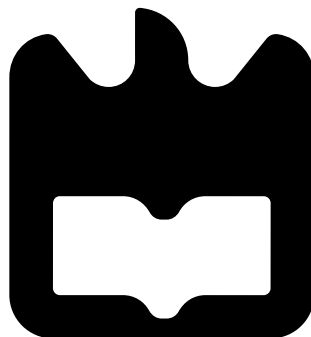




**Marco António
Macedo Pinto**

**EMOTIVA: Métodos para a Criação de uma Base
de Dados Dinâmica de Expressão de Emoções
EMOTIVA: Methods for Creating a Dynamic
Emotion Expression Database**





**Marco António
Macedo Pinto**

**EMOTIVA: Métodos para a Criação de uma Base
de Dados Dinâmica de Expressão de Emoções
EMOTIVA: Methods for Creating a Dynamic
Emotion Expression Database**

“If everyone were cast in the
same mould, there would be
no such thing as beauty.”

— Charles Darwin



**Marco António
Macedo Pinto**

**EMOTIVA: Métodos para a Criação de uma Base
de Dados Dinâmica de Expressão de Emoções
EMOTIVA: Methods for Creating a Dynamic
Emotion Expression Database**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Doutor Samuel de Sousa Silva, Investigador do Instituto de Engenharia Eletrónica e Informática de Aveiro, e da Professora Doutora Sandra Cristina de Oliveira Soares, Professora Auxiliar do Departamento de Educação e Psicologia da Universidade de Aveiro.

o júri / the jury

presidente / president

Professor Doutor Joaquim João Estrela Ribeiro Silvestre Madeira
Professor Auxiliar da Universidade de Aveiro

vogais / examiners committee

Professor Doutor Daniel Filipe Albuquerque
Professor Adjunto do Instituto Politécnico de Viseu - Escola Superior de Tecnologia e Gestão de Viseu (Arguente Principal)

Doutor Samuel de Sousa Silva

Investigador Doutorado (nível 2) da Universidade de Aveiro (Orientador)

**agradecimentos /
acknowledgements**

É com muito gosto que aproveito esta oportunidade para agradecer a todos os que me ajudaram neste longa etapa. Ficarei para sempre grato a quem me apoiou e incentivou na realização desta dissertação de mestrado.

Ao Doutor Samuel de Sousa Silva, pela orientação e disponibilidade que sempre demonstrou, por todas as críticas, opiniões e soluções que me ajudaram a ultrapassar todos os obstáculos que apareceram durante a realização do trabalho.

A todos os docentes que colaboraram na sua realização, nomeadamente a Professora Doutora Sandra Cristina de Oliveira Soares e a aluna Inês Fontes. As meus pais e irmão por todos os sacrifícios e incentivos que fizeram para que isto se tornasse possível.

À minha namorada, que sempre me deu força e que me acompanha desde que iniciei este longo percurso.

Resumo

As emoções têm um papel central no nosso dia-a-dia e, por isso, entender como expressamos e percebemos as emoções nos outros tem sido objecto de inúmeros trabalhos. Para estudar a influência que determinada emoção pode ter no indivíduo, é comum os procedimentos experimentais recorrerem ao visionamento de fotografias de pessoas a expressar emoções (e.g., medo, alegria) obtidas de bases de dados existentes. No entanto, a expressão de emoção é um processo dinâmico, i.e., uma expressão de alegria, por exemplo, exhibe uma progressão de movimentos da face ao longo do tempo, não surge, na sua máxima intensidade, instantaneamente.

Este carácter dinâmico da expressão de emoção tem também motivado estudos que tentam perceber, entre outros aspectos, que movimentos faciais caracterizam cada emoção e o momento a partir do qual conseguimos reconhecer a emoção que está a ser expressada. Para este efeito, a literatura tem recorrido, em alguns momentos, a vídeos, mas o mais usual é a consideração de técnicas de interpolação (*morphing*) para construir sequências de imagens com diferentes intensidades de cada emoção. Ainda que esta abordagem tenha sido usada com sucesso em vários trabalhos, deixa muitas questões sobre como um processo de interpolação pode afetar o realismo dos dados que se estão a analisar. Neste contexto, a existência de uma base de dados que incluisse sequências reais de expressão de emoção, seria muito útil para a comunidade e para um estudo mais amplo da expressão de emoção.

Nesse âmbito, este trabalho, desenvolvido em colaboração com o Departamento de Educação e Psicologia da Universidade de Aveiro, propõe uma metodologia e métodos computacionais para suportar a criação de uma base de dados contendo dados dinâmicos de expressão de emoção. Para esse efeito, foi proposto um protocolo experimental para, de forma sistemática, suportar a recolha dos dados de expressão de emoção. Como base nessa proposta, foram definidos os requisitos, identificados os diferentes passos a ser assegurados e desenvolvida uma ferramenta interativa para a sua articulação incluindo: 1) aquisição dos dados recorrendo a uma Kinect V2; 2) a validação dos dados adquiridos pelo participante e experimentador; 3) a anotação das sequências adquiridas para identificação dos intervalos de interesse em que a emoção está a ser expressada; 4) o pós-processamento e normalização das sequências; e 5) o armazenamento estruturado dos dados finais.

O trabalho desenvolvido permite, agora, e de forma sistemática, iniciar a recolha e processamento de dados dinâmicos de expressão de emoção.

Abstract

Emotions have a crucial role on our day-to-day life and therefore, understand how we express and interpret them in other people has been a matter of object of countless studies. To understand the influence of one particular emotion in a person, it is common, for experimental procedures, to explore the visioning of people portraits expressing emotions (e.g., fear, joy), obtained from existing databases. Nevertheless, the emotion expression is a dynamic process, i.e., an expression of joy, for example, displays a progression of face movements through the time; it does not come up immediately.

This dynamic nature of the emotion expression has also motivated studies that try to understand, among other aspects, which facial movements represent each emotion, as well as the moment from when we recognize the emotion that is being expressed. To this effect, literature has been resorting, in some moments, to videos. Even though, the most usual is to resort to the consideration of interpolation techniques (morphing) to create sequences of images with different intensities for each emotion. Although this approach has been used with success in several works, it leaves many questions regarding how an interpolation process can affect the realism of the databases that are being analyzed. In this context, the existence of a database that could include real emotion expression sequences, would be very useful to the community and to a wider study about the expression of emotion.

In that scope, this work, developed in collaboration with the Education & Psychology Department of the University of Aveiro, proposes a methodology and computing methods to support the development of a database containing dynamic data of emotion expression. To that effect, an experimental protocol was developed to, in a systematic way, support the collection of expression emotion data. Based on that proposal, the requirements were defined, the different steps and phases to be ensured were identified and an interactive tool was developed to its articulation, including: 1) data acquisition, using the Microsoft sensor - Kinect V2; 2) the validation of the acquired data by the participants and the experimenter; 3) the annotation of the acquired sequences in order to identify the interest ranges when the emotion is being expressed; 4) the post-processing and normalization of the sequences; and 5) the structured storage of the final data.

The developed work allows now, in a systematic way, the beginning of the dynamic expression emotion data collecting and processing.

Conteúdo

Conteúdo	i
Lista de Figuras	iii
Lista de Tabelas	v
1 Introdução	1
1.1 Contexto e Motivação	1
1.2 Desafios	1
1.3 Objetivos	2
1.4 Estrutura do Documento	3
2 Estado da Arte	5
2.1 Estudos Envolvendo Expressão Facial de Emoção	5
2.1.1 Expressões Faciais de Emoções	7
2.2 Protocolos Experimentais	8
2.3 Bases de Dados Existentes	10
2.3.1 Karolinska Directed Emotional Faces (KDEF)	10
2.3.2 I BU-3DFE (Binghamton University 3D Facial Expression) Database (Static Data)	11
2.3.3 II BU-4DFE (Binghamton University 3D + time): A 3D Dynamic Fa- cial Expression Database (Dynamic Data)	12
2.3.4 III. BP4D-Spontaneous: Binghamton-Pittsburgh 3D Dynamic Sponta- neous Facial Expression Database	13
2.3.5 The Bosphorus Database	14
2.4 Fases da Criação de uma Base de Dados Emocional	15
2.4.1 Aquisição	15
2.4.1.1 Microsoft Kinect	15
2.4.1.2 Sensor Microsoft Kinect for windows V2	16
2.4.1.3 Características Principais	16
2.4.1.4 Action Units do sensor Microsoft Kinect	17
2.4.1.5 Funcionalidades do Microsoft Kinect	17
2.4.2 Processamento	18
2.4.3 Caracterização Quantitativa dos Estímulos	19
2.4.3.1 OpenFace	19
2.4.3.2 Funcionalidades do OpenFace	20
2.5 Discussão	21

3	Protocolo Experimental e Requisitos	23
3.1	Conteúdo da Base de Dados	23
3.2	Protocolo Experimental	23
3.2.1	Procedimento	24
3.2.2	Setup Experimental	25
3.3	Pós-Processamento	25
3.4	Requisitos	26
3.4.1	Hardware e Software	26
3.4.2	Aquisição	27
3.4.3	Anotação	28
3.4.4	Normalização	28
3.4.5	Análise Quantitativa	28
3.5	Conclusões	28
4	Sistema Computacional de Suporte à Aquisição de uma Base de Dados Emocional	29
4.1	Visão Global	29
4.2	Configuração da Experiência	31
4.3	Método de Aquisição de Dados	33
4.3.1	Integração com Kinect	33
4.3.2	Gravação via Kinect	33
4.4	Aquisição de Dados: Expressão Emocional através de <i>Acting</i>	35
4.4.1	Ratings	35
4.4.2	Papel do Experimentador	36
4.5	Aquisição de Dados: Expressão Emocional Espontânea	36
4.5.1	Papel do Experimentador	37
4.5.2	Expressão Espontânea de Emoção	37
4.6	Armazenamento dos Dados Experimentais	38
4.6.1	Estrutura	39
4.7	Conclusões	39
5	Anotação e Pós Processamentos de Dados	41
5.1	Visão Global	41
5.2	Anotação	42
5.3	Normalização e Resultados	43
5.4	Descritores Quantitativos dos Dados	45
5.4.1	Action Units	46
5.4.2	Marcos Faciais	48
5.5	Base de Dados Resultante	49
5.5.1	Criação de Vídeos Emocionais	49
5.6	Conclusões	50
6	Conclusões	55
	Bibliografia	57

Lista de Figuras

2.1	As 7 emoções básicas.	8
2.2	Transição entre nojo e alegria.	10
2.3	Expressões faciais KDEF.	10
2.4	I BU-3DFE níveis	11
2.5	I BU-3DFE expressões femininas	11
2.6	I BU-3DFE expressões masculinas	12
2.7	Visualização dos modelos de sequências de vídeo 3D	12
2.8	II BU-4DFE expressões masculinas	13
2.9	II BU-4DFE expressões femininas	13
2.10	Expressões para as 8 tarefas	14
2.11	Exemplo com e sem textura, ambas de frente e 45°.	14
2.12	Sensor Microsoft Kinect v2.	16
2.13	17 Action units extraídas da Kinect v2	17
2.14	Valores numéricos das AUs correspondentes a várias expressões.	18
2.15	Deteção de marcos faciais utilizando a ferramenta OpenFace.	20
2.16	Tracking da cabeça utilizando a ferramenta OpenFace.	20
2.17	Reconhecimento das action units na face utilizando a ferramenta OpenFace.	21
3.1	Setup experimental	26
4.1	Fluxograma aquisição	31
4.2	Formulário inicial	32
4.3	Gravação via Kinect	34
4.4	Ratings	35
4.5	Aprovação do experimentador	36
4.6	Formulário experimentador	37
4.7	Formulário participante	38
4.8	Directório	40
4.9	Dados experimentais	40
5.1	Fluxograma dos métodos de pós-processamento	42
5.2	Anotação de dados: Selecção das sequências de frames	43
5.3	Anotação de dados: Selecção do frame Neutro Inicial	43
5.4	Anotação de dados: Selecção do frame Pico	44
5.5	Anotação de dados: Selecção do frame Neutro Final	44
5.6	Formulário de Normalização: Processo de selecção e normalização das sequências	46
5.7	Sequência de frames não normalizada para Alegria	46

5.8	Sequência de frames normalizada para Alegria	47
5.9	Sequência de frames não normalizada para Nojo	47
5.10	Sequência de frames normalizada para Nojo	48
5.11	Interface da ferramenta OpenFace	48
5.12	Action Units disponíveis com o OpenFace	51
5.13	Action Units para sequência de Alegria	52
5.14	Action Units para sequência de Medo	52
5.15	Sequência com Marcos Faciais para Alegria	53
5.16	Criação de vídeo através de sequência Normalizada	53

Lista de Tabelas

2.1	Action Units com os respetivos movimentos dos músculos faciais.	9
-----	---	---

Capítulo 1

Introdução

1.1 Contexto e Motivação

Que alterações acontecem na nossa cara quando estamos felizes ou com medo? Terão os olhos, boca ou outro elemento da face humana um peso igual na forma como expressamos a emoção que estamos a sentir? A partir de que momento da sua expressão é que reconhecemos determinada emoção nos outros? O estudo sobre emoções e suas respectivas expressões faciais é um campo de investigação com grande actividade. A necessidade de compreender e interpretar expressões emocionais tem um papel importante na comunicação e comportamento social. Que comportamento tem uma pessoa que vê outra com cara de medo? Será que a percepção de medo noutra pessoa pode influenciar a forma como agimos ou reagimos? São exemplos onde se destaca a clara importância na compreensão e interpretação das emoções expressadas perante determinadas situações.

O estudo destes aspectos em laboratório implica a consideração de estímulos emocionais que possam ser apresentados a pessoas para, por exemplo, no âmbito de um variado número de paradigmas experimentais, aferir desempenho, medir tempos de reacção, e alterações fisiológicas.

Dada a importância dos estímulos considerados e a necessidade de os poder utilizar, de forma sistemática, no âmbito de trabalhos de investigação, vários trabalhos existentes criaram e validaram bases de dados que consistem em imagens estáticas das expressões faciais de diferentes actores para diferentes emoções [25, 12, 26]. As bases de dados existentes são normalmente baseadas na teoria de emoções básicas segundo Paul Ekman [9] assumindo a existência de 6 emoções básicas (raiva, medo, nojo, surpresa, alegria e tristeza).

1.2 Desafios

As diferentes bases de dados disponíveis na literatura apresentam, no entanto, algumas limitações que impossibilitam um estudo mais amplo da expressão de emoções. Em primeiro lugar, as bases de dados são construídas pedindo a participantes que expressem a emoção que

lhes é pedida, recebendo algumas instruções de como o fazer (p.e. que músculos da cara ativar). Neste contexto, ainda que se assegurem todos os aspectos prototípicos de expressão de cada emoção, relevantes para alguns contextos experimentais, torna-se complicado perceber o impacto que esta metodologia pode ter na naturalidade da expressão e na intensidade (eventualmente exagerada) de alguns elementos (e.g., boca, olhos). Para perceber este impacto e aproximar as expressões faciais do que se observa, no dia-a-dia, seria importante dispor de estímulos em que a emoção tivesse sido expressa de forma espontânea. Isto permitiria ter estímulos com maior validade ecológica.

No entanto, um dos aspectos que mais impacto tem na forma como podemos estudar a expressão facial de emoção relaciona-se com o carácter estático das bases de dados existentes. A expressão da emoção é um fenómeno inerentemente dinâmico e é dessa forma que o observamos no nosso dia-a-dia, isto é, uma emoção não surge no seu pico de intensidade. Se quisermos estudar como essa dinâmica influencia a nossa percepção de emoções (e.g., a nossa capacidade de identificar uma emoção em particular) e a importância de cada elemento da face, em cada momento, para essa percepção, precisamos de olhar, não só para o pico de expressão, mas para toda a sequência desde uma expressão neutra, passando pelo pico de expressão e terminando na expressão neutra que se lhe siga. Adicionalmente, as expressões que observamos podem abarcar um leque alargado de intensidades, e poder estudar como diferentes níveis de intensidade de cada emoção são percebidos pode lançar pistas sobre como processamos a face para extrair essa informação. Será que são os olhos ou a boca que mais nos revelam a emoção que se está a desencadear?

Tendo estes aspectos em consideração, e a título de exemplo em trabalhos recentes, Manuel G. Calvo [11] apresentou uma base de dados dinâmica de expressão de emoção. No entanto, as sequências dinâmicas foram obtidas por *morphing*, i.e., interpolações entre uma imagem estática neutra e uma imagem da expressão máxima de cada emoção. Enquanto isto pode ser uma primeira aproximação que permita estudar a influência de diferentes intensidades na expressão de emoção, ao interpolar a evolução dos elementos da face, impossibilita a análise da sua progressão, já que aplica uma evolução linear, a todos eles, mascarando diferenças de sincronismo e coordenação dos elementos faciais.

1.3 Objetivos

Tendo em conta em conta os desafios identificados, torna-se evidente a necessidade de se construir uma base de dados em que seja possível avaliar uma emoção de forma dinâmica, capturando todos as sequências de emoções por *acting* normal, *acting* com instruções e espontaneamente, para realizar uma possível comparação, e por fim extrair/quantificar todos os elementos possíveis da face (p.e. altura das sobrancelhas). Teremos assim para análise três fases diferentes durante a experiência: normal *acting* (os participantes têm que expressar uma

emoção) e *acting* com instruções (é explicado aos participantes como proceder para expressar uma emoção) e emoção espontânea (o participante reage à emoção de forma espontânea).

Tendo isto em conta, o objetivo principal é contribuir em métodos e recursos que permitam melhorar a forma como a investigação sobre expressão facial de emoção é efectuada.

Para alcançar este objetivo, pretende-se:

- Perceber as abordagens e limitações da literatura já existente;
- Propor, em conjunto com o Departamento de Educação e Psicologia procedimentos experimentais que permitam obter dados dinâmicos de expressões faciais de emoções;
- Desenvolver métodos de aquisição e validação de dados que permitam, ao implementar os procedimentos experimentais, construir uma BD de expressões faciais de emoções;
- Propor e testar métodos que permitam obter medidas que quantifiquem a actividade facial, ao longo do tempo, para suportar a validação dos dados obtidos e futuros estudos dinâmicos.

1.4 Estrutura do Documento

O restante conteúdo deste documento está organizado em quatro capítulos. De uma forma sucinta o capítulo 2, Estado de Arte, refere-se ao trabalho que já existe nesta área de estudo, nomeadamente protocolos, base de dados existentes e a sua criação, depois de analisada a literatura, segue-se o capítulo 3, Protocolo Experimental e Requisitos, onde é explicado qual o procedimento e enumerados todos os requisitos do sistema que será implementado. Os últimos dois capítulos demonstram todo o processo na criação da base de dados, no capítulo 4, Sistema computacional de suporte à aquisição de uma base de dados emocional, é explicado todo o processo inicial de aquisição dos dados necessários e o por fim o capítulo 5, Métodos computacionais para pós-processamento, explica o métodos utilizados para processamento dos dados adquiridos anteriormente.

Capítulo 2

Estado da Arte

Este capítulo tem como objectivo proporcionar um enquadramento geral sobre a investigação em torno da expressão facial da emoção e focando, principalmente, a área da Psicologia. Ainda que o estudo da expressão emocional (e constituição de bases de dados para esse fim) tenha relevância e aplicações em outra áreas, por exemplo a de desenvolvimento de métodos computacionais para reconhecimento de emoção, o intuito foi o de focar nos aspectos relevantes para a criação de uma base de dados que sirva a investigação em Psicologia e como a questão tem sido tratada dessa perspectiva. Assim, nas secções seguintes é possível analisar essas aplicações, os protocolos existentes na Psicologia, e de que forma as bases de dados e suas características contribuem para o estudo nesta área de investigação.

2.1 Estudos Envolvendo Expressão Facial de Emoção

O face humana tem propriedades particulares, que permitem transmitir mensagens, afetos e emoções, etc. Paul Ekman, com base no trabalho de Charles Darwin (1872 / 1998) conclui que existe uma relação entre o comportamento da face humana e as emoções [13, 16, 17, 18, 20, 21, 22]. Assim, Ekman com as suas pesquisas confirmou que as emoções humanas têm expressões faciais universais, não apenas transculturais, mas como também partilhadas com algumas espécies de animais. Quase um século depois confirmou-se o que Darwin tinha proposto.

Os estudos transculturais realizados por Ekman avaliaram o reconhecimento de emoções através de poses de expressões faciais, evidenciou-se então que há unanimidade em sete emoções básicas: medo, raiva, tristeza, felicidade, surpresa, nojo e desprezo [9].

Estas descobertas sobre a face humana, abriram um vasto horizonte de possibilidades para a investigação e exploração sobre emoções no comportamento facial.

A maioria dos autores concorda que a emoção é uma reacção a um estímulo ou acontecimento externo, ou um estado de sentimento que leva a uma possível resposta psicológica.

As expressões faciais de emoções universais são provavelmente inatas ao logo dos anos, ao

contrário de gestos que fazem parte do comportamento corporal e cultural. Apesar da unanimidade na expressão das emoções básicas, há elementos físicos na face humana que variam de cultura para cultura. As evidências da universalidade das expressões faciais de emoções abriram um vasto horizonte de possibilidades para a pesquisa deste ramo, no entanto, impuseram desafios metodológicos para o seu estudo. Nesse sentido o primeiro método que permite a caracterização dos movimentos dos músculos da face foi desenvolvido em 1878, por Paul Ekman e Wallace Friesen [19], designado de *Facial Action Coding System (FACS)*. O sistema deles consiste em medir todos os diferentes movimentos faciais e codificar como é que esses movimentos dos músculos faciais influenciam a aparência da face. Os autores estudaram anatomia e encontram as associação entre a acção dos músculos, as mudanças da aparência da face. Algumas das a mudanças da aparência são o resultado de múltiplos músculos em que alguns deles podem ter mais que uma acção. A AUs (action units) são as acções realizadas por cada músculo ou combinação. O FACS possui 46 AUs (12 para a parte de cima da face, 18 para a parte de baixo, AU 1 até à 7 refere-se as sobrancelhas, testa e pálpebras, ver Tabela 2.1. As seis emoções básicas detectadas pelo FACS são a raiva, nojo, medo, felicidade, tristeza e surpresa. Foram feitas duas importantes descobertas utilizando este método, uma que contribui para a interpretação e deteção de fraudes (mentiras), através de músculos da face que podem ser facilmente manipulados, no entanto estão presentes na expressão espontânea de algumas emoções. Torna assim a avaliação da espontaneidade de uma expressão facial importante, porque analisa-se o tempo de reacção, duração e os músculos envolvidos na expressão, que sem ignorar outros factores que podem justificar as alterações do comportamento da face, podem dar-nos pistas sobre a sua interpretação. A segunda descoberta é o ato de imitar uma expressão facial de uma emoção, activa um processo fisiológico correspondente a determinada emoção. A validação desta hipótese, pode abrir um novo campo de investigação e técnicas de treino de gestão de emoções, podendo numa larga escala ter fins terapêuticos e ajudar no bem estar das pessoas no geral.

O conhecimento sobre expressões faciais é evidente em várias aplicações, como por exemplo: avaliação psicológica, inquéritos policiais, apoio, avaliação clínica, marketing, vendas, pesquisa e industria televisiva. Nos inquéritos policiais, as fraudes transmitem uma carga emocional maior que em situações normais. Contudo, as micro-expressões não nos dizem se o sujeito está a mentir, mas mostram que há divergências entre o que é dito e aquilo que o sujeito experiencia emocionalmente. Em todo o mundo, as polícias de vários países têm treino de percepção de linguagem corporal e reconhecimento de micro-expressões para poderem avaliar o comportamento do sujeito inquirido.

O interesse pelas expressões faciais na Psicologia, não deve ser só encarado pela necessidade de detectar fraudes, mas sim, pelo objetivo de acreditarmos que a Psicologia deve munir o balanço emocional das pessoas. Dessa forma, é importante perceber porquê temos emoções, que ferramentas podem melhorar a vida emocional das pessoas, como reconhecer as emoções

noutras pessoas e como lidar com essas emoções.

2.1.1 Expressões Faciais de Emoções

A compreensão das expressões faciais tem um papel importante na ênfase da comunicação, e conseqüentemente, na leitura emocional das pessoas. A análise da expressão de emoção, desta forma, é um ponto chave dessas características faciais. O nosso corpo está preparado para sentir emoções. Sentir tristeza, raiva ou alegria são repostas naturais a experiências que o nosso cérebro processa no dia a dia. Durante a história, as emoções tiveram um grande papel na evolução humana. Durante anos, o nosso cérebro desenvolveu emoções para ultrapassar certos eventos, de forma a salvaguardar espécies. Por esse motivo que, quando vemos algo que associamos como perigo, sentimos medo, o que nos leva a fugir ou lutar contra esse perigo. Podemos assumir que a expressão de emoção é o reconhecimento dessas emoções que o humano foi preparado para sentir. Naturalmente, a face humana é onde é possível ter essa percepção, vão existir alterações que nos associam a uma determinada emoção. Tomkins [1] ressuscitou as reivindicações de Darwin sugerindo que as emoções eram o alicerce das motivações humanas e que o pico dessas emoções eram na face humana. Tomkins realizou o primeiro estudo demonstrando que as expressões faciais eram fielmente associadas com determinados estados emocionais [2]. Mais tarde Tomkins recrutou Paul Ekman Carroll Izard para concluir o que hoje é conhecido como “estudos universais” ou transculturais (como é descrito na secção anterior).

Desde os originais “estudos universais”, muitos estudos examinando o comportamento de expressões faciais replicaram o reconhecimento universal de uma emoção na face humana. Para além disso um conjunto de bases de dados, avaliando o comportamento da emoção na face humana indicaram um baixo nível de mudança no reconhecimento universal de uma emoção [6]. Foi demonstrado também que essas mesmas expressões faciais são produzidas quando as emoções são elicitadas espontaneamente [3]. Estas descobertas foram importantes, visto que estes estudos foram feitos por vários investigadores de todo o mundo em laboratórios diferentes usando métodos diferentes com participantes de diferentes culturas, e todos convergiram para mesmo resultado. Portanto há uma evidência forte para a expressão facial universal de sete emoções: raiva, alegria, medo, surpresa, tristeza, nojo e desprezo, ver Figura 2.1.

As descobertas relativas a universalidade da expressão facial de emoção podem ajudar no desenvolvimento de empatia ou confiança. Podem ser úteis em tornar apreciações credíveis, avaliar veracidade e detectar enganos, e obtendo melhor informação sobre estados emocionais, fornece-se o básico para melhor cooperação e negociação. Com este propósito vários trabalhos foram realizados para analisar as características faciais para determinadas emoções, Ekman [4] conclui que as expressões faciais são primariamente geradas por contracção ou relaxamento dos músculos faciais. Isto causa uma mudança na localização dos pontos na face humana

(marcos faciais). A relação entre movimento dos músculos e emoção foi também estudada e definida pela *emotional facial actioncoding system* [8], onde, por exemplo, a emoção de alegria é representada pela combinação do “levantamento do queixo” e “levamento do canto da boca”.



Figura 2.1: As 7 emoções básicas (Alegria (*Joy*), Surpresa (*Surprise*), Desprezo (*Contempt*), Tristeza (*Sadness*), Raiva (*Anger*), Nojo (*Disgust*) e Medo (*Fear*)) e a sua Expressão Universal. Imagem adaptada de [3].

2.2 Protocolos Experimentais

Os estudos experimentais, em Psicologia, que envolvem o impacto da expressão de emoções nas pessoas e na sua capacidade, por exemplo, de desempenhar tarefas de forma rápida, fazem uso, tipicamente, de imagens de pessoas a expressar cada emoção. O trabalho de Kurakova [10], por exemplo, o participante teve que lembrar ou imaginar uma situação em qual ele experiencie 2 emoções básicas numa sequência, e expressar na sua face a transição entre essas emoções mantendo um olhar direto. A sua face é gravada com uma máquina fotográfica, num fundo neutro. O processo é repetido para seis emoções. A Figura 2.2 mostra um exemplo de sequência.

Na generalidade, os estímulos considerados nestes estudos, dadas as vantagens de não ter de os criar de raiz, e de os utilizar de forma sistemática, são obtidos de bases de dados de expressões de emoção já existentes, o que facilita a validação e replicação dos estudos. Essas bases de dados são descritas de seguida.

AU	Nome FACS	Músculo
1	Levantador de Sobrancelha Interna	Frontalis (pars medialis)
2	Levantador de Sobrancelha Externa	Frontalis (pars lateralis)
4	Abaixador de Sobrancelha	depressor glabellae, depressor supercilii, corrugator supercilii
5	Levantador de Pálpebra Superior	Levator pálpebra superior
6	Levantador de Bochechas	Orbicularis oculi (pars orbitalis)
7	Apertador de Pálpebra	Orbicularis oculi (pars palpebralis)
9	Enrugador de Nariz	Levator labii superioris alaeque nasi
10	Levantador de Lábio Superior	Levator labii superioris, caput infraorbitalis
11	Aprofundador Nasolabial	Zygomaticus minor
12	Puxador de Canto do Lábio	Zygomaticus major
13	Inchador de Bochecha	Levator anguli oris (also known as Caninus)
14	Fazedor de Covinhas	Buccinator
15	Depressor de Canto do Lábio	Depressor anguli oris (also known as Triangularis)
16	Depressor do Lábio Inferior	Depressor labii inferioris
17	Levantador de Queixo	Mentalis
18	Franzedor de Lábio	Incisivii labii superioris and Incisivii labii inferioris
19	Língua para Fora	
20	Esticador de Lábio	Risorius w/ platysma
21	Endurecedor de Pescoço	Platysma
22	Afunilador de Lábio	Orbicularis oris
23	Endurecedor de Lábio	Orbicularis oris
24	Pressionador de Lábio	Orbicularis oris
25	Separador de Lábios	Depressor labii inferioris ou relaxation do Mentalis, ou Orbicularis oris
26	Queda de Mandíbula	Masseter, Temporalis relaxada e pterygoid interno
27	Esticção da Boca	Pterygoids, Digastric
28	Sucção de Lábios	Orbicularis oris
29	Projeção de Mandíbula	
30	Movimentação Lateral da Mandíbula	
31	Jaw Clencher	Masseter
32	Mordida do Lábio	
33	Inflar de Bochecha	
34	Bufar de Bochecha	
35	Sucção de Bochecha	
36	Arqueamento da Língua	
37	Limpeza do Lábio	
38	Dilatador das Narinas	
39	Compressor das Narinas	
41	Abaixamento da pálpebra	
42	Fenda	
43	Olhos Fechados	Relaxamento do Levator palpebrae superioris; Orbicularis oculi (pars palpebralis)
44	Olhos Semicerrados	
45	Piscar	Relaxamento do Levator palpebrae superioris; Orbicularis oculi (pars palpebralis)
46	Piscada	Relaxamento do Levator palpebrae superioris; Orbicularis oculi (pars palpebralis)

Tabela 2.1: Action Units com os respectivos movimentos dos músculos faciais.



Figura 2.2: Exemplo de uma transição da emoção nojo para alegria, com respectivos marcos faciais. Imagem adaptada de [10].

2.3 Bases de Dados Existentes

Existem bases de dados para reconhecimento e detecção de expressões faciais criadas por vários investigadores e organizações, que podem ser usadas para o estudo de emoções. Existem diferentes regras e requisitos no que toca ao uso delas. Em particular, a maior parte das bases de dados estão apenas disponíveis para propósitos não comerciais.

Em seguida serão selecionadas algumas dessas bases de dados que são usadas maioritariamente para estudo de emoções.

2.3.1 Karolinska Directed Emotional Faces (KDEF)

Uma das mais conceituadas bases de dados existentes, é a Karolinska Directed Emotional Faces. Consiste num conjunto de 4900 imagens de expressões de emoção de faces humanas. Foi desenvolvida originalmente para ser usada no estudo psicológico e médico. Desenvolvido em 1998 por Daniel Lundqvist, Anders Flykt, o Professor Arne Öhman no Instituto Karolinska, Departamento de Neurociência clínica, secção de psicologia, Estocolmo, Suécia [25].

Consiste num conjunto de 70 pessoas, cada uma expressa 7 emoções diferentes, sendo que cada uma delas é fotografada duas vezes em 5 ângulos diferentes, ver Figura 2.3.

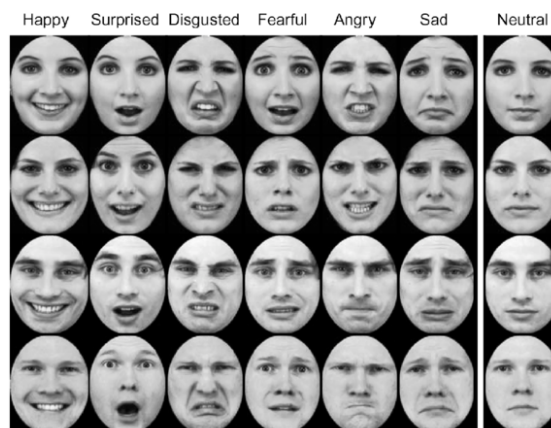


Figura 2.3: Exemplo das imagens de expressões faciais usadas na base de dados KDEF. Imagem adaptada de [25]

2.3.2 I BU-3DFE (Binghamton University 3D Facial Expression) Database (Static Data)

Desenvolvida na Universidade de Binghamton, esta base de dados de expressões faciais em 3D consiste num conjunto de 100 pessoas exprimindo 2500 expressões faciais. Está disponível para a comunidade de investigação (computação, computação visual, interação humano-computador, segurança, biomedicina e psicologia) [26].

Das 100 pessoas 56 são mulheres e 44 homens, com idades compreendidas entre os 18 e 70 anos, com várias etnias.

Cada pessoa expressou 6 expressões (alegria, nojo, raiva, surpresa, medo e tristeza) através de um scanner 3D. Com a exceção da expressão neutra, cada uma das 6 expressões inclui 4 níveis de intensidade. Assim cada pessoa tem 25 modelos de expressões 3D estáticas. Ver Figuras 2.4, 2.5 e 2.6.

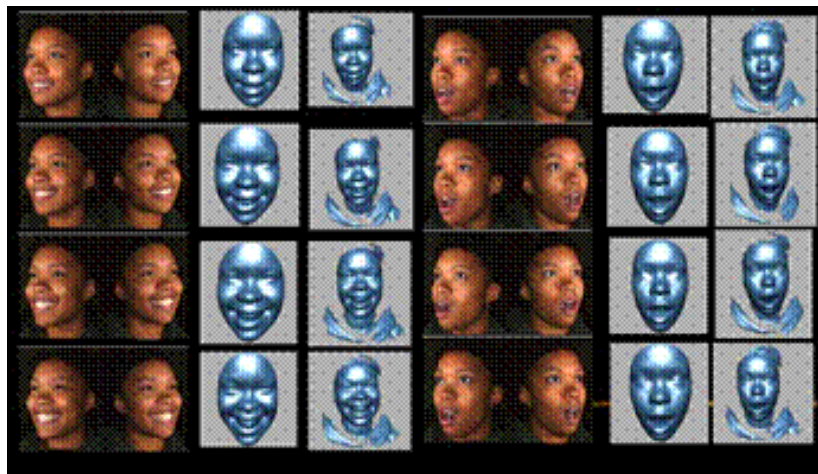


Figura 2.4: Quatro níveis de expressões faciais do mais baixo para o mais alto e respetivos modelos. Imagem adaptada de [26].

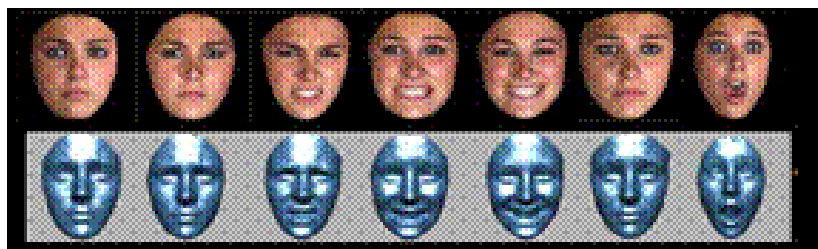


Figura 2.5: Sete expressões femininas (neutro, raiva, nojo, medo, alegria, tristeza e surpresa) e respetivos modelos. Imagem adaptada de [26].

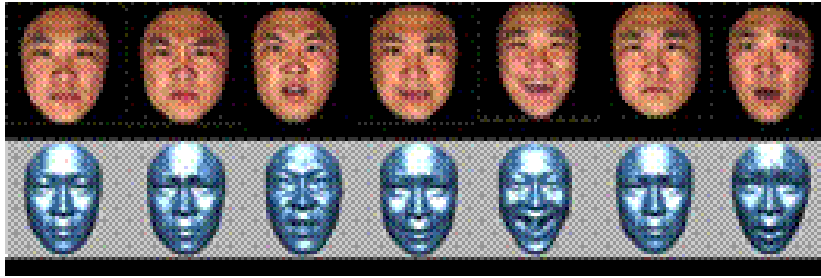


Figura 2.6: Sete expressões masculinas (neutro, raiva, nojo, medo, alegria, tristeza e surpresa) e respectivos modelos. Imagem adaptada de [26].

2.3.3 II BU-4DFE (Binghamton University 3D + time): A 3D Dynamic Facial Expression Database (Dynamic Data)

É uma base de dados estendida da anterior, que consiste na análise do comportamento entre o espaço estático e o dinâmico [27].

Foi criada uma base de dados nova com expressões faciais 3D dinâmicas, disponível para toda a comunidade científica de investigação. As expressões foram capturadas com um *frame rate* de 25 frames por segundo. Cada pessoa representa 6 sequências de expressões para cada uma das 6 expressões faciais. Cada sequência contém 100 frames. A base de dados contém 606 modelos de sequências de vídeo 3D, capturadas de 101 pessoas, com o total de aproximadamente 60600 frames, ver Figuras 2.7 2.8 e 2.9.

Consiste em 58 mulheres e 43 homens de várias etnias.

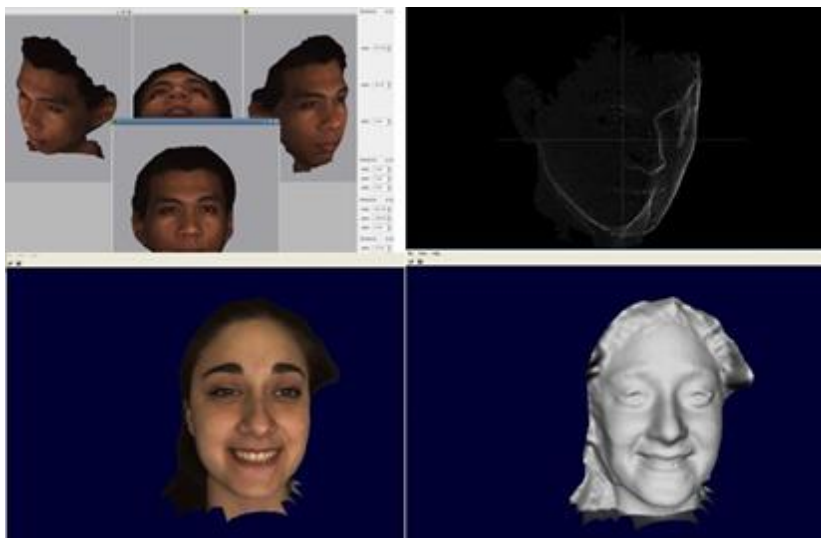


Figura 2.7: Visualização dos modelos de sequências de vídeo 3D. Imagem adaptada de [27].



Figura 2.8: Sete expressões masculinas (neutro, raiva, nojo, medo, alegria, tristeza e surpresa e respetivos modelos. Imagem adaptada de [27].

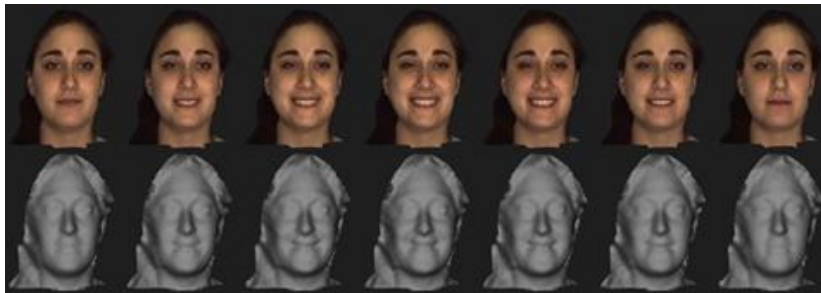


Figura 2.9: Sete expressões femininas (neutro, raiva, nojo, medo, alegria, tristeza e surpresa e respetivos modelos. Imagem adaptada de [27].

2.3.4 III. BP4D-Spontaneous: Binghamton-Pittsburgh 3D Dynamic Spontaneous Facial Expression Database

Expressões faciais representadas ou não representadas (espontâneas) podem diferir em vários aspetos como, por exemplo, a sua complexidade e tempo. Por isso a importância de criar um base de dados com vídeos espontâneos é necessária. Esta base de dados foi criada com esse intuito. Consiste assim numa BD de vídeo de expressões faciais espontâneas [28].

Foram usadas instruções validadas para elicitare expressões de uma emoção. As acções faciais for obtidas através do FACS. As características faciais foram monitorizadas em domínios 2D e 3D usando uma abordagem genérica ou pessoa específica.

O objetivo deste trabalho é explorar as características 3D em subtis expressões faciais, perceber a relação entre representar/posar e movimento dinâmico em action units faciais e perceber a ação natural que ocorre na face.

A base de dados inclui 41 participantes (23 mulheres e 18 homens). Idades entre 18-29. O protocolo de elicitação da emoção foi desenvolvida eficazmente para elicitare as emoções nos participante através de uma entrevista com oito tarefas e uma série de actividades para elicitare as emoções.

A base de dados está estruturada por participantes. Cada um deles está associado com as 8 tarefas. Cada tarefa tem dois vídeos 2D e 3D. Inclui ainda dados sobre as AUs (FACS AU). Ver figura 2.10.

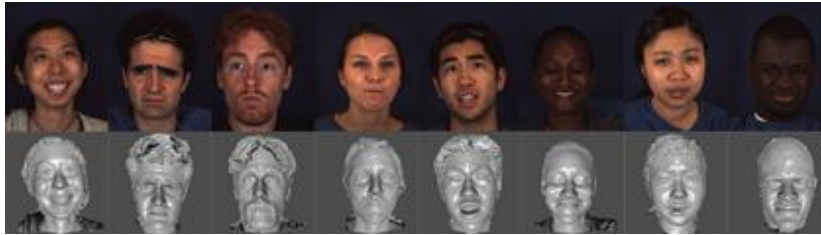


Figura 2.10: Expressões para as 8 tarefas. Imagem adaptada de [28]

2.3.5 The Bosphorus Database

Base de dados criada para investigação em processamento facial humano 2D e 3D, incluindo reconhecimento facial, detecção de AUs faciais, estimar intensidade de AUs faciais, reconhecimento faciais sobre condições adversas, modelagem facial e reconstrução facial 3D.

Existem 105 pessoas e 4666 faces na base de dados, e é única em três aspectos: (ver Figura 2.11).

1. Reportório rico de expressões:
 - Até 35 expressões por pessoa;
 - FACS scoring (inclui intensidade para cada AU);
 - Um terço das pessoas são profissionais/atores.
2. Variação da postura da cabeça;
3. Tapar elementos da face (barba e bigode, cabelo, mãos, óculos);

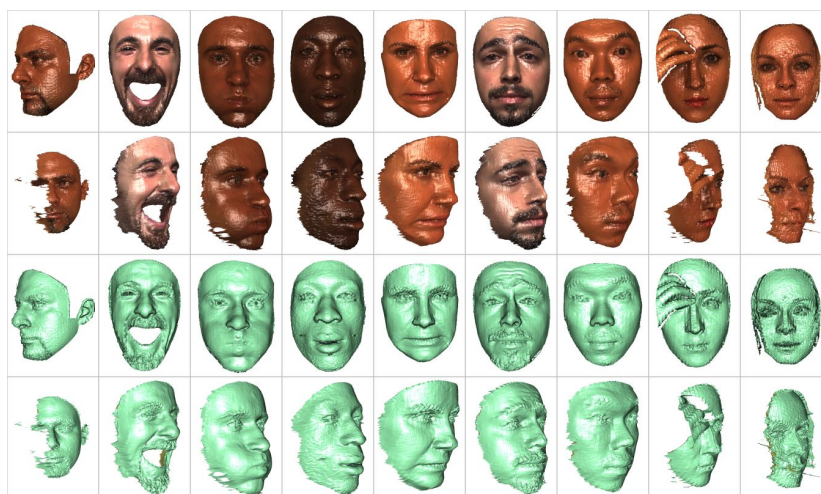


Figura 2.11: Exemplo com e sem textura, ambas de frente e 45°.

2.4 Fases da Criação de uma Base de Dados Emocional

A criação das bases de dados descritas na sessão anterior necessitam de um suporte computacional abrangente. A aquisição de uma foto, é realizada, na maioria das bases de dados, através de uma máquina fotográfica ou de um qualquer outro dispositivo que capture imagem, como por exemplo a Microsoft Kinect, que será descrita com maior atenção nas secções seguintes. Após a aquisição dos dados é necessário que sejam transferidos para um computador para processamento e normalização dos mesmos, normalmente utilizando aplicações como Adobe Photoshop. Por fim, torna-se muitas vezes relevante o cálculo de medidas das características das expressões faciais capturadas para, por exemplo, validar a base de dados.

Nas secções seguintes é feita uma descrição sucinta dos principais aspectos envolvidos na criação destas bases de dados.

2.4.1 Aquisição

Estas bases de dados são criadas através de métodos de gravação da face de uma pessoa para uma determinada emoção. Para a gravação é necessário que aspectos como iluminação, *background* ou distância da pessoa para o dispositivo de gravação sejam tomados em conta. É importante que todas as imagens sejam uniformes para que o processamento das mesmas sejam mais eficiente. Estas bases de dados, por norma estão estruturadas por participantes e emoções.

O equipamento utilizado para a gravação das fotos na base de dados KDEF [25] foi uma câmara Pentax LX, 3 lâmpadas de 500W e uma lente de 135 mm. Todos os participantes tiveram que estar a um distância de 3 metros da câmara, sendo que as faces dos participantes foram ajustadas a uma grelha existente no ecrã da câmara, já a base de dados *The Bosphorus Database* utilizou um lâmpada de halogéneo de 1000W numa sala escura para obter uma luminosidade homogénea para todas as imagens que foram capturadas o sensor Inspeck Mega Capturor II 3D, em que o participante se tinha que colocar a 1.5 metros de distância.

Outros estudos utilizaram o sensor Microsoft Kinect para a aquisição de imagens faciais. Entre eles está uma base de dados criada no âmbito da computação visual, FaceWarehouse [5] que utiliza a câmara RGB do sensor para capturar todas as imagens.

2.4.1.1 Microsoft Kinect

O Microsoft Kinect é um dispositivo de entrada multimédia muito versátil que pode ser usado como scanner 3D. Originalmente lançado como uma extensão à consola de jogos Xbox. Permitia controlo de movimentos usando um sistema de câmaras dedicado sem a necessidade de controlos remotos.

Existem 4 tipos de dispositivos Kinect:

1. Kinect for Xbox 360

2. Kinect for Windows
3. Kinect for Xbox One
4. Kinect for Windows v2

2.4.1.2 Sensor Microsoft Kinect for windows V2

Para este trabalho será utilizado o quarto tipo, figura 2.12. Com este dispositivo é possível detectar pontos faciais característicos num sistema de coordenadas 3D. As expressões faciais derivadas da atividade de um músculo específico são descritas com coeficientes especiais, denominadas de AU (Action Units), medidas usada pelo método de caracterização física de expressão de emoções chamado Facial Action Coding System (FACS).

O dispositivo possui um sensor de profundidade, uma câmara RGB, emissor de infravermelhos e microfone. Tem uma maior profundidade, fidelidade que a versão anterior o que torna a visualização e desempenho mais significativa.

A Microsoft fornece um kit de desenvolvimento (SDK 2.0) com novas funcionalidade, drivers, ferramentas, interfaces e muitos exemplos de código em C#, C++ e JAVA para ajudar os desenvolvedores. Com este novo SDK é possível rastrear até 6 pessoas em simultâneo.



Figura 2.12: Sensoe Microsoft Kinect v2.

2.4.1.3 Características Principais

O sensor possui uma câmara RGB com uma resolução de 1920x1080 (16:9) e um frame rate de 60 fps. Uma sensor de profundidade (em 3D) com uma resolução de 512x424. O software utilizado é o Kinect Studio v2 e a ferramenta Face Tracking SDK.

2.4.1.4 Action Units do sensor Microsoft Kinect

O sensor Microsoft Kinect consegue captar 17 Action Units na face humana, como podemos verificar na Figura 2.13 em que 14 delas são expressas com um valor entre 0 e 1, e os restantes 3 (*Jaw Slide Right*, *Right Eyebrow Lowerer* e *Left Eyebrow Lowerer*,) variam entre -1 e $+1$, representando o deslocamento de uma AU de uma expressão neutra. Podem ser representados num vector da seguinte forma:

$$a = (AU0, AU1, AU2...AU16)$$

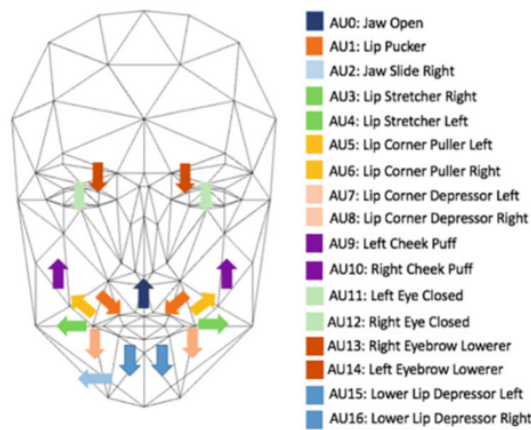


Figura 2.13: 17 Action units extraídas da Kinect v2 e o movimento do músculo correspondente.

Através da Figura 2.14 verificamos que por exemplo na emoção de Alegria (*Joy*) os valores do AU5 e AU6, que correspondem ao levantamento do lábio esquerdo e direito, os valores estão muito próximos de 1, que representa o sorriso da pessoa.

2.4.1.5 Funcionalidades do Microsoft Kinect

O sensor Microsoft Kinect é um dispositivo versátil que, ligado a um computador com Windows, oferece um conjunto de funcionalidades. O Microsoft Face Tracking Software Development Kit para a Kinect (Face Tracking SDK), juntamente com o Software Development Kit (Kinect for Windows SDK), permite-nos criar aplicações que podem rastrear faces em tempo real.

O Face Tracking SDK consegue, através da câmara detectar a posição da face e as suas características dependendo dos pontos que foram rastreados. É possível detectar 87 pontos

	neutral	joy	surprise	anger	sadness	fear	disgust
AU0	0.045	0.160	0.337	0.088	0.091	0.229	0.197
AU1	0.231	0.000	0.347	0.216	0.392	0.246	0.289
AU2	0.057	-0.019	0.018	-0.055	0.001	-0.067	-0.037
AU3	0.133	0.204	0.010	0.101	0.052	0.015	0.022
AU4	0.115	0.161	0.032	0.101	0.151	0.188	0.030
AU5	0.000	0.820	0.023	0.020	0.010	0.024	0.084
AU6	0.015	0.903	0.023	0.025	0.014	0.002	0.013
AU7	0.040	0.056	0.174	0.118	0.421	0.241	0.116
AU8	0.044	0.080	0.134	0.036	0.437	0.277	0.193
AU9	0.034	0.022	0.019	0.026	0.035	0.020	0.022
AU10	0.031	0.010	0.016	0.018	0.025	0.021	0.024
AU11	0.201	0.179	0.047	0.132	0.055	0.083	0.304
AU12	0.216	0.127	0.033	0.044	0.046	0.034	0.366
AU13	0.340	-0.049	-0.246	0.156	-0.118	-0.201	0.117
AU14	0.371	0.063	-0.221	0.376	-0.025	-0.085	0.258
AU15	0.000	0.550	0.036	0.045	0.017	0.001	0.137
AU16	0.000	0.482	0.026	0.044	0.018	0.005	0.114

Figura 2.14: Valores numéricos das AUs correspondentes a várias expressões.

2D na face e 13 outros pontos, referentes aos cantos da boca, centro de cada olho, centro do nariz e a delimitação da cabeça.

Os 87 pontos são: 16 pontos para os olhos (0-15, 8 para cada olho) 20 pontos para as sobrancelhas (16-35, 10 para cada sobrancelha) 12 pontos para o nariz (36-47) 20 pontos para o lábios (48-67, 12 para os lábios exteriores e 8 para os interiores) 19 pontos para o queixo (68-86).

Estes pontos são representados através de um array.

2.4.2 Processamento

A criação de um conjunto de estímulos de expressão facial de emoção exige, para que estes possam ser usados em contexto experimental, que se assegurem algumas características dos estímulos de forma transversal. Este princípio pretende uniformizar, entre estímulos, todas as características que não influenciem a sua função, i.e., fornecer expressões, de modo a que estas não possam resultar em enviesamentos da percepção por parte dos participantes. Exemplos simples são o de garantir uma área de ocupação da imagem que seja equilibrada entre actores e uma localização das diferentes regiões da cara (e.g., olhos, nariz, boca) em zonas pré-estabelecidas da imagem. Neste contexto, depois da aquisição das imagens são aplicados métodos de processamento que permitam normalizar as características de interesse.

No que concerne às bases de dados apresentadas na secção anterior, destacam-se procedimentos de normalização e pós-processamento. Por exemplo, no caso da base de dados KDEF [25] as faces foram ajustadas novamente a uma grelha digital e cortadas a um tamanho de 562 pixels de largura e 762 pixels de comprimento, utilizando o software Adobe Photoshop 4 para o respectivo processamento. Todas as fotos foram divididas em directórios específicos

baseados no participante, género e emoção. No caso da base de dados *The Bosphorus Database* destaca-se o procedimento manual da localização dos pontos fulcrais na face. Em cada imagem capturada da face são marcados 24 pontos. No caso referido na sub-secção anterior, para a base de dados FaceWarehouse [5] é utilizado o algoritmo Microsoft Fusion para reduzir o ruído das imagens, agregando múltiplas imagens. O Microsoft Fusion funciona unindo todos os dados do sensor de profundidade num único modelo, na qual é depois aplicado o algoritmo de *Ray Tracing* para gerar um mapa de profundidade para um frame escolhido.

Outras normalizações podem ainda ser efectuadas no que respeita a propriedades de baixo-nível das imagens (e.g., intensidade média e contraste), de modo a garantir que estas são iguais ao longo de todos os estímulos.

2.4.3 Caracterização Quantitativa dos Estímulos

Depois de os estímulos terem sido normalizados, torna-se importante reunir um conjunto de medidas caracterizadoras que podem ser usadas quer para validação da normalização, quer como ponto de partida para análises de como diferentes características de configuração facial podem estar relacionadas com determinadas emoções (e.g., enrugamento da testa na expressão de raiva).

As medidas quantitativas das características das imagens e expressões podem ainda servir de base a outras áreas, por exemplo, para o desenvolvimento de métodos computacionais para reconhecimento de emoções (e.g., o trabalho de G. R. Vinnetha, C. Sreeji, and J. Lentin [14], que apresentam um método de reconhecimento facial 3D utilizando o sensor Kinect, onde os seus resultados mostraram que expressões como tristeza e nojo eram mais difíceis de reconhecer que outra ou de A. Youssef, S. F. Aly, A. Ibrahim, and A. Lynn [15], onde propuseram um sistema que tentava reconhecer expressões faciais usando o sensor 3D Kinect). No entanto, no contexto deste trabalho, serão fundamentalmente abordadas medidas de caracterização da expressão facial com base na anatomia.

Para além de características das imagens (e.g., intensidade média dos pixels) dois tipos de características têm sido considerados por alguns autores para estudar as bases de dados existentes: marcos faciais e activação de action units.

2.4.3.1 OpenFace

Ferramenta originalmente desenvolvida por Tadas Baltrušaitis [29] em colaboração com o CMU MultiComp Lab liderado pelo professor Louis-Philippe Morency Github. Alguns dos algoritmos originais foram criados pelo Rainbow Group, Cambridge University.

Esta ferramenta é capaz de realizar a deteção da face e os seus marcos, estimar posição da cabeça, reconhecimento as actions units da face e entre outras funcionalidades.

É possível através de uma simples *webcam* obtermos todos estes dados em tempo real, sem a necessidade de nenhum especialista de hardware, bem como utilizar sequências de imagens

ou vídeos normalizados.

2.4.3.2 Funcionalidades do OpenFace

- Detecção de marcos faciais [30]. Ver Figura 2.15
- Tracking de marcos faciais e posição da cabeça [30]. Ver Figura 2.16
- Reconhecimento das Action Unit [31]. Ver Figura 2.17



Figura 2.15: Detecção de marcos faciais utilizando a ferramenta OpenFace.

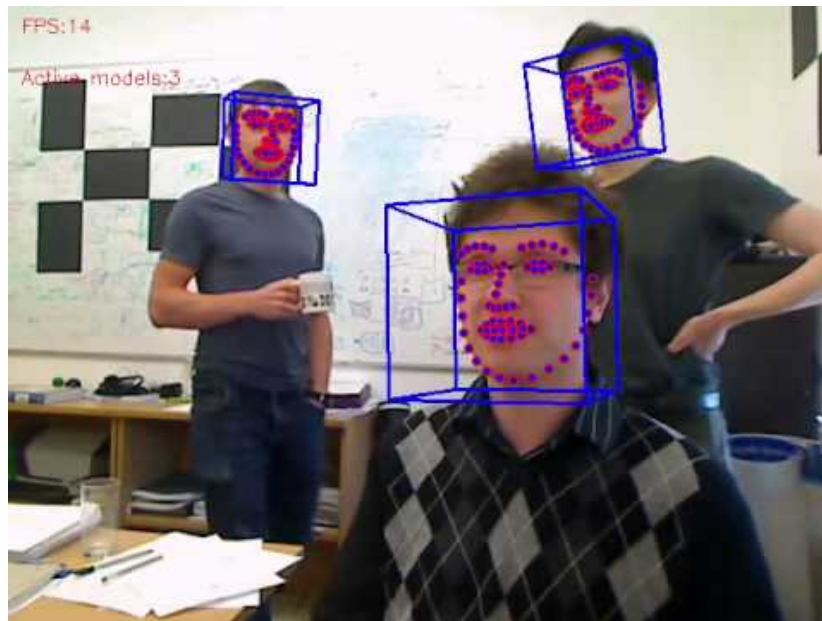


Figura 2.16: Tracking da cabeça utilizando a ferramenta OpenFace.

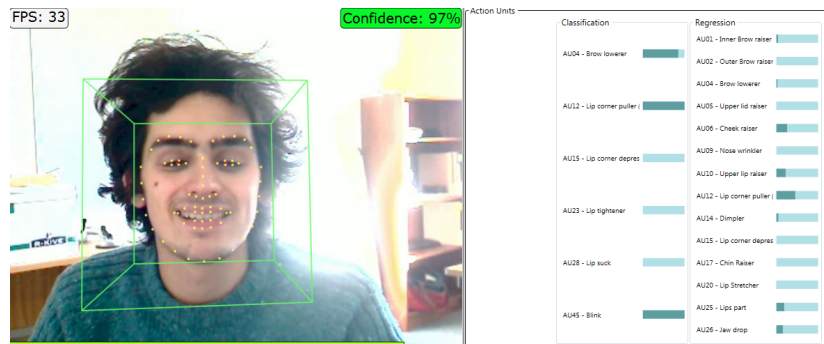


Figura 2.17: Reconhecimento das action units na face utilizando a ferramenta OpenFace.

2.5 Discussão

Da análise às bases de dados existentes de expressão facial de emoção, apresentada neste capítulo, identificam-se diferentes aspectos que devem ser abordados para melhorar suportar a investigação nesta área.

Ainda que os aspectos dinâmicos de expressão facial de emoção sejam relevantes enquanto objectos de estudo, existe apenas uma base de dados que disponibiliza esse tipo de dados, ainda que através de *morphing*. Este método para obter sequências da faces pode não ser totalmente preciso, isto porque, durante as transições, como por exemplo de Neutro para Alegria, o comportamento da expressão da face resulta de uma interpolação de pontos que não garante que a progressão das alterações da face corresponda ao real. A maior parte dos estudos nesta área usa imagens estáticas ou sequências deste género. Para além disso as bases de dados focam-se muito em expressões obtidas através de *acting*. Outro dos problemas é a falta de medidas objectivas sobre a actividade das expressões faciais para permitir estudos quantitativos.

Com isso, temos a necessidade de construir uma base de dados que suporte o estudo de aspetos dinâmicos de emoções de uma expressão facial em ambos contextos: espontâneo e de *acting*. E de contribuir para a quantificação de emoções, incluindo na base de dados um conjunto de novas características que descrevam a atividade facial.

Propor métodos que contribuam para abordar estes desafios contribui também com uma base que pode permitir uma evolução dos métodos computacionais de reconhecimento de emoções. Há uma grande necessidade de criar sistemas autónomos que permitam a análise e compreensão do comportamento facial, por exemplo, para adaptar sistemas à emoção expressa por um utilizador. Como em qualquer prática computacional alguns obstáculos têm de ser superados, de modo a que haja um maior aproveitamento do potencial dos métodos implementados. O facto de que as bases de dados de emoção, usadas em muitos destes estudos não serem naturais, pode conduzir a que algoritmos treinados nessas expressões vejam o seu desempenho penalizado quando aplicados a expressões naturais.

Capítulo 3

Protocolo Experimental e Requisitos

Após a análise da literatura e bases de dados existentes, estabeleceu-se a necessidade de criar uma base de dados dinâmica para expressão facial de emoção. De modo a que se possam definir os requisitos computacionais necessários para o efeito, foi desenhado, em colaboração com o Departamento de Educação e Psicologia da Universidade, um protocolo experimental para a aquisição dos dados. Os seus principais detalhes são apresentados, neste capítulo, e permitem estabelecer os requisitos para o trabalho a realizar no âmbito desta dissertação.

3.1 Conteúdo da Base de Dados

O principal objectivo é conter sequências dinâmicas de expressão facial de emoção, obtidas por instruções ou espontaneamente, seguindo uma metodologia experimental de forma sistemática e que assegure a sua utilidade para investigação em três níveis: estímulos visuais que possam ser usados no âmbito de protocolos experimentais; estímulos que permitam estudar as dinâmicas de expressão facial de emoção; e dados que suportem o desenvolvimento e teste de métodos computacionais de reconhecimento de emoção.

3.2 Protocolo Experimental

A construção de uma base de dados dinâmica para expressão de emoções implica a definição de um procedimento experimental que permita a obtenção, de forma controlada e replicável, dos dados necessários. Nesse sentido, foi desenhado, em conjunto com o Departamento de Educação e Psicologia da Universidade de Aveiro um procedimento experimental que deve ser assegurado para esse fim. Este, deverá ditar os requisitos para o(s) sistemas computacionais de suporte. Apresenta-se e explica-se, de seguida, o protocolo experimental a todos os participantes, através de um documento com a descrição do procedimento. O

próximo passo é a aquisição dos dados, onde os participante realizam a gravação das sequências para todas as fases nas duas sessões. Após a aquisição é necessário o processamento dos dados, isto é, normalizar e criar os vídeos das sequências juntamente com as medidas quantitativas que servirão de suporte à base de dados.

Terminados estes procedimentos temos tudo o que é necessário para a criação de uma base de dados dinâmica, isto é, criação de vídeos das expressões faciais para cada uma das emoções e respectivas medidas quantitativas.

3.2.1 Procedimento

Para a tarefa experimental serão avaliados, a considerar, cerca de 20 participantes (dividido em dois para cada género) com idades entre os 18 e 30 anos. Alguns fatores de exclusão que a ter em conta são o uso de óculos, maquilhagem, barba em excesso, joalheria bem como pessoas com psicopatologias e/ou alexitimia.

Os participantes são sujeitos a duas sessões em dias diferentes, em que têm que proceder a gravação de acordo com as três condições da experiência. A experiência é realizada em dias diferentes para que não haja um desfasamento entre as gravações de emoções, nomeadamente quando a gravação é espontânea e com instruções.

O procedimento consiste então em três fases, ou condições, distintas:

1. *Acting normal:*

Os participantes tem que expressar todas as seis emoções básicas (alegria, medo, surpresa, nojo e raiva), que irão aparecer num ecrã, uma por uma, por ordem aleatória. O participante deve expressar a emoção da forma mais natural que conseguir. O processo de gravação é repetido para cada uma das seis emoções e o seu início e fim será ativo pelo participante ao clicar na tecla esquerda do rato. No final de cada emoção o participante tem que avaliar a intensidade da sua performance através de uma escala visual analógica (de 0 a 100), que será aprovada ou não pelo experimentador. Se a sua avaliação for menor que 80, a expressão para essa emoção será novamente repetida. Caso for maior que 80 e aprovada pelo experimentador o participante passa à gravação da próxima emoção.

2. *Acting com instruções:*

Nesta fase usamos instruções como base no trabalho do Karolinska Directed Emotional Faces Database [25]. As instruções possuem a descrição das expressões faciais (músculos e movimentos envolvidos), para cada uma das seis emoções, para que os participantes possam pratica-las, por exemplo, em casa em frente ao um espelho. Estas

instruções são entregues a cada participante após a conclusão da primeira deslocação. processo de gravação é repetido para cada uma das seis emoções e o seu início e fim será ativo pelo participante ao clicar na tecla esquerda do rato. No final de cada emoção o participante tem que avaliar a intensidade da sua performance através de uma escala visual analógica (de 0 a 100), que será aprovada ou não pelo experimentador. Se a sua avaliação for menor que 80, a expressão para essa emoção será novamente repetida. Caso for maior que 80 e aprovada pelo experimentador o participante passa à gravação da próxima emoção.

3. Expressão emocional espontânea:

É reproduzido um vídeo com conteúdo emocional a cada participante com a intenção de eliciar a emoção. Os vídeos serão apresentados no ecrã em frente a eles. Irão ser usados apenas vídeo para eliciar alegria, medo e o estado neutro. O vídeo neutro será apresentado sempre primeiro que um dos outros vídeos para que o participante se sinta num estado neutro. O experimentador controla o início e fim da reprodução do vídeo e da respectiva gravação e é realizada apenas uma emoção em cada sessão. Repete-se o processo de rating.

3.2.2 Setup Experimental

A disposição dos diferentes equipamentos para a aquisição de dados foi organizado de acordo com a Figura 3.1. A experiência é realizada numa sala no Departamento de Psicologia da Universidade de Aveiro. O participante fica sentado numa cadeira, numa parte da sala, de frente para o sensor Microsoft Kinect e um monitor. De outro lado da sala, e fora da vista do participante de forma a não influenciar a sua performance, está o experimentador responsável pela recolha de dados, de frente para outro monitor e computador onde é executado o sistema computacional. Um dispositivo BIOPAC é usado, pela Psicologia, para adquirir dados psicofisiológicos durante a realização da experiência.

3.3 Pós-Processamento

Ao contrário das bases de dados em que os dados são estáticos ou interpolados, numa base de dados dinâmica, é necessário definir pontos chave nas sequências de expressão de emoção, isto é, definir onde se encontra o estado neutro no início da emoção, onde se atinge o pico e novamente o estado neutro no fim da expressão. O processo de anotação tem assim uma grande importância na criação de uma base de dados dinâmica, pois permite criar sequências com uma lógica temporal.

Como discutido anteriormente, é também importante agilizar o cálculo de medidas que permitam uma descrição mais quantitativa dos estímulos, e.g., marcos faciais, e que possa, no

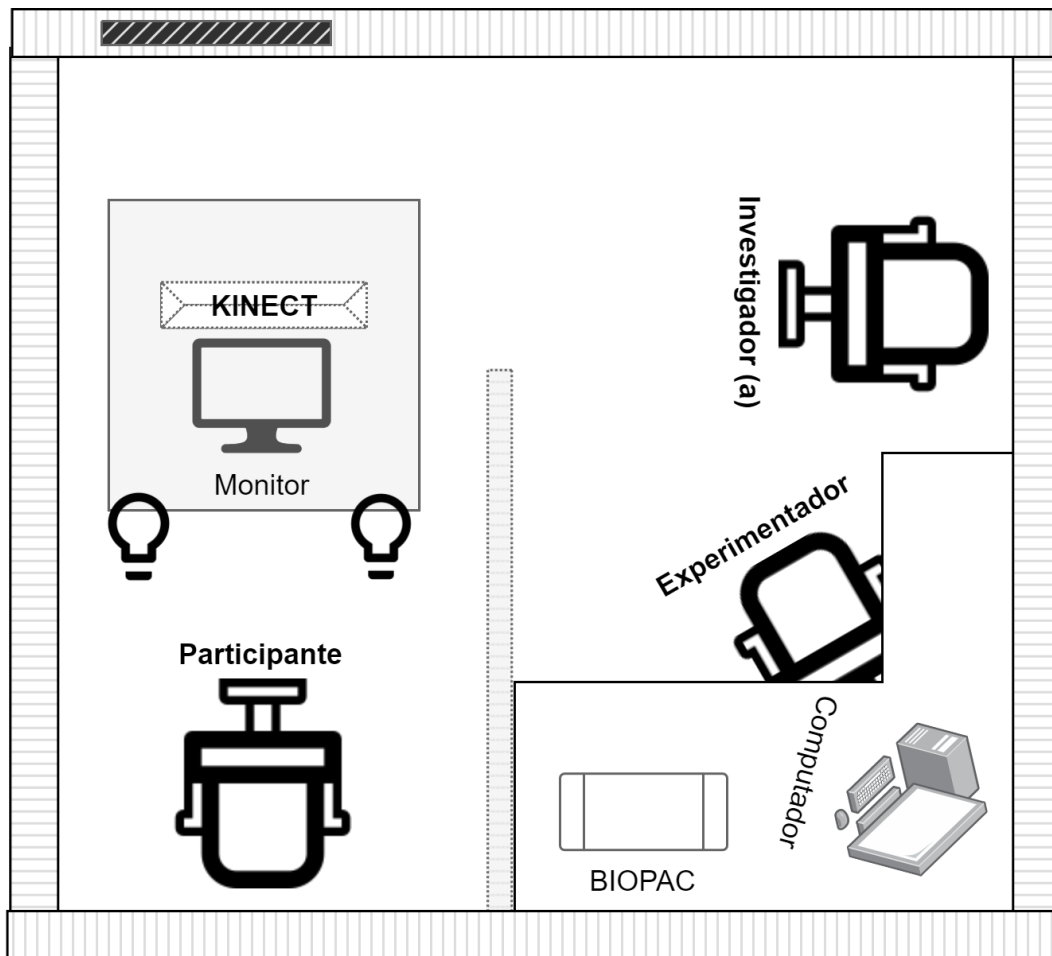


Figura 3.1: Disposição dos diferentes equipamentos e participantes para a configuração destinada à aquisição de dados experimentais. O experimentador fica fora da vista do participante de forma a não influenciar a sua performance.

futuro, ser adaptada para incluir outras medidas consideradas relevantes para serem aplicadas a toda a base de dados.

3.4 Requisitos

A lista de requisitos tem como base, para além dos requisitos de hardware e software, quatro pontos importantes: Aquisição, Normalização, Anotação e Análise quantitativa.

3.4.1 Hardware e Software

- 1 computador
- 2 monitores

- Kinect for Windows v2
- BIOPAC
- O sistema deverá correr em Windows
- O sistema deverá ser usado pelo gestor e participante

3.4.2 Aquisição

- O sistema tem que permitir registrar os dados do participante: ID e género.
- O sistema tem que estar dividido em duas partes correspondentes às deslocações do participante e respetivas fases.
- Para cada fase o sistema tem que gravar o participante com a Kinect.
- Para a fase um e dois, acting normal e acting com instruções o sistema tem que solicitar a emoção ao participante de forma aleatória.
- Para a fase três, elicitação de emoção, o experimentador tem que escolher qual o vídeo que vai reproduzir.
- O início e fim da gravação para as fase 1 e 2 têm que ser efectuado pelo participante através de um clique no rato do computador ou por defeito ao fim de 10 segundos.
- O início e fim da gravação para as fase 3 é controlado pelo experimentador.
- O sistema após o fim de cada gravação deverá permitir que sejam feitos ratings, pelo participante, sobre cada expressão que efectuou.
- Cada expressão de emoção deverá ser armazenada num diretório designado de “raw”.
- O diretório para cada sequência tem que ter o seguinte caminho:
“..Desktop\database\raw\session\fase\emotion”
- O nome de cada ficheiro deverá ser auto-explicativo, ou seja, ou seja normalizado por sessão, fase, género, id, emoção. A sessão pode ser A ou B (corresponde às duas sessões que o participante tem que realizar), a fase pode ser 1 ou 2, o género pode ser M ou F (Masculino e Feminino), a emoção pode ser: AL (Alegria), TR (Tristeza), ME (Medo), NO (Nojo), RA (Raiva), SU (Surpresa) e NE (Neutro) e por fim o número do frame.
- O sistema deverá guardar num ficheiro .csv os dados referentes ao participante bem como a sequência final de emoções que expressou.

3.4.3 Anotação

- O sistema deverá permitir que os frames armazenados sejam visualizados por participante e emoção.
- O sistema deverá permitir seleccionar qual o frame referente ao Neutro Inicial, o Pico e o Neutro Final de intensidade da expressão.

3.4.4 Normalização

- O sistema deverá ajustar os frames a uma grelha.
- O sistema deverá uniformizar a sequência de frames.
- O sistema deverá copiar todos os frames entre o inicial e o final e respetivo pico para um novo diretório igual ao *raw*, mas denominado *normalized*.
- O sistema deverá permitir seleccionar os frames normalizadas e converter para vídeo.

3.4.5 Análise Quantitativa

- O sistema deverá permitir utilizar as sequências de frames ou vídeos armazenados para obter dados quantitativos.
- O sistema deverá armazenar as action units de todas as faces para cada emoção e criar um ficheiro excel com esses dados.

3.5 Conclusões

Sumariamente, ficam assim apresentadas as ideias para a criação de uma base de dados dinâmica, os passos e meios para o conseguir. É importante respeitar os passos do protocolo e respetivos requisitos tornando-se assim possível avançar com o sistema computacional de suporte à aquisição da base de dados emocional. O capítulo seguinte contém as informações necessárias sobre este processo.

Capítulo 4

Sistema Computacional de Suporte à Aquisição de uma Base de Dados Emocional

Com base no protocolo experimental e nos principais objetivos desta experiência é necessário então um sistema computacional que permita realizar as tarefas necessárias para a criação da base de dados. A primeira tarefa é assegurar o suporte à aquisição, isto é, criar um sistema que permita obter todos os dados relevantes à experiência, como por exemplo, e mais importante, as sequências das expressões de emoções. Esta tarefa e todos os seus passos são descritos nas seções seguintes.

4.1 Visão Global

Resumidamente o processo de aquisição de dados divide-se em três fases já descritas anteriormente, acting espontâneo, acting com instruções e elicitación de emoção via vídeo. O processo de aquisição para a fase 1 e fase 2 são muito similares com exceção da existência de instruções na fase 2, que computacionalmente não altera nada. A fase 3 é de igual modo muito idêntica as outras fases só neste caso há um novo bloco para a visualização do vídeo que elicit a emoção, que apenas será visualizada pelo participante, enquanto o experimentador inicia e termina a gravação. Este controlo por parte do experimentador não é único, ele terá acesso a todos os passos que o participante realiza durante a gravação dos vídeos. Tem um papel importante nos ratings realizados pelo participante, pois tem que aprovar ou chumbar os mesmos, caso ache que a performance submetida pelo participante não é a correta. Cada participante terá que realizar duas sessões, em dias diferentes, uma para a fase 1 e 3 e outra para fase 2 e 3. Por fim são armazenados dados como as sequências de emoções e dados experimentais por cada participante.

O diagrama da Figura 4.1 mostra-nos o fluxograma do sistema de suporte à aquisição da base de dados emocional que é criada. Em seguida é descrito todos os blocos do diagrama:

- Configuração da experiência: Configuração de alguns dados iniciais que são necessários para a experiência. Tais como as emoções que são utilizadas ou os vídeos que são utilizados na fase 3. Ver secção 4.2.
- Dados iniciais: Este bloco inicial dá-se após a obtenção dos dados configuráveis e divide-se em 3 blocos referentes as três fases da experiência.
- Acting espontâneo: É a partir deste bloco que se realiza toda gravação das expressões emocionais bem como o seu armazenamento.
- Acting com instruções: Similiar ao bloco anterior, apenas com a diferença indireta no sistema das instruções.
- elicitación emoción - vídeo: Gravação das expressões emocionais com base na visualização de vídeos que elicitam a emoção no participante.
- Lista de emoções random: A lista de emoções é obtida através da configuração da experiência e armazenada num array. As emoções são então sorteadas aleatoriamente antes de cada gravação.
- Lista vídeos: Lista de vídeos é também obtida através da configuração da experiência.
- Gravação via Kinect: Bloco que diz respeito à gravação das expressões de emoções para cada participante. É realizada através do sensor Microsoft Kinect. Bloco comum para a fase 1 e 2. A diferença para a fase 3 é ação do experimentador que inicia o vídeo de elicitación e a respetiva gravação.
- Visualização vídeo: O vídeo para elicitación é reproduzido. Este é controlado pelo experimentador através do bloco anterior.
- Rating: No final de cada gravação, para cada emoção, o participante tem que avaliar a sua performance.
- Experimentador: O experimentador tem que aprovar ou chumbar a avaliação realizada pelo participante. Se o rating que o participante avaliou for maior que 80 e for aprovado pelo experimentador a emoção em questão é removida do array. Caso contrário, se for menor que 80 ou chumbada pelo experimentador, a emoção volta para o array para ser expressada novamente.
- Armazenamento RAW: As sequências de frames das emoções são armazenadas na pasta RAW.

- **Dados experimentais:** Dados experimentais para cada participante são também armazenados num ficheiro CSV como dados como ID, Sequência de emoções gravada e sessão.

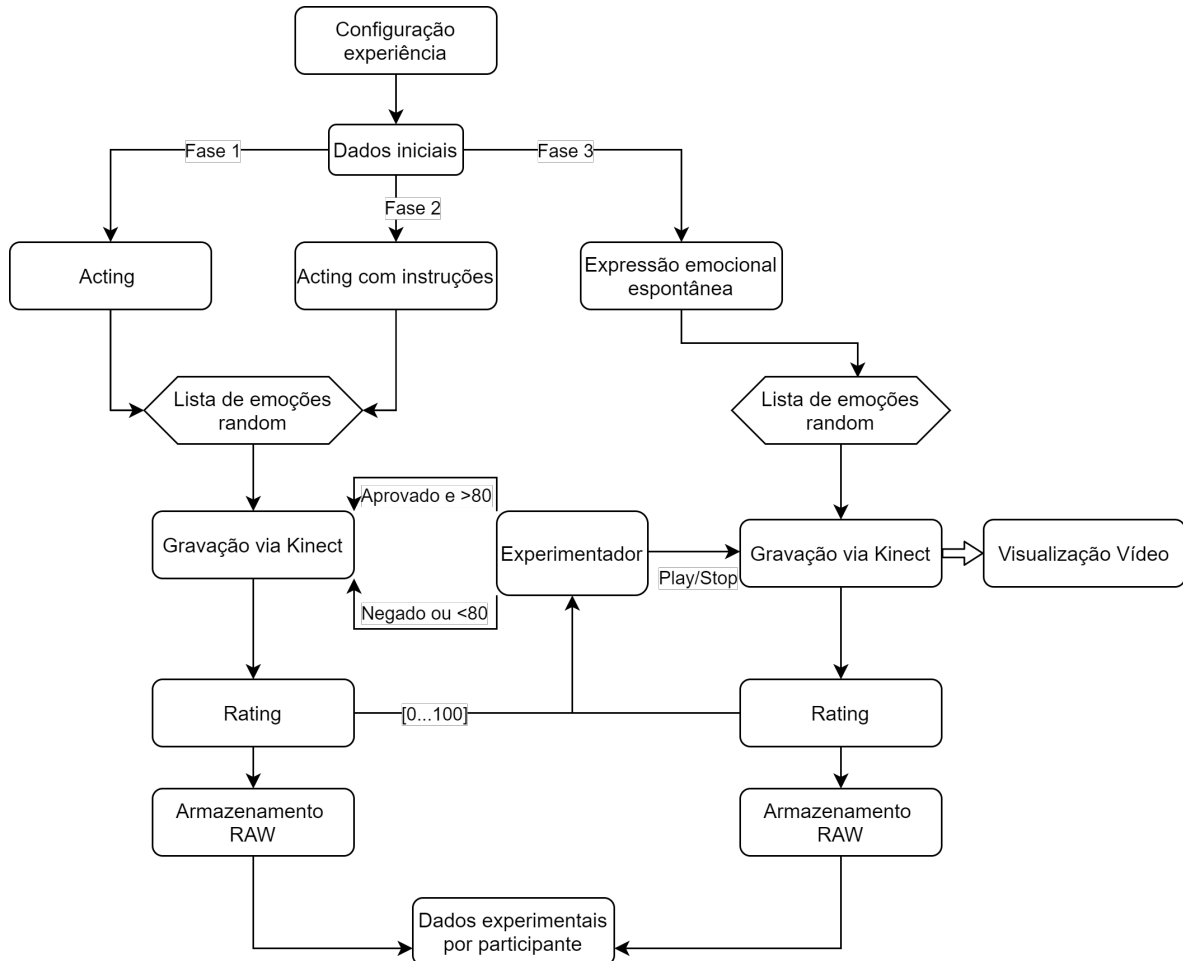


Figura 4.1: Visão global do fluxograma do sistema de suporte à aquisição da base de dados emocional.

O formulário inicial, Figura 4.2, mostra-nos as duas sessões que o participante terá que realizar, para cada uma delas a respectiva fase que irá ser realizada. A selecção destas fases é seleccionada pelo experimntador. Antes de iniciarmos a aquisição das imagens e vídeos é necessário introduzir no sistema dados sobre o participantes, nomeadamente o seu ID (que lhe é entregue durante a experiência) e seu género. Estes dados são importantes para a análise final dos dados.

4.2 Configuração da Experiência

A necessidade de existir um ficheiro de configuração torna-se bastante importante caso seja necessário actualizar ou adicionar novas funcionalidades ao sistema sem a alteração de código.

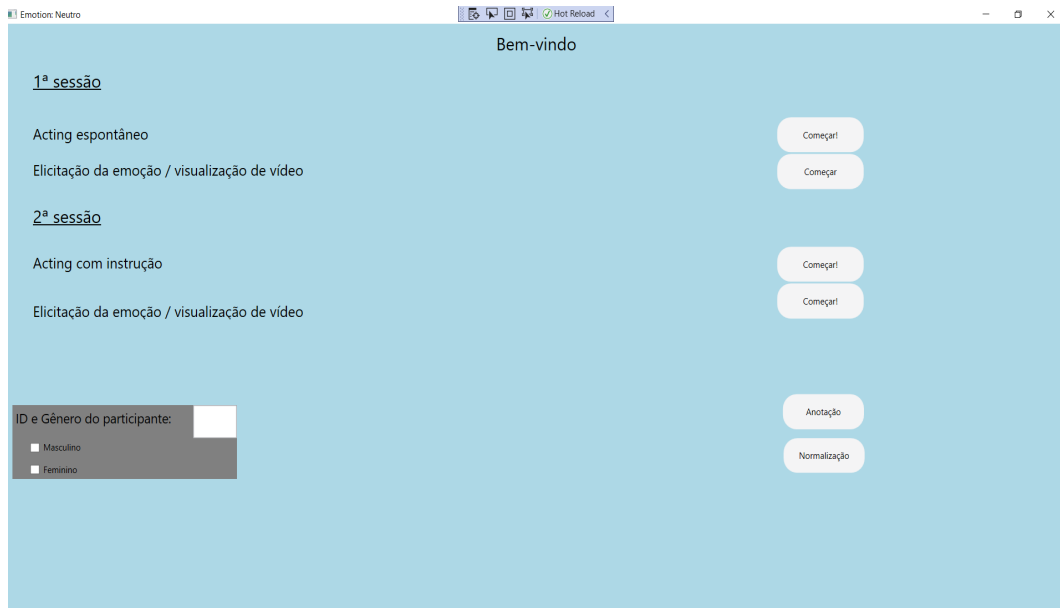


Figura 4.2: Formulário inicial com as opções para cada fase e introdução dos dados do participante.

Isto permite ao experimentador fazer pequenos ajustes a alguns dos aspetos do protocolo, por exemplo, decorrentes de alterações julgadas necessárias após os primeiros testes do sistema (e.g., tempo de gravação) sem uma intervenção do desenvolvedor.

Nesta experiência foi utilizado XML (Extensible Markup Language) pois possui uma estrutura padronizada, baseado em texto simples e auto-documentado ao que permite uma fácil leitura e percepção. O ficheiro XML contém todas as emoções utilizadas para a experiência, permitindo assim adicionar ou remover outras no futuro, o tempo de duração da gravação e os caminhos para os vídeos de elicitação de emoções. O experimentador sempre que inicia uma fase é criado um array com todas as emoções do XML e é gerado uma emoção aleatória do mesmo para a gravação da emoção. Isto permite que a actualização das emoções via rating não interfira com o XML.

A lista seguinte mostra um exemplo para a configuração XML das emoções cada uma delas é identificada pelo nome e um id (abreviação) próprio.

Lista 4.1: Exemplo código XML para as emoções

```
<emotions>
  <emotion name="Neutro" id="NE"/>
  <emotion name="Alegria" id="AL" />
  <emotion name="Medo" id="ME" />
  <emotion name="Surpresa" id="SU" />
  <emotion name="Nojo" id="NO" />
  <emotion name="Raiva" id="RA" />
</emotions>
```

```
<emotion name="Tristeza" id="TR" />
</emotions>
```

4.3 Método de Aquisição de Dados

A gravação dos dados é realizada através do sensor Kinect, esta é uma parte central do sistema desenvolvido, as sequências armazenadas são um ponto fulcral na aquisição e posteriormente na normalização e obtenção das medidas quantitativas. Nas secções seguintes podemos ver como é feita a integração da Kinect com o sistema, que ferramentas, linguagens e softwares foram utilizados, bem como é desenvolvida a gravação das expressões de emoções.

4.3.1 Integração com Kinect

Para o desenvolvimento do sistema computacional foi utilizada do sensor Microsoft Kinect v2. Para a sua integração foi necessário instalar a ferramenta de desenvolvimento para Windows. Neste caso foi usado o Kinect for Windows SDK 2.0, que permite criar aplicações em ambiente Windows, usando o sensor.

O Kinect for Windows SDK 2.0 inclui as drivers para o uso dos sensores Kinect v2 em computadores com Windows 8 (x64), Windows 8.1 (x64), e Windows Embedded Standard 8 (x64), interfaces para dispositivos e APIs e exemplos de código. Como requisito de software temos o Microsoft Visual Studio 2019.

Após instalação o sensor Kinect v2 está pronto para ser utilizado pelo sistema computacional, desenvolvido utilizando a linguagem de programação C# e WPF pertencente à plataforma .NET da Microsoft. O namespace utilizado é o Microsoft.Kinect.

Para a sua integração no sistema, resumidamente, é necessário inicializar o sensor e os seus componentes, criar um codificador no formato PNG e um frame do bitmap obtido pelo sensor adicionando-o a esse mesmo codificador. Por fim uma função lida com os dados que chegam do sensor Kinect para que seja possível a sua visualização.

4.3.2 Gravação via Kinect

Assim que o experimentador selecciona uma fase correspondente a uma das condições, ver Figura 4.2, um formulário novo é aberto e, a emoção que é aleatoriamente gerada, aparecerá no topo do formulário. A partir deste momento o participante tem as instruções necessárias para proceder a gravação dos frames. Após a posição correta por parte do participante, este tem que clicar no botão esquerdo do rato para iniciar a gravação e novamente para terminar, ver Figura 4.3.

Este processo é exactamente o mesmo para as fases 1 e 2, acting espontâneo e acting com instruções, a única diferença é que o participante na primeira fase apenas tem que expressar

as emoções de forma espontânea e na segunda fase, que apenas se realiza na segunda sessão, terá instruções sobre que músculos e processos tem que fazer para expressar a emoção.

Na fase 3, em que o participante tem que visualizar um vídeo para elicitare a emoção nele próprio a gravação é realizada de forma diferente. Haverá duas telas, uma associada ao experimentador, ver secção 4.4.2, e outra ao participante. Deste modo, é necessário um formulário para o participante, apenas contendo um player de vídeo, onde o participante não necessita de efectuar qualquer acção para além de visualizar o mesmo. Este vídeo é manipulado pelo experimentador que o inicia e o termina no segundo formulário assim que o participante está pronto. No segundo formulário é possível também visualizar a gravação via Kinect, desta vez não é necessário clicar no rato para iniciar e terminar a gravação. Esse processo é realizado sempre que se inicia ou termina o vídeo.

Sempre que termina a gravação, todos os frames capturados são armazenados num directório específico, com dados RAW. Esse directório é criado no Ambiente de Trabalho do sistema e os sub directórios seguem algumas regras de acordo com a secção 4.6.

Após o final da gravação de uma emoção, aparecerá um formulário onde o participante terá que avaliar a sua performance da sua expressão, ver secção 4.4.1, e assim que a avaliação é realizada e submetida pelo participante e aprovada ou não pelo experimentador, o sistema voltará para o formulário anterior da gravação via Kinect, onde uma nova emoção aparecerá. O processo repete-se até que não hajam mais emoções no array.



Figura 4.3: Formulário correspondente à gravação via Kinect, com a emoção que o participante tem que expressar.

4.4 Aquisição de Dados: Expressão Emocional através de *Acting*

Correspondem às fases 1 e 2 e como foi descrito na secção anterior a gravação é feita via Kinect, ver Figura 4.3, o participante tem expressar a emoção que lhe aparece no ecrã por acting, na primeira sessão sem instruções e na segunda sessão com as instruções que lhes foram entregues na primeira. Terá que repetir este processo para todas as emoções que se encontram configuradas no ficheiro XML e no fim avaliar a sua performance. Essa tarefa é descrita na secção seguinte 4.4.1.

4.4.1 Ratings

Assim que o participante termina de gravar uma emoção, um novo formulário aparece, ver Figura 4.4, onde tem que avaliar a sua performance. Para isso tem que deslizar um slider baseado numa escala visual analógica onde o participante tem que submeter a sua escala de avaliação de 0 a 100, se esta for maior que 80 e posteriormente aprovada pelo experimentador 4.4.2, a emoção é removida do array de emoções que ele tem que expressar e volta novamente à gravação de uma nova emoção Figura 4.2, caso essa avaliação for menor que 80 ou chumbada pelo experimentador, não será removida do array e será novamente aleatoriamente pedida ao participante que a expresse no formulário de gravação. Este processo é realizado no final de todas as gravações de emoções e sempre que este processo termina são armazenados todos os frames capturados pela gravação via Kinect e os dados experimentais do participante, ver secção 4.6.

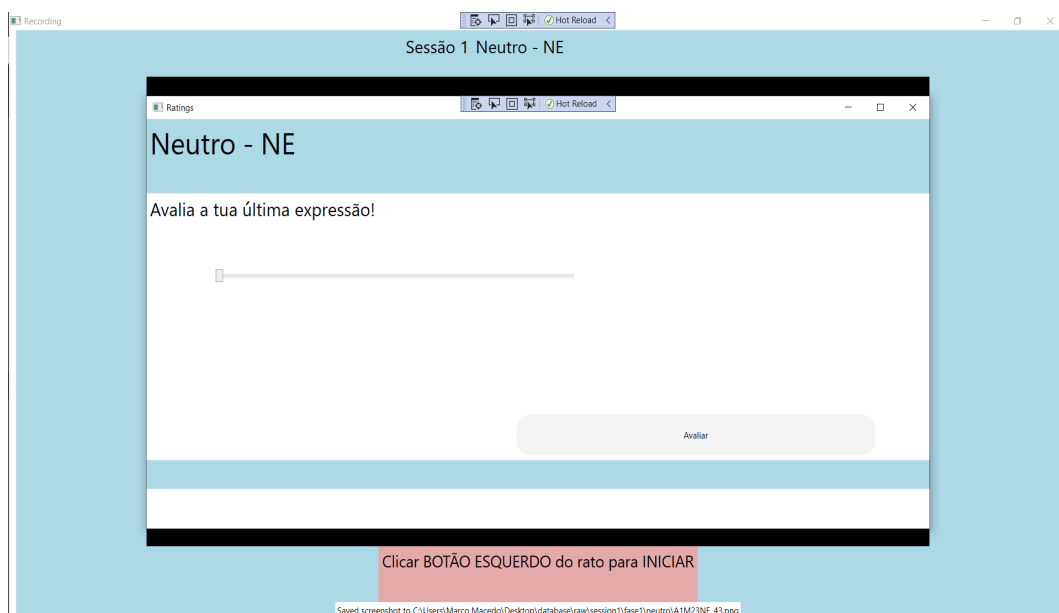


Figura 4.4: Avaliação via rating com escala visual analógica.

4.4.2 Papel do Experimentador

Todo o processo desde que o participante inicia a utilização do sistema até que termina é monitorizado pelo experimentador com uso de dois ecrãs. Partes do sistema utilizam dois formulários distintos em cada um dos ecrãs. Um formulário para o participante e outro para o experimentador. Para fase 1 e 2, como já foi descrito na secção anterior, ver secção 4.4.1, os ratings são um caso onde isso acontece, quando o participante submete o seu rating, num segundo ecrã o experimentador pode aprovar ou chumbar a avaliação submetida se achar que a performance não foi suficiente ou explicitamente bem avaliada pelo participante, ver 4.5. O Participante nunca tem acesso à avaliação do experimentador. Assim que o que experimentador aprova, ou não, e segundo os parâmetros descritos na secção anterior, ver secção 4.4.1 uma nova emoção é exibida para gravação.

O experimentador tem assim também, um papel importante, em todo o processo, pois é capaz de seguir todos os passos que o participante toma, para o caso necessitar rectificar ou ajudar ou em algum dos passos.

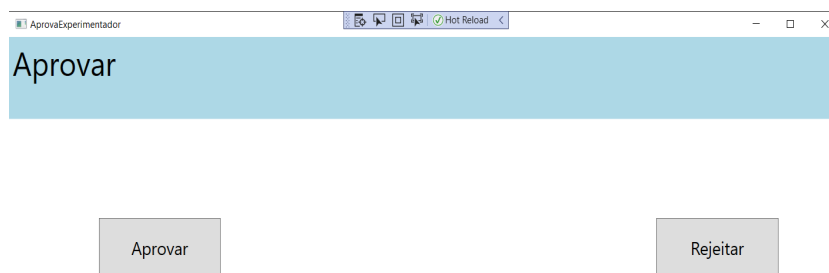


Figura 4.5: O experimentador aprova ou rejeita a avaliação acabada de submeter pelo participante.

4.5 Aquisição de Dados: Expressão Emocional Espontânea

Corresponde à fase 3 do processo. Este processo assemelha-se aos processos da fase 1 e 2. O participante terá que visualizar um vídeo que elicitará emoção nele próprio de forma espontânea, enquanto a Kinect grava da sua expressão facial. Uma das diferenças, nesta fase do protocolo, será o papel mais activo do experimentador que, ao contrário das outras fases, terá que controlar o início e fim da gravação, tal como descrito na secção seguinte. Adicionalmente, esta fase em que se pretende que a expressão facial de emoção seja espontânea, durante o visionamento de vídeos especificamente escolhidos para o efeito, o número de emoções considerado é menor, estando limitado a alegria e medo.

4.5.1 Papel do Experimentador

Nesta fase, há um formulário para o participante, figura 4.7, que apenas contém o player com o vídeo de elicitación e um segundo, para o experimentador, figura 4.6, que contém a gravação via Kinect e os botões para iniciar e terminar o vídeo do primeiro formulário, assim o participante apenas tem que esperar que o vídeo se inicie sem realizar nenhuma acção. O início e fim gravação via Kinect funciona com os mesmos botões que controlam o vídeo. Desta forma o experimentador tem um papel mais activo nesta fase da experiência, sendo que é o responsável pelo o processo todo.

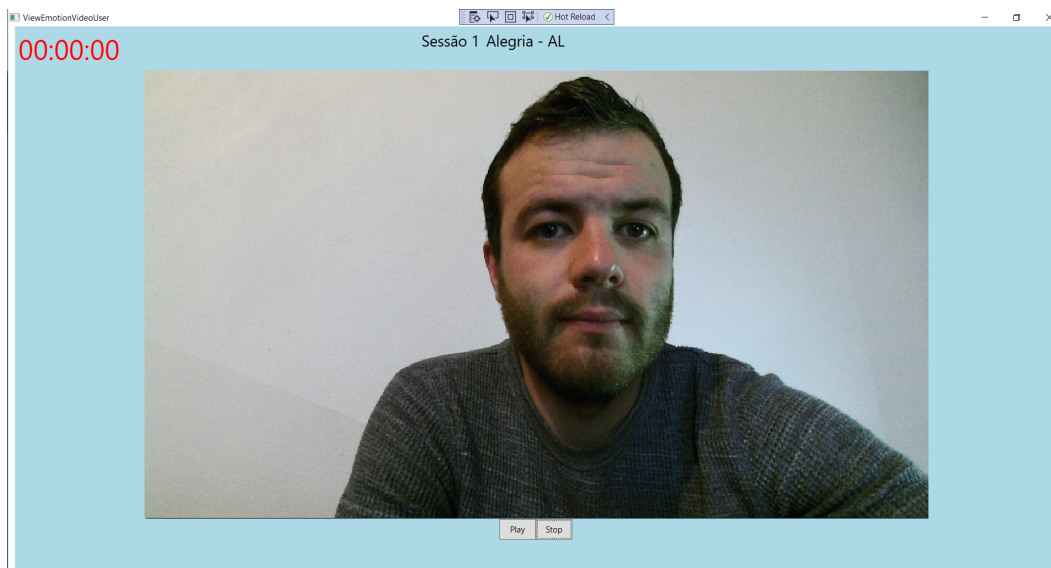


Figura 4.6: Janela com o formulário do experimentador. Botões Play e Stop para iniciar vídeo e gravação com Kinect.

Terá de igual forma um papel importante na aprovação ou não dos ratings submetidos pelos participantes no final de cada gravação, ver secção 4.4.1.

4.5.2 Expressão Espontânea de Emoção

De forma a elicitare uma expressão de emoção que seja espontânea, o protocolo experimental descrito no capítulo 3 prevê a apresentação de vídeos ao participante. Para o efeito pretendido de criação da base de dados tratada neste trabalho são considerados vídeos para elicitare medo e alegria, assim como vídeos neutros. Estes foram previamente validados, em laboratório, de forma a aferir a sua capacidade de elicitare as emoções pretendidas. Os vídeos podem ser partes de filmes, por exemplo de comédia para elicitare alegria, ou de de terror para o medo. O vídeo neutro poderá ser algo que transmita tranquilidade ao participante, como ouvir sons de ondas ou de pássaros a cantar.

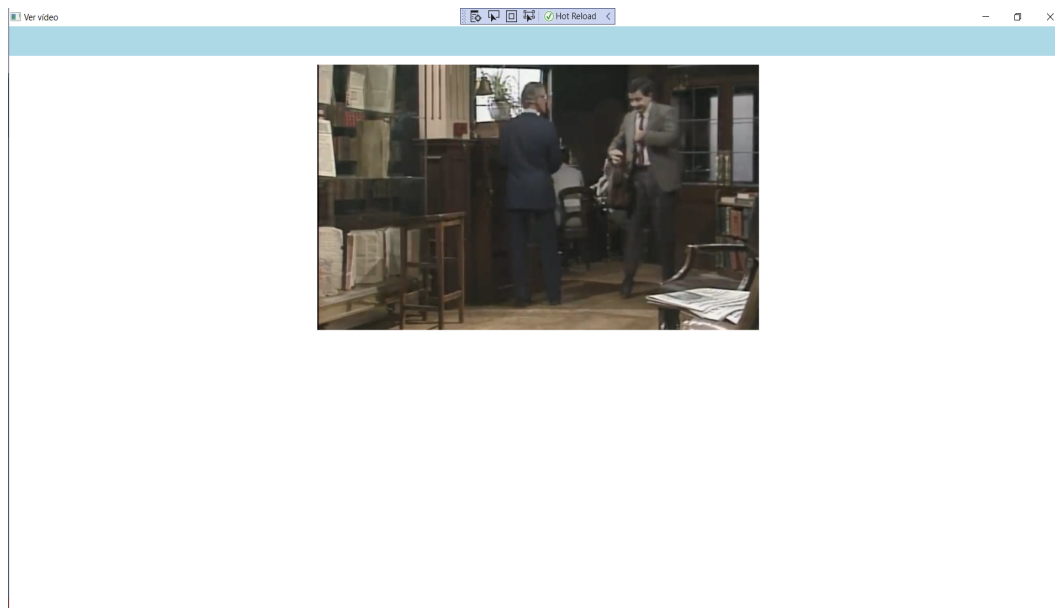


Figura 4.7: Janela com o formulário do participante. Vídeo é reproduzido para elicitare a emoção.

O ficheiro XML de configuração da experiência contém a lista dos vídeos que correspondem as emoções de medo, alegria e neutro que poderá ser ajustada à medida das necessidades, por exemplo, para incluir mais vídeos. Eles serão seleccionados pelo experimentador consoante a sessão que o participante está a realizar. Se na primeira sessão o vídeo visualizado for o de alegria, na segunda sessão o vídeo que visualizará será o de medo, sendo que antes de qualquer uma deles o participante tem que visualizar o neutro para o colocar num estado emocional completamente isento. Este procedimento é importante para que, caso medidas psicofísicas sejam adquiridas, durante a aquisição, e.g., ECG, haver um período de estabilização das mesmas (baseline) que permita, depois, aferir alterações relevantes.

A duração dos vídeos poderá variar de uns para os outros mas, posteriormente, serão alvo de anotação e pós processamento para seleccionar apenas o trecho de gravação onde se observa expressão facial da emoção pretendida.

4.6 Armazenamento dos Dados Experimentais

O armazenamento de todos os dados é feito num directório gerado automaticamente no Ambiente de Trabalho de qualquer computador que use o sistema computacional desenvolvido. É necessário e importante ter um directório claro, para que seja o mais perceptível possível aos investigadores que usem a base de dados. Os nomes dos ficheiros também seguem um determinado padrão, pelo mesmos motivos, ver secção 4.6.1.

Os dados que são armazenados durante a aquisição são as sequências de frames capturadas

durante a gravação via Kinect e os dados experimentais, gravados num ficheiro CSV, onde é possível visualizar as informações da experiência e do participante que a realiza, isto é, o seu ID, a sessão, fase e sequência de emoções gravadas por ele.

4.6.1 Estrutura

O nome de cada ficheiro é composto a partir da sessão (cada participante vem duas vezes ao laboratório: A ou B, fase: 1 ou 2, género, id, emoção: AL, TR, ME, NO, RA, SU e NE e número do frame. Por exemplo: A1M1000ME_100. O formato dos frames é em PNG e os vídeos AVI. A Figura 4.8 ilustra o directório completo correspondente à base de dados.

Todos os ficheiros acabados de adquirir inicialmente são armazenados na pasta RAW, isto porque estes dados ainda não estão normalizados. Após normalização serão movidos para a pasta NORMALIZED (discutido no capítulo seguinte).

Já os dados experimentais, de cada participante, serão armazenados num ficheiro CSV designado de “data.csv” onde contém os dados referentes a cada sessão, como a fase, id, género e ordem de sequência de emoções gravadas pelo participante, a figura 4.9 mostra a estrutura do ficheiro. Como é possível ver, por exemplo para o participante com ID 1000, na fase 1, a sequência de emoções gravadas foi a MEDO -> RAIVA -> ALEGRIA -> TRISTEZA -> SURPRESA -> NOJO -> NEUTRO.

4.7 Conclusões

Neste capítulo foi apresentada uma plataforma computacional para suportar a aquisição dos dados de expressão de emoção seguindo o protocolo descrito no Capítulo 3. Partindo de um ficheiro de configuração, o sistema desenvolvido possibilita os diferentes passos a seguir assim como o armazenamento dos dados em disco para posterior processamento. O passo seguinte permite tratar estes dados em bruto, através de anotação e normalização, para os tornar parte integrante da base de dados e consequentemente alcançar o resultado final de uma base de dados dinâmica, com dados quantitativos que a suportam. O capítulo seguinte mostra-nos os métodos necessários para o pós-processamento dos dados armazenados durante a aquisição.

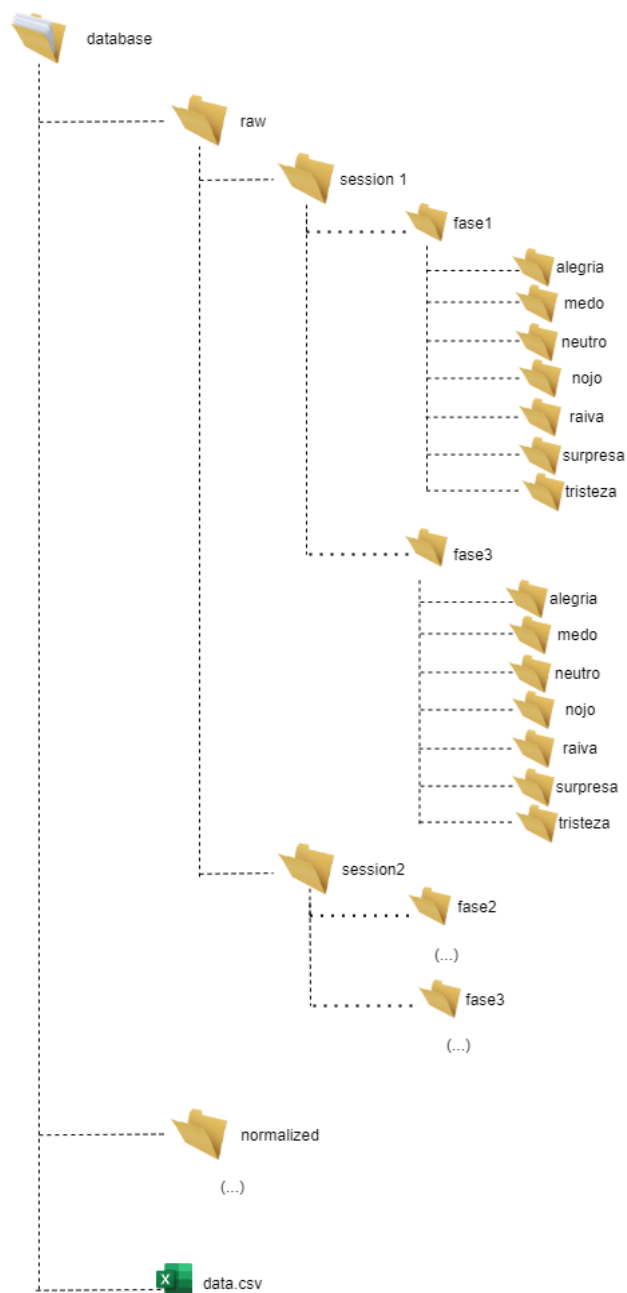


Figura 4.8: Directório da base de dados simplificada.

	A	B	C	D
1	Fase	Person Id	Género	Seq. Emoções
2	1	13123	M	ME->RA
3	1	13123	M	ME->RA->AL
4	1	13123	M	ME->RA->AL->TR
5	1	13123	M	ME->RA->AL->TR->SU
6	1	13123	M	ME->RA->AL->TR->SU->NO
7	1	13123	M	ME->RA->AL->TR->SU->NO->NE

Figura 4.9: Ficheiro CSV com dados experimentais por participante.

Capítulo 5

Anotação e Pós Processamentos de Dados

Para garantir uma base dados uniforme e dinamicamente organizada é necessário, após a aquisição de dados da experiência, realizar vários métodos de processamento nas imagens das faces capturadas dos participantes durante a expressão das emoções, com o intuito de que todas elas tenham proporções e formatação iguais. Este processo é denominado por normalização.

Para iniciarmos o processo de normalização, é previamente realizado um outro processo, chamado de anotação, que consiste em seleccionar os frames, para cada expressão de emoção, referente a sua intensidade, isto é onde se inicia, Neutro Inicial, o Pico e onde termina, Neutro Final. É assim garantido que as sequências são armazenadas por essa ordem de evolução da intensidade.

É após finalizar este passo que se inicia a normalização das sequências de imagens, as faces são detectadas e ajustadas a uma grelha padrão de tamanho igual para todas, garantindo que não haja condicionantes que não permitam uma análise precisa dos dados quantitativos.

5.1 Visão Global

A ordem dos métodos computacionais necessários para pós-processamento dos dados é descrita no Diagrama da Figura 5.1. De um forma sucinta, os dados adquiridos e armazenados no directório RAW são alvos do processo de anotação, onde são seleccionadas as intensidades (Neutro Inicial, Pico, Neutro Final) de em todas as sequências. Após este passo é iniciado o tratamento de dados, isto é, a normalização, para todas as sequências de imagens anotadas anteriormente. Estas sequências são então, armazenadas no directório NORMALIZED, onde são posteriormente utilizadas para obter as medidas quantitativas, action units e marcos faciais, através da ferramenta OpenFace. Estas medidas são enviadas para um ficheiro CSV que será armazenado no directório NORMALIZED juntamente com a sua sequência da expressão

de emoção. É através do mesmo directório que são obtidos e criados os vídeos emocionais, sendo eles sem ou com os marcos faciais visíveis.

Por fim são obtidas as medidas quantitativas para posterior análise. Para além disso há também a criação de vídeo através das sequências de frames armazenados e adição dos marcos faciais em movimento no vídeo.

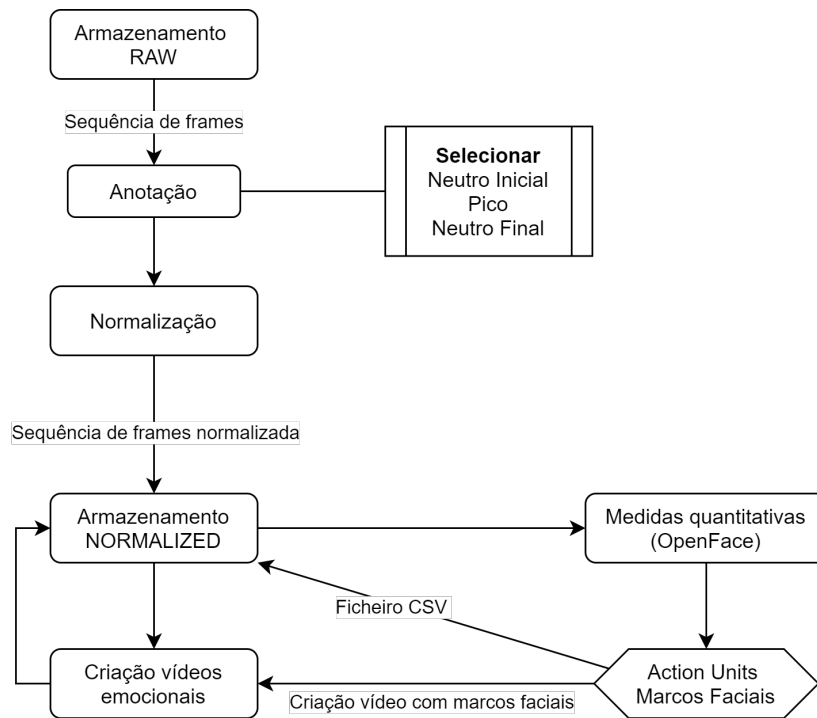


Figura 5.1: Fluxograma dos métodos de pós-processamento.

5.2 Anotação

Depois de os dados para as várias sequências de expressão facial de emoção serem adquiridos, torna-se necessário fazer uma anotação destes para identificar, em cada expressão de emoção gravada, apenas a sequência de imagens que deve integrar a base de dados. Uma vez que interessa disponibilizar, na base de dados, a evolução desde a expressão facial neutra até uma expressão de máxima intensidade de uma emoção e de volta à expressão neutra, é importante identificar, em cada sequência, cada um destes momentos. Uma identificação destes instantes, depois da aquisição, permite que os participantes não sejam forçados a cumprir rotinas de gravação tão apertadas, o que origina inícios, picos e fins da expressão facial variáveis, mas potencialmente mais naturais

A selecção de frames, ver Figura 5.2, é realizada a partir directório RAW, onde se selecciona os frames que desejamos consoante cada participante e emoção. Após essa selecção é possível visualizar a sequência de frames seleccionada através de um *slider*, que nos permite ver frame

a frame e perceber qual a intensidade da emoção para cada um deles, de seguida podemos seleccionar qual é o Neutro inicial 5.3 e sinaliza-lo para que seja possível deslizar o *slider* e identificar aquele frame como o Neutro inicial mesmo que pretendamos mais tarde altera-lo. O processo repete-se para a selecção do Pico 5.4 e o Neutro Final 5.5.

Posto isto o passo final é finalizar o processo clicando no botão Guardar. As seqüências são gravadas e estão prontas para o seguinte passo, a sua Normalização.

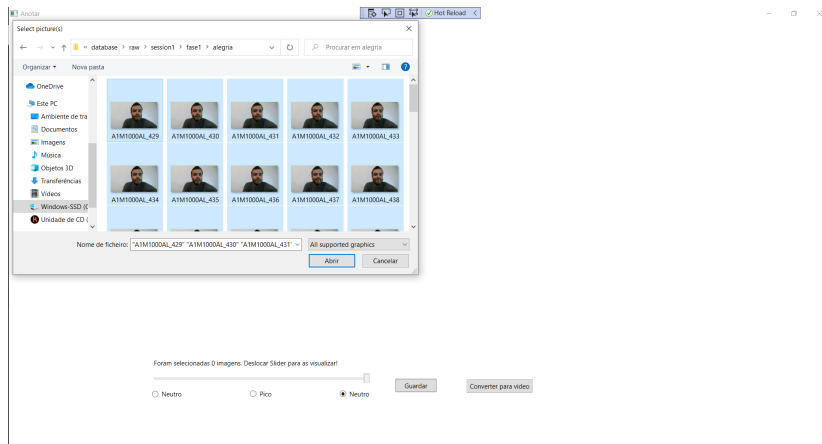


Figura 5.2: Anotação dos dados: Seleccionar as seqüências de frames capturadas por ID e emoção.

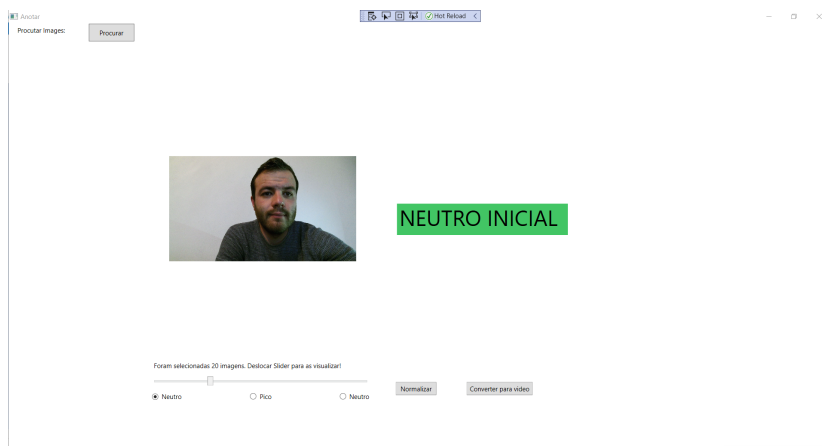


Figura 5.3: Anotação dos dados: Seleccionar o frame correspondente ao Neutro inicial.

5.3 Normalização e Resultados

Para que os dados adquiridos e anotados possam ser usados em contexto experimental, é importante garantir que algumas das suas características sejam uniformes para todas as seqüências adquiridas de um mesmo participante e entre participantes. Desta forma, pretende-

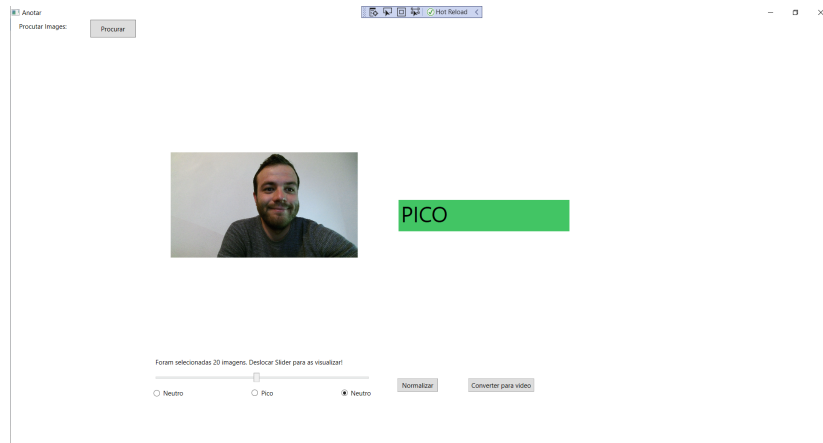


Figura 5.4: Anotação dos dados: Seleccionar o frame correspondente ao Pico.

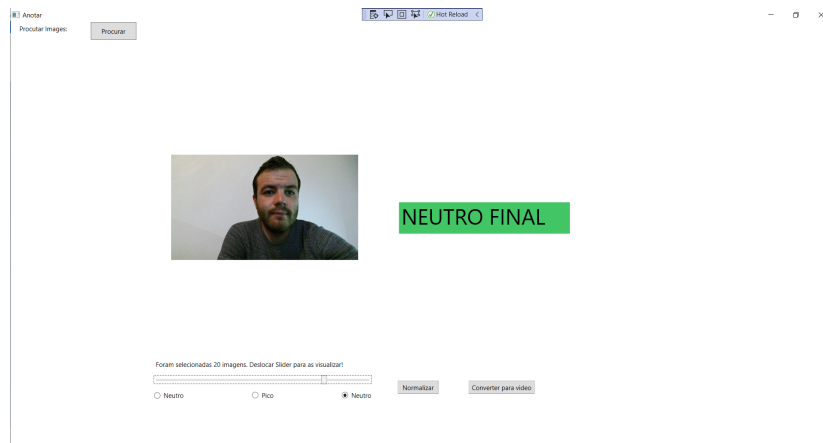


Figura 5.5: Anotação dos dados: Seleccionar o frame correspondente ao Neutro final.

se minimizar as diferenças entre sequências que, não estando directamente relacionadas com a expressão facial, podem influenciar como o conjunto é percebido por quem vê a sequência. Por exemplo, se em algumas sequências a cara aparecer com dimensões diferentes e posicionamentos diferentes, no enquadramento, isso pode influenciar a quantidade de detalhe observável e, conseqüentemente, o seu impacto.

A normalização é feita com o auxílio de um wrapper .NET multiplataforma para a biblioteca de processamento de imagens OpenCV. Permite que funções de OpenCV sejam chamadas em linguagens .NET.

O processo de normalização inclui as seguintes fases: Selecionar iterativamente uma sequência de imagens anotadas, detectar zonas rectangulares na face das imagens, aplicar o corte da imagem, aplicar o contraste e guardar a sequência normalizada.

A primeira fase é consiste em seleccionar uma sequência de expressão de emoção anotada e adicionar-lhe uma lista, que de forma iterativa vai percorrer todas as imagens. Na seguinte

fase, com um auxílio do classificador *Haar Cascades* [34] para detecção de faces, que é numa abordagem baseada em *machine learning* onde uma função em cascata é treinada através várias imagens positivas e negativas, e com base nesse treino é então usado para detectar as faces nas imagens anotadas. Este processo permite obter zonas rectangulares em redor das faces da sequência de imagens. Os rectângulos vão ter coordenadas diferentes para cada imagem da sequência, visto que posição da face varia ao longo da expressão. Para evitar este problema, vamos utilizar a mesma dimensão do rectângulo de todas as imagens, isto é, sempre que uma iteração é realizada verificamos se as coordenadas X,Y superiores à esquerda e coordenadas X,Y inferiores à direita são menores e maiores, respectivamente, que as coordenadas do rectângulo da imagem anterior na iteração, e guardamos esses valores. Por fim, é aplicado o corte da porção de imagem consoante os valores da menor coordenada X,Y superior à esquerda e a maior coordenada X,Y inferior à direita do rectângulo da face com uma largura e altura fixa de 248 e 340, respectivamente. Para além disso há também um ajuste de contraste, é definido um parâmetro limite (*threshold*) entre -100 e 100, onde valores positivos aumentam o contraste e negativos diminuem. Baseado no valor (para este trabalho foi definido como 10) é calculado o nível de contraste com base na formula $C = ((100 + T)/100)^2$, onde o C representa o contraste e o T o *threshold*. Por fim é necessário ajustar o contraste por todos os pixels de cor da imagem, que é realizada através das seguintes formulas:

$$R = (((R1/255.0) - 0.5) * C) + 0.5 * 255.0$$

$$G = (((G1/255.0) - 0.5) * C) + 0.5 * 255.0$$

$$B = (((B1/255.0) - 0.5) * C) + 0.5 * 255.0$$

Em que o R, G, B representam a cor de contraste ajustada, Vermelho, Verde e Azul (Red, Green, Blue), R1, G1, B1 os valores originais de cor e C o nível de contraste calculado antes. A sequência normalizada é guardada imediatamente a seguir a selecção das sequências anotadas. O formulário permite ver as sequências de imagens, não normalizadas e normalizadas, ver Figura 5.6.

A figura 5.7 e a figura 5.9 mostram-nos dois exemplos entre uma sequência de 3 frames não normalizados para a emoção de alegria e nojo, respectivamente. Após serem normalizadas, obtemos o resultado das mesmas sequências na figura 5.8 e figura 5.10. Como é possível verificaram-se algumas alterações, como por exemplo a posição das faces, que agora se encontram centradas.

5.4 Descritores Quantitativos dos Dados

Um dos aspetos importantes para permitir a validação da base de dados assim como investigação futura relacionada com análises computacionais da dinâmica da expressão facial de emoção é o cálculo de descritores das características julgadas pertinentes. Neste trabalho, pretendeu-se criar a infraestrutura de suporte a esse cálculo e ilustrar a sua funcionalidade

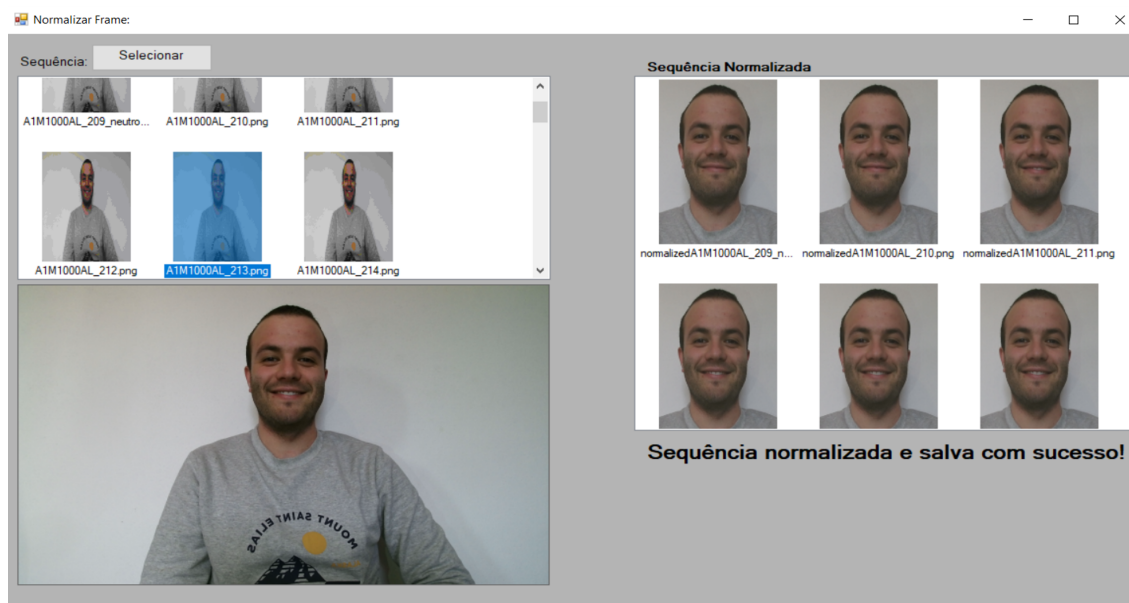


Figura 5.6: Formulário de Normalização: À esquerda uma lista que contém uma sequência de frames anotadas e não normalizados e à direita a sequência de frames já normalizada.



Figura 5.7: Exemplo de sequência de 4 frames não normalizados para emoção de Alegria.

com duas medidas comumente consideradas na literatura (e.g. [25]): os marcos faciais e as action units.

Será utilizada a ferramenta OpenFace [29] descrita na seção 2.4.3.1. Através da Windows GUI que é fornecido (OpenFaceOffline.exe) é possível adicionar as sequências ou vídeos das emoções que foram gravados e obter dados como action units ou marcos faciais de todas as faces 5.11.

Nas seções seguintes é possível visualizar estas medidas bem como a análises das mesmas.

5.4.1 Action Units

Através da ferramenta conseguimos obter valores de 18 action units, como ilustra a figura 5.12, nomeadamente as action units 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45, de acordo com o FACS (Facial Action Coding System) [19]. Todos essas medi-



Figura 5.8: Exemplo de sequência de 4 frames após normalização para emoção de Alegria.

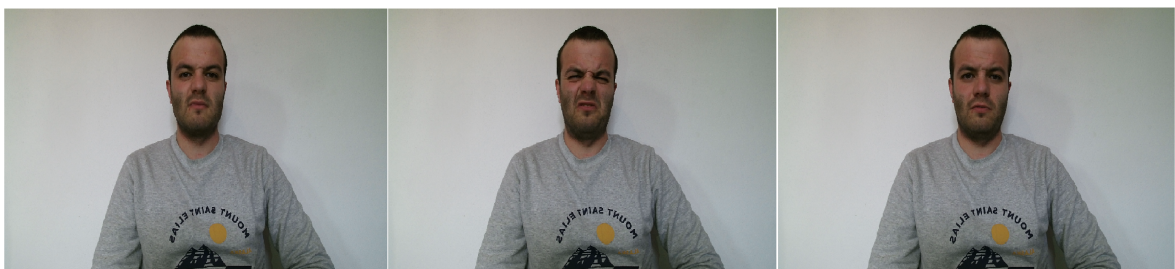


Figura 5.9: Exemplo de sequência de 4 frames não normalizados para emoção de Nojo.

das são armazenados num ficheiro CSV após adicionarmos a sequência de emoções desejada, como indica a figura 5.11. Os valores da intensidade variam entre 0 e 5. De seguida podemos visualizar e comparar graficamente os valores obtidos para sequências de frames da expressão de emoção de Alegria e Medo. É possível verificar na ferramenta que alterações existem em todas as AU ativadas durante a expressão de uma das emoções.

A sequência de expressão de emoção de Alegria contém 55 frames. Há alterações nas action units AU01, AU02, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26 e AU45 (Maior parte delas estão associadas aos lábios e a covinha da cara).

Foram seleccionadas 5 dessas action units onde a intensidade é maior, nomeadamente a AU06 (Levantamento da bochecha), AU07 (tensor da pálpebra), AU12 (cantos dos lábio para cima), AU14 (cavinha), AU15 (Depressão do canto da boca), para a criação, a título ilustrativo, do gráfico 5.13.

A sequência de expressão de emoção de Medo contém 49 frames, sendo que as action units que sofreram alterações durante a expressão da emoção foram a AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26 e AU45. Para a criação do exemplo gráfico 5.14 foram seleccionadas apenas 5 dessas action units, nomeadamente a AU04 (sobrancelha baixa), AU07 (tensor da pálpebra), AU10 (Levantamento lábio superior), AU17 (Levantamento do queixo) e AU25 (Separação dos lábios), pois foi onde se notou uma maior intensidade de valores.



Figura 5.10: Exemplo de sequência de 4 frames após normalização para emoção de Nojo.

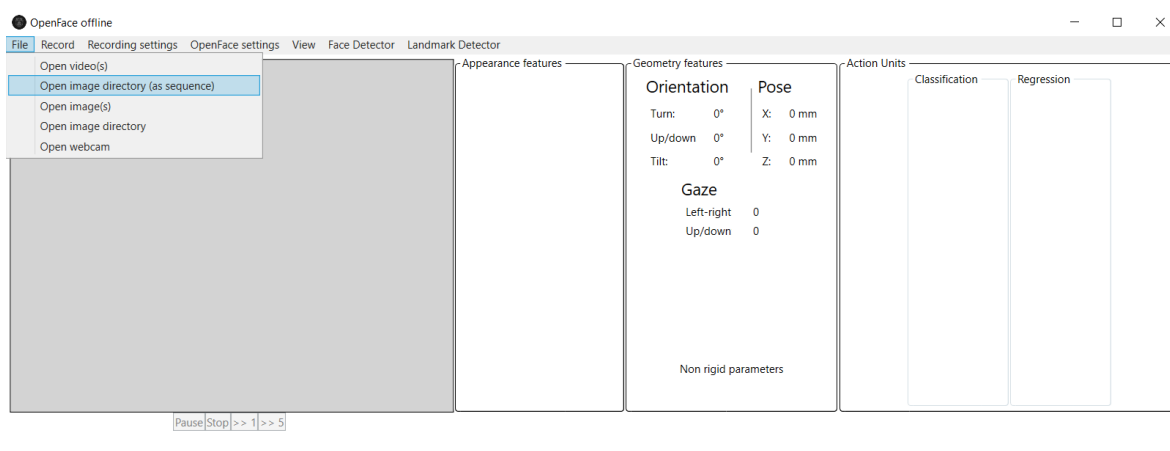


Figura 5.11: Ferramenta OpenFace: Windows GUI OpenFaceOffline e as suas opções. Permite abrir vídeos, diretórios como sequências ou apenas imagens. Após essa selecção é possível ver os marcos faciais e action units.

Naturalmente quando se expressa alegria o que se destaca é o sorriso da pessoa. Segundo os dados obtidos é possível ver que é isso que acontece, o o levantamento dos lábios e das bochecha da face (AU12 e AU06) é notório, características normais de quando alguém sorri ou se sente alegre. Em comparação com a emoção de medo os action units que se ativam maioritariamente são o abaixamento das sobrancelhas (AU04) e a separação dos lábios, mais um vez marcos normais de quando alguém se sente com medo.

5.4.2 Marcos Faciais

Para além das action units, outras das medidas quantitativas que armazenamos, com o auxílio do OpenFace são os as sequências dos frames com os marcos faciais visíveis. Isto permite visualizar diretamente na imagem capturada que marcos faciais estão ativos durante a expressão da emoção. Permitindo assim, ver secção 5.5.1, criar vídeos das sequências armazenadas com a visualização em tempo real dos marcos faciais na face do participante.

A seguinte figura 5.15 são exemplos de uma pequena sequência frames com os marcos faciais

visíveis na face.

5.5 Base de Dados Resultante

Finalmente temos os métodos computacionais para a criação da base de dados emocional dinâmica praticamente todos assegurados faltando apenas processar os dados para criar vídeos das várias sequências de frames, descrito na sub-secção seguinte. Juntamente com os vídeos a base de dados possui todas as sequências de expressões de emoção, as medidas quantitativas e os dados da experiência por participante. Está dividida em dois diretórios principais, o RAW e NORMALIZED, na primeira temos os dados por normalizar e na segunda todos os dados finais que constituem a base de dados normalizada, é aqui que é possível visualizar os vídeos, sequências e as medidas quantitativas.

Alguns detalhes da base de dados são indicados de seguida:

- Resolução dos frames normalizados: 250x350px
- Tamanho médio dos frames: 165KB
- Formato dos frames: PNG
- Formato vídeos: AVI
- Tamanho médio dos vídeos normalizados: 100KB
- Resolução dos vídeos: 250x350px
- Formato dos ficheiros das medidas quantitativas e dados da experiência: CSV
- Exemplo nome ficheiro para frames: A1M1000ME_100.JPG
- Exemplo nome ficheiro para vídeo: A1M1000ME.AVI

5.5.1 Criação de Vídeos Emocionais

A criação do vídeo final com a expressão da emoção é feito através da conversão dos frames normalizados para um ficheiro de vídeo. Para a sua realização é necessário seleccionar os frames, ver Figura 5.16, e de seguida carregar no botão Converter. É possível também visualizar os frames no formulário. Esta conversão é efectuada utilizando a ferramenta AForge.NET Framework [32] utilizando o namespace AForge.Video.FFMPEG, ver documentação [33].

Serão criados vídeos com as sequências de frames dos dados normalizados, bem como vídeos com os marcos faciais sinalizados na face do participante.

5.6 Conclusões

A obtenção das medidas quantitativas finais referentes a cada emoção é um ponto fulcral para permitir validação posterior dos dados, pois permite que investigadores ou pessoas da área da psicologia estudem a base de dados da melhor forma possível. O processo de pós-processamento é deste modo, também um ponto de grande importância na construção da base dados, a anotação dos picos nas sequências e armazenamento as medidas quantitativas através dos frames normalizados permite criar uma grande vantagem perante outras bases de dados já existentes, sendo uma grande ajuda para futuros estudos na área. A visualização das sequências de emoções via vídeo juntamente com os marcos faciais, adicionam também uma mais valia a base de dados, permitindo visualizar o desenvolver da emoção desde o seu início até ao seu final.

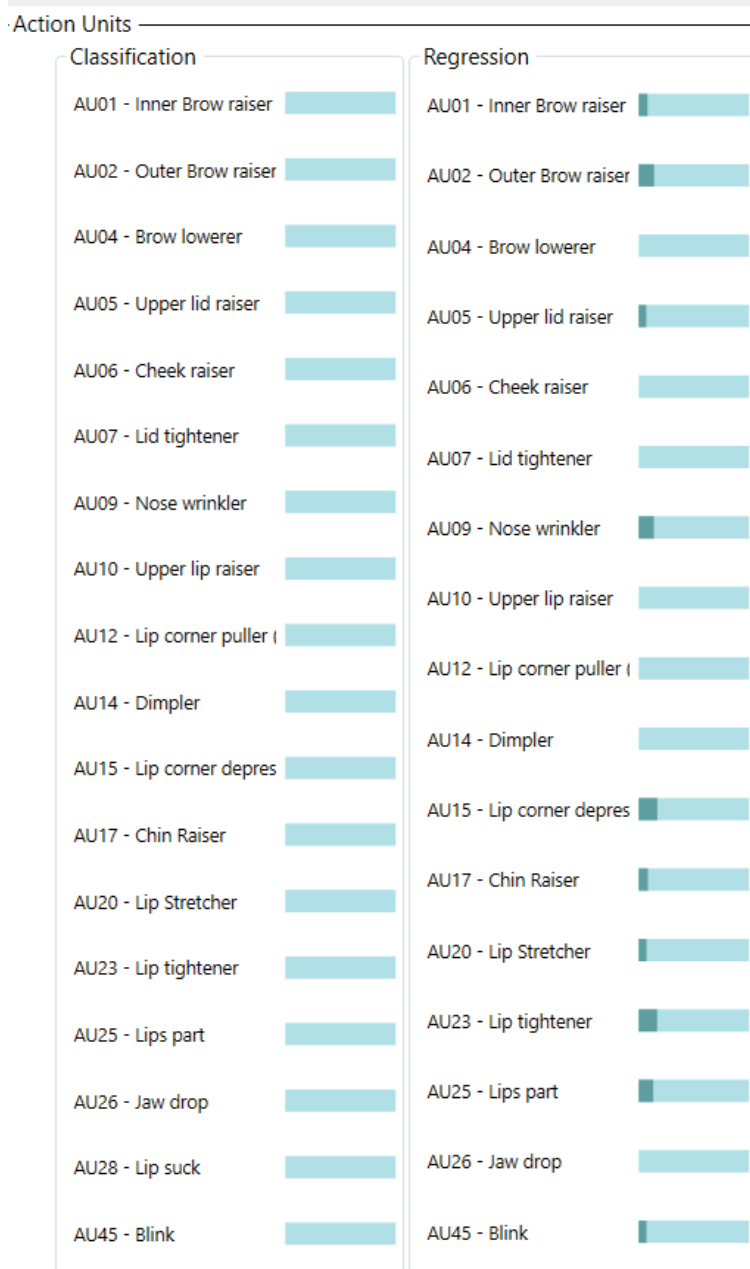


Figura 5.12: Action Units disponíveis com o OpenFace.

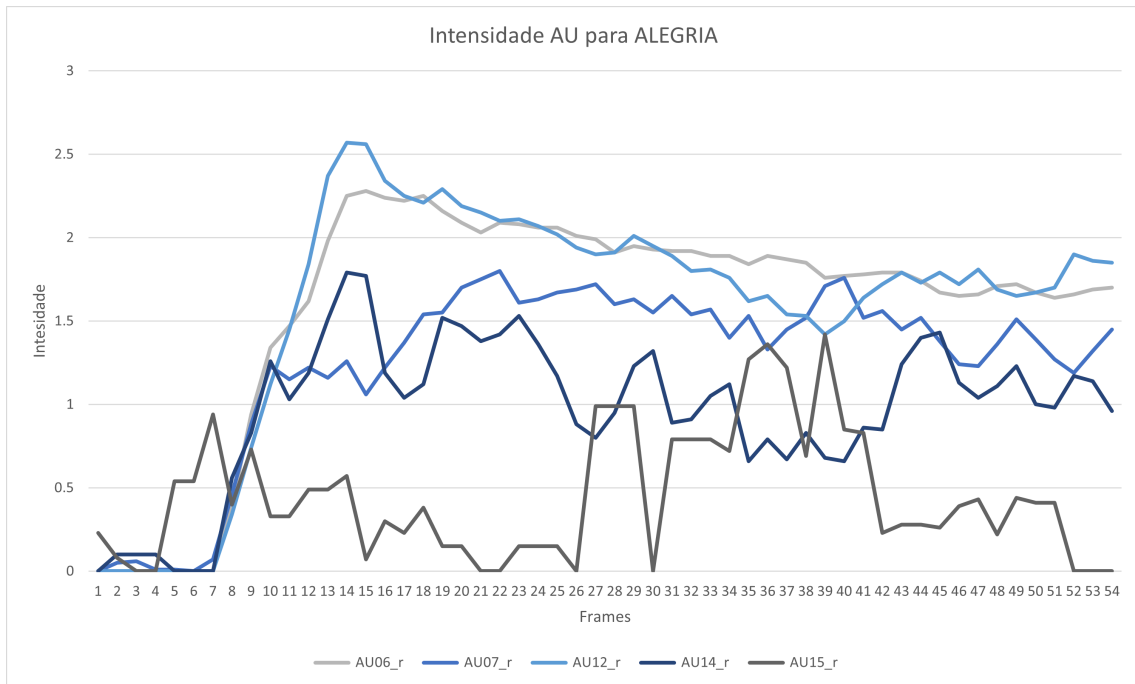


Figura 5.13: Gráfico de valores de algumas das action units para a emoção Alegria.

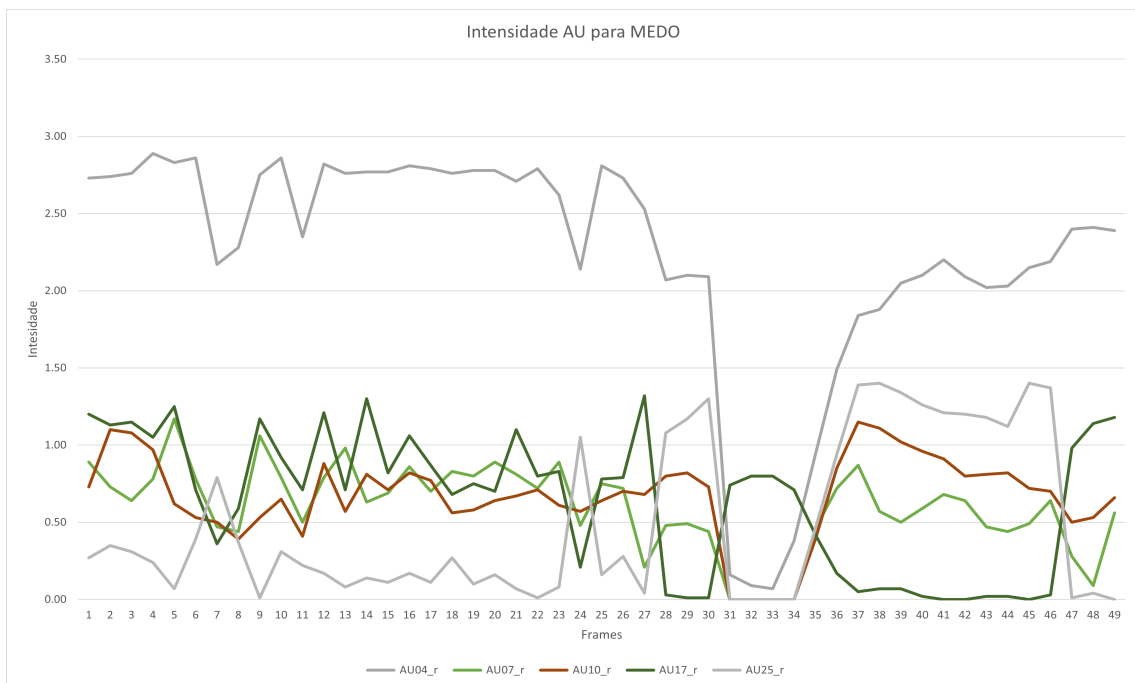


Figura 5.14: Gráfico de valores de algumas das action units para a emoção Medo.

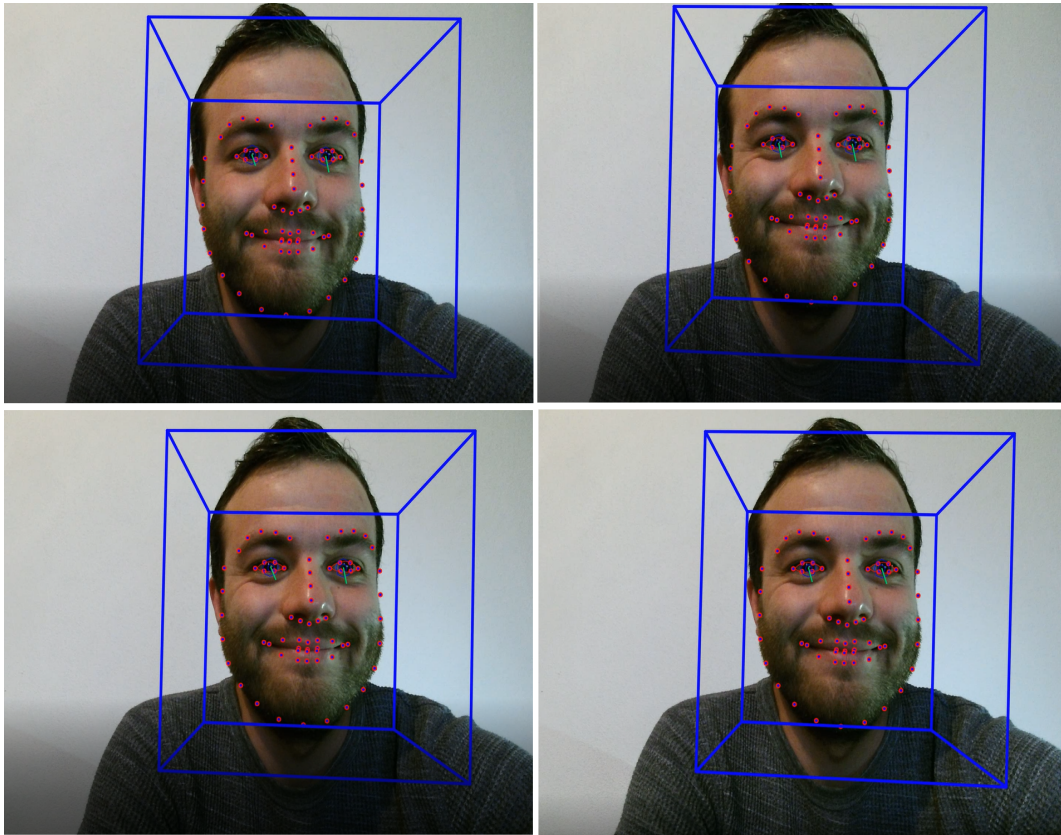


Figura 5.15: Sequência de frames com marcos faciais para a emoção de Alegria.

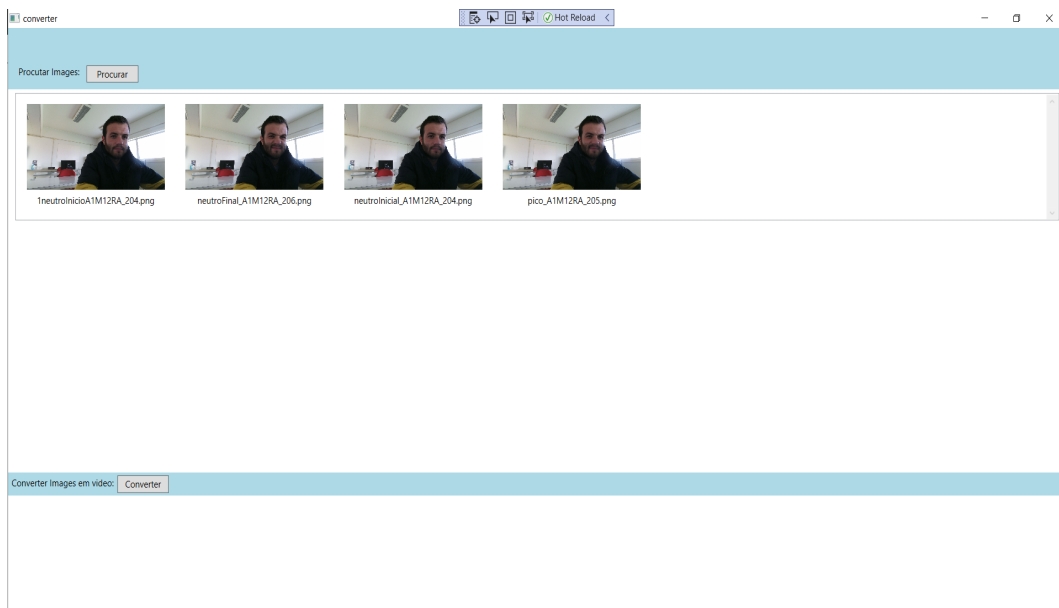


Figura 5.16: Criação de vídeo: Seleccionar os frames normalizados e converter para vídeo.

Capítulo 6

Conclusões

Este trabalho tinha como objetivos criar condições para que uma base de dados de expressões emocionais possa ser adquirida e que correspondesse aos desafios que existem nesta área de estudo, isto é, uma base de dados que fosse totalmente dinâmica e para além disso associar medidas quantitativas que estão associadas às expressões de emoções, que permitam a análise desses dados no futuro.

Nesse contexto, o sistema computacional desenvolvido permite, desde a aquisição à obtenção das medidas quantitativas, um trabalho que se adapta a esses objetivos. É possível através dele, adquirir as sequências faciais de emoções para cada uma das três fases, de forma a obtermos resultados concretos para cada uma delas, ter um sistema de rating onde o participante pode avaliar as suas performances, com a avaliação do experimentador, um sistema de anotação e normalização que permite o pós-processamento dos dados adquiridos e um sistema de conversão dos frames a para vídeo, com sequências de de emoções por normalizar, normalizadas e com os marcos faciais visíveis. Para além disso, é armazenado dados experimentais por cada participante e medidas quantitativas faciais, isto é, action units e marcos faciais para todas as sequências. A obtenção das medidas quantitativas é conseguida através da utilização da ferramenta OpenFace, utilizando as sequências adquiridas pelo sistema desenvolvido.

Os objectivos principais foram conseguidos, temos as condições criadas para obter uma base de dados dinâmica como projectado inicialmente, há contudo, alguns pormenores que ficaram aquém daquilo que foi pensado inicialmente, nomeadamente, melhorias na normalização dos frames, onde poderia existir um maior processamento de forma a obter uma maior precisão dos dados resultantes. Inicialmente foi pensado também obter as medidas quantitativas, como por exemplo action units, diretamente pelo sensor Kinect, no entanto houve a opção de utilizarmos a já pronta ferramenta OpenFace, todavia poderá ser um aspeto útil para o futuro. Há funcionalidades que podem ser adicionadas no futuro, tornando o trabalho ainda mais amplo e completo. A adaptação da ferramenta OpenFace ao sistema computacional criado é um dos casos, a possibilidade da integração num sistema único não cria a necessidade de cruzamento de dados entres dois softwares diferentes. A capacidade de adici-

onar novas medidas quantitativas à experiência, permitiria também alargar os estudos sobre características faciais. A configuração da experiência, criada para este trabalho, permite também a adição de novos dados configuráveis. Com a Kinect é possível extrair mais dados durante gravação, nomeadamente, a utilização do sensor de profundidade, para projectar a face tridimensionalmente, analisando assim modificações que podem acontecer nela.

Perante isto, este trabalho abre portas a novos temas de investigação, temos um sistema e uma base de dados que permite a investigadores de qualquer ramo, desde da psicologia ao estudo criminal, entre outras vertentes, estudar traços faciais de várias emoções, todos com o mesmo objetivo, melhorar a vida quotidiana de todos.

Bibliografia

- [1] Tomkins, S. S. (1962-1963) *Affect, imagery, consciousness* (Vol. 1-2). New York: Springer. doi:10.1037/a0014612
- [2] Silvan S. Tomkins, Robert McCarter What and Where are the Primary Affects? Some Evidence for a Theory, First Published February 1, 1964 Research Article <https://doi.org/10.2466/pms.1964.18.1.119>
- [3] Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008) Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (p. 211–234). The Guilford Press.
- [4] Ekman, P. and Friesen W. V. Constants across cultures in the face and emotion. 1971. In *Journal of Personality and Social Psychology*. 17, 2 (1971) 124.
- [5] Cao Chen, Yanlin Weng, Shun Zhou, Yiyong Tong, Kun Zhou. "FaceWarehouse: a 3D Facial Expression Database for Visual Computing", *IEEE Transactions on Visualization and Computer Graphics*, 20(3): 413-425, 2014
- [6] Hillary Anger Elfenbein, Nalini Ambady Universals and Cultural Differences in Recognizing Emotions, First Published October 1, 2003 Research Article, <https://doi.org/10.1111/1467-8721.01252>
- [7] Ekman, P. *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Times Books. (2003)
- [8] Friesen, W. and Ekman, P. 1983. EMFACS-7: Emotional Facial Action Coding System. (1983) Unpublished manual, University of California, California. (2003).
- [9] Paul Ekman. Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (p. 45–60). John Wiley & Sons Ltd (1999)
- [10] Kurakova, O. A. (2012). A new database of natural transitions between basic emotional-facial expressions. In V. A. Barabanschikov, A. A. Demidov, & D. A. Diveev (Eds.). *Human face as a means of communication: Interdisciplinary approach* (pp. 287–309). Moscow: IPRAS Publ.

- [11] Calvo MG, Fernández-Martín A, Recio G and Lundqvist D Human Observers and Automated Assessment of Dynamic Emotional Facial Expressions: KDEF-dyn Database Validation. *Front. Psychol.* 9:2052. doi: 10.3389/fpsyg.2018.02052 (2018)
- [12] Mislav Grgic, Kresimir Delac, Sonja Grgic SCface - surveillance cameras face database, *Multimedia Tools and Applications Journal*, Vol. 51, No. 3, February 2011, pp. 863-879
- [13] Pfister, Tomas.Li, Xiaobai., Zhao, Guoying.,eAinen, MattiPietik)(SD). Recognising Spontaneous Facial Micro-expressions. Machine Vision Group, Department of Computer Science and Engineering .University of Oulu, Finland, 16 de Julho de 2014.
- [14] Vineetha, G.R., Sreeji, C., Lentin, J Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network. *Trends in Innovative Computing* (2012)
- [15] Youssef, A.E., Aly, S.F., Ibrahim, A.S., Abbott, A.L. Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD T
- [16] Paul Ekman Group Micro Expressions, 2014a.
- [17] Paul Ekman Group Universality of facial expressions, 2014b.
- [18] David Matsumoto and Hyi Sung Hwang Science Brief: Reading facial expressions of emotion.
- [19] Ekman P, Friesen W Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press. (1978).
- [20] Ekman, Paul., Friesen, Wallace., & Ancoli, Sonia. (1980). Facial Signs of Emotional Experience. *Journal of Personality and Social Psychology.* 39 (6, 1125-1134) 14 de Agosto de 2014.
- [21] Ekman, P. (Ed.). Darwin and facial expression; a century of research in review. New York: Academic Press, 1973.
- [22] Darwin, C. (1965). The expression of the emotions in man and animals. Chicago: University of Chicago Press. (Original work published 1872)
- [23] Bettadapura, Vinay (SD). Face Expression Recognition and Analysis: The State of the Art. Georgia Institute of Technology.
- [24] Donato, Gianluca., Bartlett, Marian., Hager, Joseph C., Ekman, Paul and Sejnowski, Terrence. (1999). Classifying Facial Actions. *Ieee transactions on pattern analysis and machine intelligence*, 21.

- [25] Lundqvist, D., Flykt, A., & Öhman, A. The Karolinska Directed Emotional Faces: KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9, 1998.
- [26] Lijun Yin; Xiaozhou Wei; Yi Sun; Jun Wang; Matthew J. Rosato. "A 3D Facial Expression Database For Facial Behavior Research" 7th International Conference on Automatic Face and Gesture Recognition, 10-12 April 2006 P:211 – 216
- [27] Lijun Yin; Xiaochen Chen; Yi Sun; Tony Worm; Michael Reale. "A High-Resolution 3D Dynamic Facial Expression Database": The 8th International Conference on Automatic Face and Gesture Recognition, 17-19 September 2008 (Tracking Number: 66)
- [28] Xing Zhang, Lijun Yin, Jeff Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu, and Jeff Girard "BP4D-Spontaneous: A high resolution spontaneous 3D dynamic facial expression database": Image and Vision Computing, 32 (2014), pp. 692-706 (special issue of the Best of FG13)
- [29] OpenFace 2.0: Facial Behavior Analysis Toolkit Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency, *IEEE International Conference on Automatic Face and Gesture Recognition, 2018*
- [30] Convolutional experts constrained local model for facial landmark detection A. Zadeh, T. Baltrušaitis, and Louis-Philippe Morency. *Computer Vision and Pattern Recognition Workshops, 2017*
- [31] Cross-dataset learning and person-specific normalisation for automatic Action Unit detection Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson *Facial Expression Recognition and Analysis Challenge, IEEE International Conference on Automatic Face and Gesture Recognition, 2015*
- [32] <http://www.aforgenet.com/framework/downloads.html>
- [33] <http://www.aforgenet.com/framework/docs/html/4ee1742c-44d3-b250-d6aa-90cd2d606611.htm>
- [34] https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html

