nos Gráficos 5 e 6.

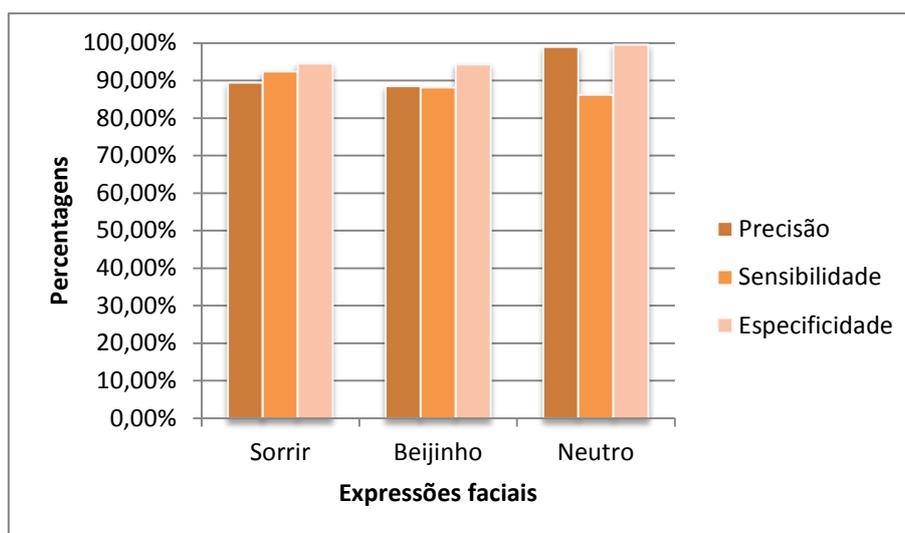


Gráfico 5 - Resultados da precisão, sensibilidade e especificidade do estudo de caso.

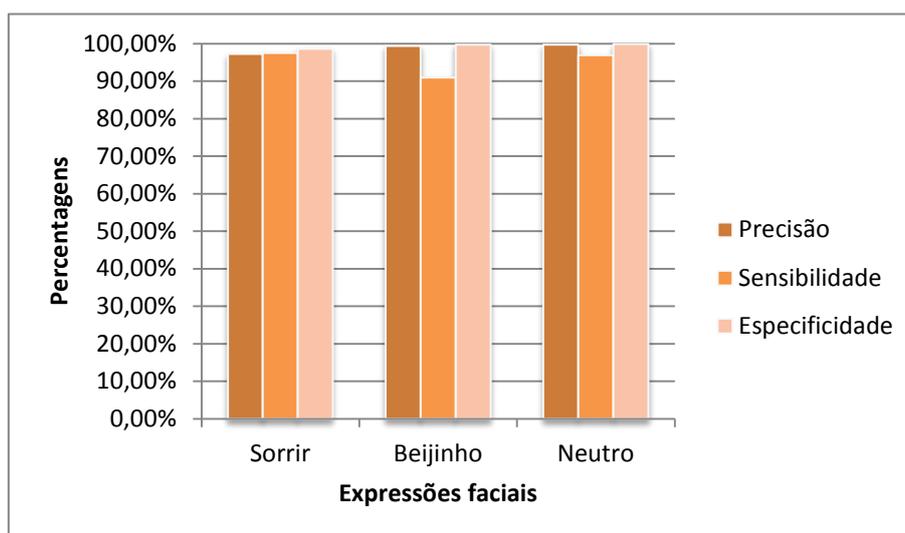


Gráfico 6 - Resultados da precisão, sensibilidade e especificidade dos testes das pessoas sem deficiência.

Como finalização desta comparação, é possível fazer algumas conclusões acerca dos resultados. Entre a expressão sorrir e beijinho, a expressão sorrir tem melhores resultados no estudo de caso. Em pessoas sem deficiência estas expressões tiveram resultados idênticos. A expressão neutra teve resultados muito positivos em ambos os casos, concluindo assim que as expressões são excluídas com sucesso. Os resultados foram melhores no caso de pessoas sem deficiência, mostrando assim a complexidade do estudo de caso do projeto. Apesar desta conclusão, é possível dizer que os resultados para o estudo de caso foram muito bons.

5.5 Conclusão

Neste capítulo foram apresentados os resultados obtidos pelos testes realizados à pessoa do estudo de caso e às pessoas sem deficiência. Primeiramente é feita uma apresentação da amostra e do procedimento, explicando como foram feitas as filmagens e o que foi feito durante os testes. De seguida fez-se a apresentação dos resultados seguindo-se da sua análise. Estas foram divididas em duas partes: resultados dos testes do estudo de caso e dos testes das pessoas sem deficiência. Nestes resultados são apresentadas as matrizes de confusão e um modelo de avaliação que inclui os cálculos da precisão, sensibilidade e especificidade. Por fim fez-se uma comparação de resultados para perceber as possíveis conclusões a tirar com os testes que foram desenvolvidos com o protótipo.

6. Considerações finais e trabalho futuro

O presente trabalho centrou-se no desenvolvimento de uma aplicação que faz a deteção facial e posterior reconhecimento de expressões faciais.

Para o desenvolvimento do sistema foi pensado num modelo fácil de usar, que pudesse ser universal, ou seja, que possa usado por todos e de baixo custo. O único aspeto que muda de utilizador para utilizador são as imagens da base de dados. O restante conteúdo fica igual. O protótipo está separado em dois métodos. Para a deteção facial, da boca e olho esquerdo é utilizado o algoritmo Viola e Jones. O método PCA é usado para o reconhecimento de expressões faciais.

Tendo em conta todos os objetivos previstos inicialmente e os resultados obtidos pelos testes realizados, o protótipo cumpre com sucesso todas as metas.

Apesar de já existirem inúmeros trabalhos feitos nesta área (reconhecimento de expressões faciais), não foram encontrados estudos que tenham realizado testes em pessoas com deficiência. Apesar de na teoria dizerem que o trabalho que desenvolveram pode servir, isso nem sempre acontece. Exemplo disso é o estudo de caso deste trabalho que por ter a língua sempre de fora dificulta a deteção da boca e conseqüente a expressão facial. Por isso, este trabalho apresenta não só um protótipo como também testes realizados com uma pessoa com uma doença neuromuscular degenerativa. Para compreender se há diferenças nos resultados entre pessoas sem deficiência e o nosso estudo de caso, então foi decidido aplicar o mesmo protótipo em cinco pessoas sem qualquer tipo de limitações. No final da análise dos resultados é possível ver que, na sua globalidade, os resultados foram melhores nos testes das pessoas sem deficiência do que nos testes do estudo de caso. Com isso conclui-se que é complicado criar sistemas de reconhecimento de expressões faciais para pessoas com deficiência física. No entanto, apesar de o estudo de caso deste trabalho ter complicado (a língua de fora foi a maior dificuldade) o desenvolvimento do protótipo, pode-se concluir que os resultados foram positivos e que é importante criar interfaces entre a pessoa com deficiência e o computador.

6.1 Trabalho futuro

Embora este trabalho tenha cumprido todos os objetivos é necessário otimizar a ferramenta de interação com o computador. Um dos aspetos que poderá ser adicionado ao sistema é a possibilidade de ter mais que duas expressões a serem reconhecidas e assim puder

ter mais funções no computador a realizar. Outro aspeto será dar a escolher ao utilizador qual a função que quer para cada expressão. O *design* da aplicação também deve ser melhorado.

Bibliografia

- Aleksic, P., & Katsaggelos, A. (2006). Automatic Facial Expression Recognition using Facial Animation Parameters and Multi-Stream HMMS. *Information Forensics and Security*, 3 - 11.
- Almeida, O. C. (2006). *Técnicas de processamento de imagens para localização e reconhecimento de faces*. São Carlos: Instituto de Ciências Matemáticas e de Computação.
- Arya, S., Neeraj, P., & Bhatia, K. (2015). Future of Face Recognition: A Review. *Procedia Computer Science*, 578-585.
- Bakshi, U., & Singhal, R. (2014). A survey on face detection methods and feature extraction techniques of face recognition. *International Journal of Emerging Trends & Technology in Computer Science* , 233-237.
- Bakshi, U., & Singhal, R. (2014). A survey on face detection methods and feature extraction techniques of face recognition. *International Journal of Emerging Trends & Technology in Computer Science*, 233-237.
- Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior. *Computer Vision and Pattern Recognition*, 568-573.
- Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006). Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 22-35.
- Bersch, R., & Machado, R. (2007). *Atendimento educacional especializado*. Brasília: SEESP / SEED / MEC.
- Cai, J., & Goshtasby, A. A. (1999). Detecting human faces in color images. *Image Vision Comput*, 63-75.

- Chen, Q., Wu, H., & Yachida, M. (1995). Face detection by fuzzy pattern matching. *Proceedings of the Fifth International Conference on Computer Vision*, 591.
- Dai, Y., & Nakano, Y. (1996). Face-texture model based on sgld and its application in face detection in a color scene. *Pattern Recognition*, 1007-1017.
- Deng, H.-B., Jin, L.-W., Zhen, L.-X., & Huang, J.-C. (2005). A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA. *International Journal of Information Technology*, 86-96.
- Dhawan, S., & Dogra, H. (2012). Feature Extraction Techniques for Face Recognition. *International Journal of Engineering, Business and Enterprise Applications (IJEBA)*, 1-4.
- Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2005). *Human-Computer Interaction*.
- Dongre, S., & Patil, S. (2015). A Face as a Mouse for Disabled Person. *A Monthly Journal of Computer Science and Information Technology*, 156-160.
- Filipe, S. B. (2008). *Detecção de faces humanas em tempo real*. Covilhã: Universidade da Beira Interior.
- Guerreiro, T. J., & Jorge, J. A. (2006). Controlo Miográfico de Dispositivos Móveis para Tetraplégicos. *Desenvolvimento de Software para melhoria de Acessibilidade e combate à Infoexclusão – DSAI*.
- Jamshidnezhad, A., & Nordin, M. j. (2011). A Classifier Model based on the Features Quantitative Analysis for Facial Expression Recognition . *Proceeding of the International Conference on Advanced Science, Engineering and Information Technology*, 391-394.
- Jeng, S.-H., Liao, H. Y., Han, C. C., Chern, M. Y., & Liu, Y. T. (1998). Facial feature detection using geometrical face model: An efficient approach. *Pattern Recognition*, 273-282.
- Kim, G. J. (2015). *Human-Computer Interaction: Fundamentals and Practice*. Boca Raton: Taylor & Francis Group.

- Kim, H. S., Kim, E. Y., Hwang, S. W., & Kim, H. J. (2000). Object-based human face detection. *Digest of Technical Papers. International Conference on*, 354–355.
- Kotsia, I., & Pitas, I. (2007). Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. *IEEE Transactions on Image Processing*, 172-187.
- Koutlas, A., & Fotiadis, D. (2008). A Region Based methodology for facial expression recognition. *Systems, Man and Cybernetics*, 662 - 666.
- Lau, B. T. (2010). Portable real time emotion detection system for the disabled. *Expert Systems with Applications*, 6561–6566.
- Leão, P. L., Bezerra, S. J., Matos, N. L., & Nunes, M. A. (2012). Detecção de expressões faciais: uma abordagem baseada em análise do fluxo óptico. *GEINTEC*, 472-489 .
- Lin, C., & Fan, K.-C. (2001). Triangle-based approach to the detection of human face. *Pattern Recognition*, 1271–1284.
- Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J., & Movellan, J. (2006). Dynamics of Facial Expression Extracted Automatically from Video. *Image and Vision Computing*, 615–625.
- Majaranta, P., & Bulling, A. (2014). *Eye Tracking and Eye-Based Human–Computer Interaction*. London: Springer-Verlag.
- MathWorks. (2016, 04 29). *vision.CascadeObjectDetector System object*. (MathWorks) Retrieved 04 29, 2016, from <http://www.mathworks.com/help/vision/ref/vision.cascadeobjectdetector-class.html?refresh=true>
- Nagarajan, R., Hariharan, M., & Satiyan, M. (2011). Luminance Sticker Based Facial Expression Recognition Using Discrete Wavelet Transform for Physically Disabled Persons. *Journal of Medical Systems*, 2225-2234.
- Parker, J. (2001). Rank and Response Combination From Confusion Matrix Data. *Elsevier*.

- Pedro, L., & Ribeiro, J. L. (2010). Implicações da situação profissional na qualidade de vida em indivíduos com esclerose múltipla. *Simpósio Nacional de Investigação em Psicologia*, 1286-1294.
- Ren, Z., Meng, J., & Yuan, J. (2011). Depth Camera Based Hand Gesture Recognition and its Applications in Human-Computer-Interaction. *8th International Conference on Information, Communications & Signal Processing*, (pp. 1-5). Singapura.
- Rito, M. (2006). Doença de Parkinson: Instrumentos Avaliativos. *Revista Portuguesa de Fisioterapia*, 27-45.
- Santanãs, L. M., Gomes, F. R., & Santos, T. S. (2015). Processo de deteção facial, utilizando Viola;Jones. *Interfaces Científicas - Exatas e Tecnológicas*, 35-40.
- Saúde, O. M. (2004). *Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF)*. Lisboa: Direcção-Geral da Saúde.
- Silva, R. K. (2013). *Interfaces naturais e o reconhecimento das línguas de sinais*. São Paulo: Universidade Católica de São Paulo.
- Silva, V. A. (2008). *Comparação entre Técnicas de Reconhecimento de Faces para Controle de Acesso a Computadores*. Campo Grande: Universidade Católica Dom Bosco.
- Sumathi, C. P., Santhanam, T., & Mahadevi, M. (2012). Automatic Facial Expression Analysis: a Survey. *International Journal of Computer Science & Engineering Survey*, 47-59.
- Thai, L. H., Nguyen, N. D., & Hai, T. S. (2011). A Facial Expression Classification System Integrating Canny, Principal Component Analysis and Artificial Neural Network. *International Journal of Machine Learning and Computing*, 388-393.
- Tong, Y., Wang, Y., Zhu, Z., & Ji, Q. (2007). Robust facial feature tracking under varying face pose and facial expression. *Pattern Recognition*, 3195–3208.
- Valverde, R. N., Pereira, F. G., & Vassallo, R. F. (2011). Reconhecimento de gestos dinâmicos usando modelos ocultos de Markov para interação homem máquina. (pp. 438 - 443). São João del-Rei: Simpósio Brasileiro de Automação Inteligente.

- Varona, J., Manresa-Yee, C., & Perales, F. (2008). Hands-free vision-based interface for computer accessibility. *Journal of Network and Computer Applications*, 357–374.
- Vasanthan, M., Nagarajan, R., Murugappan, M., & Ilias, B. (2012). Facial Expression Based Computer Cursor Control System for Assisting Physically Disabled Person. *Communication, Networks and Satellite (ComNetSat)*, 172-176.
- Viola, P., & Jones, M. (2001). Robust Real-time Face Detection. *International Conference on Computer Vision*, 1.
- Wang, J., & Tan, T. (2000). A new face detection method based on shape information. *Pattern Recognition Letters*, 463-471.
- Yang, P., Liu, Q., & Metaxas, D. (2009). Boosting encoded dynamic features for facial expression recognition. *Pattern Recognition Letters*, 132–139.
- Yeasin, M., Bullot, B., & Sharma, R. (2006). Recognition of Facial Expressions and Measurement of Levels of Interest From Video. *IEEE Transactions on Multimedia*, 500-508.
- Yow, K., & Cipolla, R. (1997). Feature-based human face detection. *Image and vision computing*, 713–735.

Anexos

Anexo A – Tabela dos resultados da análise das expressões faciais na boca em pessoas sem deficiência

Vídeos	Sorrir	Beijinho	Neutro	Não Identificado	Duração (frames)	Expressão facial
Vídeo 1	30	0	0	0	30	
Vídeo 2	30	0	0	0	30	
Vídeo 3	30	0	0	0	30	
Vídeo 4	30	0	0	0	30	
Vídeo 5	30	0	0	0	30	
Vídeo 6	30	0	0	0	30	
Vídeo 7	30	0	0	0	30	
Vídeo 8	30	0	0	0	30	
Vídeo 9	30	0	0	0	30	
Vídeo 10	30	0	0	0	30	
Vídeo 11	30	0	0	0	30	
Vídeo 12	30	0	0	0	30	
Vídeo 13	30	0	0	0	30	Sorrir
Vídeo 14	30	0	0	0	30	
Vídeo 15	30	0	0	0	30	
Vídeo 16	25	0	0	5	30	
Vídeo 17	22	0	0	8	30	
Vídeo 18	30	0	0	0	30	
Vídeo 19	30	0	0	0	30	
Vídeo 20	30	0	0	0	30	
Vídeo 21	30	0	0	0	30	
Vídeo 22	30	0	0	0	30	
Vídeo 23	28	0	2	0	30	
Vídeo 24	30	0	0	0	30	
Vídeo 25	26	4	0	0	30	
Vídeo 26	0	30	0	0	30	
Vídeo 27	15	15	0	0	30	
Vídeo 28	6	24	0	0	30	
Vídeo 29	0	30	0	0	30	

Vídeo 30	0	30	0	0	30
Vídeo 31	0	30	0	0	30
Vídeo 32	0	18	0	12	30
Vídeo 33	0	20	0	10	30
Vídeo 34	0	30	0	0	30
Vídeo 35	0	30	0	0	30
Vídeo 36	0	24	0	6	30
Vídeo 37	0	20	0	10	30
Vídeo 38	0	30	0	0	30
Vídeo 39	0	30	0	0	30
Vídeo 40	0	30	0	0	30
Vídeo 41	0	30	0	0	30
Vídeo 42	0	30	0	0	30
Vídeo 43	0	24	0	6	30
Vídeo 44	0	27	0	3	30
Vídeo 45	0	30	0	0	30
Vídeo 46	0	30	0	0	30
Vídeo 47	0	30	0	0	30
Vídeo 48	0	30	0	0	30
Vídeo 49	0	30	0	0	30
Vídeo 50	0	30	0	0	30
Vídeo 51	0	0	30	0	30
Vídeo 52	0	0	30	0	30
Vídeo 53	0	0	28	2	30
Vídeo 54	0	0	29	1	30
Vídeo 55	0	0	28	2	30
Vídeo 56	0	0	30	0	30
Vídeo 57	0	0	30	0	30
Vídeo 58	0	0	30	0	30
Vídeo 59	0	0	29	1	30
Vídeo 60	0	0	30	0	30

Beijinho

Vídeo 61	0	0	30	0	30	Neutro
Vídeo 62	0	0	30	0	30	
Vídeo 63	0	0	30	0	30	
Vídeo 64	0	0	30	0	30	
Vídeo 65	0	0	30	0	30	
Vídeo 66	0	0	20	10	30	
Vídeo 67	0	0	30	0	30	
Vídeo 68	0	0	23	7	30	
Vídeo 69	0	0	30	0	30	
Vídeo 70	0	0	30	0	30	
Vídeo 71	0	0	30	0	30	
Vídeo 72	0	0	30	0	30	
Vídeo 73	0	0	30	0	30	
Vídeo 74	0	0	30	0	30	
Vídeo 75	0	0	30	0	30	