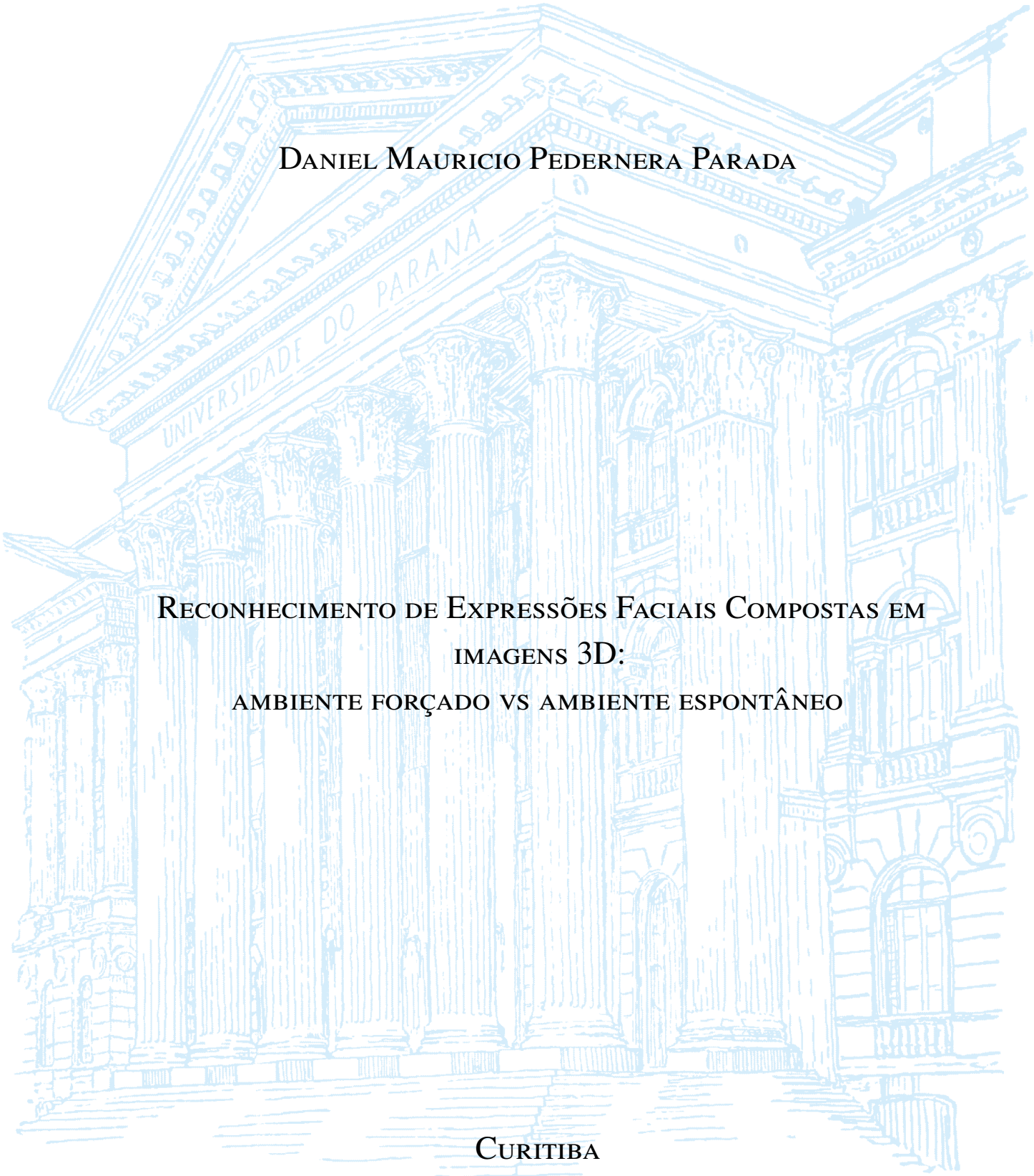
Universidade Federal do Paraná

Daniel Mauricio Pedernera Parada



Reconhecimento de Expressões Faciais Compostas em imagens 3D:
ambiente forçado vs ambiente espontâneo

Curitiba
2017

Daniel Mauricio Pedernera Parada

Reconhecimento de Expressões Faciais Compostas em imagens 3D:
ambiente forçado vs ambiente espontâneo

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática, no Programa de Pós-Graduação em Informática, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Profa. Dra. Olga Regina Pereira Bellon.

Curitiba
2017

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS/UFPR
BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

P222r Parada, Daniel Mauricio Pedernera
Reconhecimento de expressões faciais compostas em imagens 3D: ambiente forçado vs ambiente espontâneo / Daniel Mauricio Pedernera Parada. – Curitiba, 2017.
62 p. : il. color. ; 30 cm.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2017.

Orientadora: Olga Regina Pereira Bellon.

1. Expressões faciais compostas. 2. Detecção de Aus. 3. Ambiente forçado. 4. Ambiente espontâneo. I. Universidade Federal do Paraná. II. Bellon, Olga Regina Pereira. III. Título.

CDD: 006.6

Bibliotecária: Romilda Santos - CRB-9/1214

MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO INFORMÁTICA

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **DANIEL MAURICIO PEDERNERA PARADA** intitulada: **Reconhecimento de Expressoes Faciais Compostas em imagens 3D: ambiente forçado vs ambiente espontâneo**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 18 de Dezembro de 2017.

OLGA REGINA PEREIRA BELLON
Presidente da Banca Examinadora (UFPR)

LUCIANO SILVA
Avaliador Interno (UFPR)

HENRIQUE SERGIO GUTIERREZ DA COSTA
Avaliador Externo (UFPR)

A Deus, a minha Senhora mãe, meu Senhor pai, minha Muñeca, irmãos, parentes e amigos que me apoiaram sempre.

Agradecimentos

Primeiramente a Deus por me dar a força e a sabedoria para encontrar o caminho para conseguir alcançar meus objetivos, sem importar os desafios que tive no percusso.

Aos meus pais Janette del Carmen Parada Mendoza e Juan Carlos Pedernera Canteros, minha Muñeca, e meus irmãos Chrisitan Gabriel Pedernera Parada e Carlos Rodrigo Pedernera Parada; pelo incentivo, pelo constante apoio, pela paciência, pela grande confiança, e por não ter me faltado nunca, mesmo à distância.

À Universidade Federal do Paraná (UFPR), por ter me brindado a oportunidade de desenvolver minha carreira de mestrado, acolhendo-me como a filho próprio, dando-me os meios necessários, tanto materiais quanto educacionais, para o desenvolvimento desta pesquisa.

Aos meus professores, Olga Bellon e Luciano Silva, por me guiarem na jornada desta pesquisa, e me oferecem os recursos para a sua elaboração.

Aos meus colegas do grupo IMAGO, os que me colaboraram, resolveram dúvidas, e estiveram comigo, oferecendo a sua parceria e amizade, em especial ao Giovani Bertolla , quem foi revisor do português deste texto, e um grande amigo-irmão.

A minha grande família, amigos e todos os que estiveram comigo, preocupando-se por minha formação e perguntando constantemente por mim. E um agradecimento especial para Paola Umaña, Neffer Gomes, e Elías Alvizuri que me salvaram quando meu computador morreu.

Finalmente, e mais importante, a todos aqueles que acreditem este pequeno aporte, apresentado em forma de dissertação de mestrado, possa colaborar de alguma maneira na sua formação ou carreira.

Resumo

Neste trabalho, realiza-se o reconhecimento de Expressões Faciais Compostas (EFCs), em imagens 3D, nos ambientes de captura: forçado e espontâneo. Explora-se assim, uma moderna categorização de expressões faciais, diferente das expressões faciais básicas, por ser construída pela combinação de duas expressões básicas. A pesquisa se orienta através da utilização de imagens 3D por conta de suas vantagens intrínsecas: não apresentam problemas decorrentes de variações de pose, iluminação e de outras mudanças na aparência facial. Consideram-se dois ambientes de captura de expressões: forçado (quando o sujeito é instruído para realizar a expressão) e espontâneo (quando o sujeito produz a expressão por meio de estímulos). Isto, com a intenção de comparar o comportamento dos dois em relação ao reconhecimento de EFCs, já que, diferem em várias dimensões, incluindo dentre elas: complexidade, temporalidade e intensidade. Por fim, propõe-se um método para reconhecer EFCs. O método em questão representa uma nova aplicação de detectores de movimentos dos músculos faciais já existentes. Esses movimentos faciais detectar são denotados no sistema de codificação de ação facial (FACS) como Unidades de Ação (AUs). Consequentemente, implementam-se detectores de AUs em imagens 3D baseados em padrões binários locais de profundidade (LDBP). Posteriormente, o método foi aplicado em duas bases de dados públicas com imagens 3D: Bosphorus (ambiente forçado) e BP4D-Spontaneus (ambiente espontâneo). Nota-se que o método desenvolvido não diferencia as EFCs que apresentam a mesma configuração de AUs, sendo estas: "felicidade com nojo", "horror" e "impressão", por conseguinte, considera-se essas expressões como um "caso especial". Portanto, ponderaram-se 14 EFCs, mais o "caso especial" e imagens sem EFCs. Resultados obtidos evidenciam a existência de EFCs em imagens 3D, das quais aproveitaram-se algumas características. Além disso, notou-se que o ambiente espontâneo, teve melhor comportamento em reconhecer EFCs tanto pelas AUs anotadas na base, quanto pelas AUs detectadas automaticamente; reconhecendo mais casos de EFCs e com melhor desempenho. Pelo nosso conhecimento, esta é a primeira vez que EFCs são investigadas em imagens 3D.

Palavras-chave: Expressões faciais compostas, FACS, Detecção de AUs, Ambiente forçado, Ambiente espontâneo.

Abstract

The following research investigates Compound Facial Expressions (EFCs) in 3D images captured in the domains: forced and spontaneous. The work explores a modern categorization of facial expressions, different than basic facial expressions, but constructed by the combination of two basic categories of emotion. The investigation used 3D images because of their intrinsic advantages: they do not present problems due to variations in pose, lighting and other changes in facial appearance. For this purpose, this research considers both forced (when the subject is instructed to perform the expression) and spontaneous (when the subject produces the expression by means of stimuli) expression caption domains. This has the intention of comparing the behavior of both domains by analyzing the recognition of EFCs, because they differ in many dimentions, including complexity, time and intensity. Finally, a method for EFCs recognition is proposed. The method in question represents a new application of existing detectors of facial muscle movements. These facial movimentes to detect are denoted in the Facial Action Coding System (FACS) as Action Units (AUs). Consequently, 3D Facial AUs Detectors are developed based on Local Depth Binary Patterns (LDBP). Subsequently, the method was applied to two public databases with 3D images: Bosphorus (forced domain) and BP4D-Spontaneous (spontaneous domain). Note that the developed method does not differentiate the EFCs that present the same AU configuration: "sadly disgusted", "appalled" and "hateful", therefore, these expressions are considered a "special case". Thus, 14 EFCs are observed, plus the "special case" and the non-EFCs images. The results confirm the existence of EFCs in 3D images, from which some characteristics were exploit. In addition, noticed that the spontaneous environment was better at recognizing EFCs by the AUs annotated at the database and by the AUs detected; recognizing more cases of EFCs and with better performance. From our best knowledge, this is the first time that EFCs are explored for 3D images.

Keywords: Coumpound facial expression, FACS, AUs detection, posed domain, spontaneous domain.

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

APDI Imagens de Projeção de Distância Azimutal
AU Unidade de Ação
AUC Área sob a curva
BP4D *Binghamton-Pittsburgh 3D Dynamic Spontaneous Facial Expression Database*
CS-3DLBP Padrões binários locais de centro simétrico
EFCs Expressões facias compostas
FACS Sistema de codificação de ação facial
HOG Histogramas orientados a gradientes
LDBP Padrões binários locais de profundidade
LNBP Padrões binários locais Normalizado
NCMML Classificador de média de classe mais próxima
ROC Característica de operação do receptor
SVM Máquinas de suporte vetorial
VTK *Visualisation Toolkit*

Sumário

1 Introdução

A principal via de comunicação não-verbal entre grandes primatas, incluindo a espécie humana, acontece pela interpretação e/ou execução de padrões musculares faciais [8, 9]. As pessoas, independentemente das caraterísticas culturais que possuem [10], podem ter uma comunicação não-verbal e expressar suas emoções por meio de articulações naturais, que envolvem contrações de grupos musculares na face. De acordo com Ekman [8], estes movimentos musculares tornam-se em componentes não-verbais muito importantes para ter uma boa comunicação, pois diminuem o desentendimento da mensagem a ser transmitida. No entanto, no processo de transmitir emoções, o individuo deve ser capaz de reconhecer e/ou expressar as ações dos músculos da face de forma correta; caso esse processo analítico falhe, qualquer outro processo envolvido na comunicação estará comprometido, evitando que as verdadeiras emoções transmitidas pela face sejam corretamente interpretadas.

Além disso, nota-se que estes componentes da comunicação não-verbal colaboram na comunicação verbal. Segundo Mehrabian [1], as pessoas conseguem comunicar-se por meio da fala, mas são as expressões faciais que transmitem a emoção da mensagem. Em uma conversação entre dois sujeitos, por exemplo, uma mensagem transmitida pelo locutor, é compreendida pelo ouvinte, existindo três fatores diferentes que influenciam o entendimento da mesma: (1) Pelas expressões faciais, (2) pela maneira que as palavras são ditas, (3) e pelas próprias palavras ditas. Consequentemente, cada um desses fatores tem um grau de influência no entendimento da mensagem, sendo que 55% do entendimento da frase dita é influenciado pelas expressões faciais (1), enquanto que pelo modo que as palavras são faladas (2) influencia em 38%, e as palavras em si (3) têm a importância de apenas 7% na compreensão da mensagem dita. Isto reflete-se no diagrama da Figura 1.1.

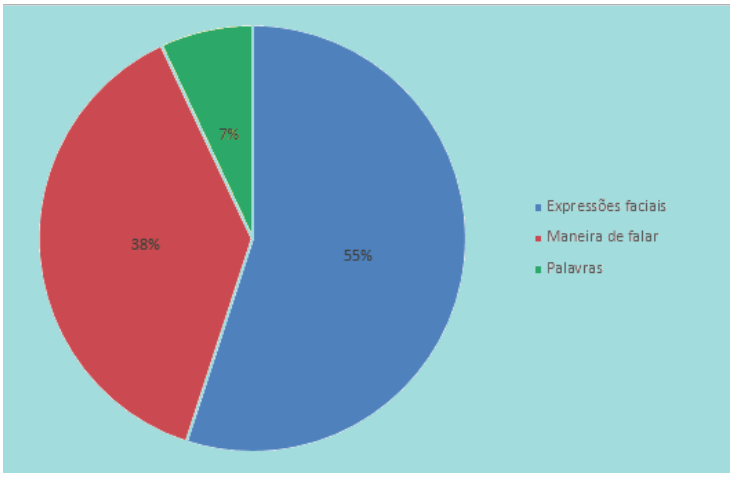


Figura 1.1: Componentes de influência em uma mensagem transmitida oralmente segundo [1], onde claramente observa-se que expressões faciais tem um papel muito importante no entendimento da mensagem (um 55% de influência)

Expressões faciais são consideradas a maneira mais indicativa, forte e natural de conhecer o estado psicológico de uma pessoa durante a comunicação. Isto quer dizer que expressões faciais jogam um papel principal na comunicação humana, mas também são importantes em áreas como: psicoterapia, educação, animação gráfica, etc. Por conta disso, abordagens de reconhecimento de expressões faciais humana estão sendo cada vez mais consideradas em processamento de imagens e interação humano-computador [11, 5].

Observando a importância que desempenham as expressões faciais na comunicação de emoções, Ekman e Friesen definem o Sistema de Codificação de Ação Facial (FACS) [12, 13, 14, 15, 8] que é utilizado para analisar a produção das diferentes categorias de expressões faciais. O FACS contribui em uma representação clara e compacta da ativação muscular nas expressões faciais por meio das Unidades de Ação (AUs), que são os movimentos visualmente menores e discrimináveis dos músculos faciais individuais ou grupais. Resultando estas unidades em um "alfabeto" muito importante a dominar, sendo este o primeiro passo para utilizar na comunicação não-verbal [16, 17]. Um exemplo da representação com AUs de movimentos musculares pode-se observar na Figura 1.2.

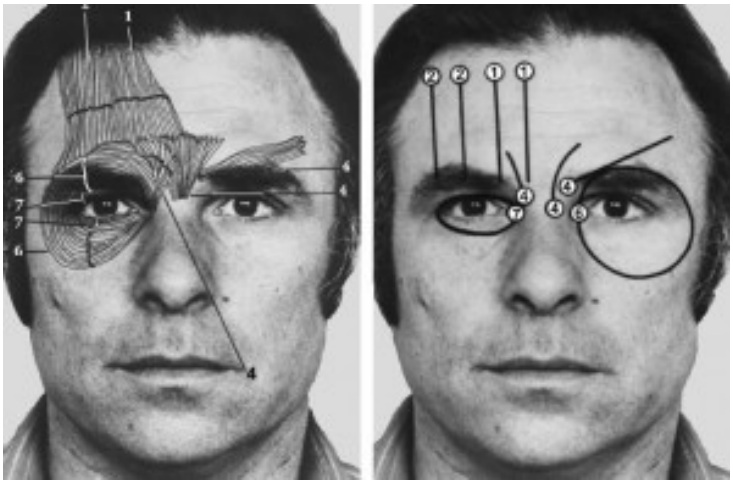


Figura 1.2: Exemplo de algumas Unidades de Ação (AUs), anotadas em uma imagem 2D com um rosto humano [2]

Portanto, por meio da aplicação do FACS, Ekman e Friesen identificam seis expressões faciais básicas (felicidade, tristeza, surpresa, medo, raiva, nojo). Na literatura pode-se observar que existe um grande grupo de pesquisas focadas nessas expressões [18, 19, 20]. Por básicas entende-se que são emoções discretas para eventos da mesma categoria de emoção, mas não significa que essas categorias sejam menos complexas que outras. Um exemplo da expressão básica de "felicidade"pode-se notar a esquerda da Figura 1.3.

Além disso, por meio de uma configuração padrão de AUs, são identificadas 17 Expressões Faciais Compostas (EFCs) (felicidade com surpresa, felicidade com nojo, tristeza com medo, tristeza com raiva, tristeza com surpresa, tristeza com nojo, medo com raiva, medo com surpresa, medo com nojo, raiva com surpresa, raiva com nojo, nojo com surpresa, horror, ódio, impressão, felicidade com medo e felicidade com tristeza) [3, 21], assim como as expressões básicas, que são expressões humanas de emoção, mas representam um conjunto maior de emoções comuns entre indivíduos humanos. Estas expressões são compostas por expressões básicas. A direita da Figura 1.3 é possível observar um exemplo da expressão composta "felicidade com supressa".

Percebe-se na literatura vários trabalhos focados em utilizar apenas imagens 2D [3, 11, 22, 23], apesar dessas imagens apresentarem limitações devido a variações de pose e

Figura 1.3: Exemplos de expressões faciais em 2D: (esq) expressão básica de felicidade, (dir) expressão composta de felicidade com surpresa [3]

iluminação própria e outras mudanças na aparência facial (como maquiagem, cabelo, ou barba) [14]. A fim de lidar com esses problemas, imagens 3D e 4D (3D Dinâmico) são cada vez mais utilizadas em pesquisa e análise de expressões [24]. A face é um objeto 3D, que apresenta muitos sinais comunicativos que envolvem mudanças na profundidade e na rotação da cabeça, sendo assim, a inclusão de informação 3D contribui com dados importantes para a solução de problemas que imagens 2D têm dificuldade para resolver [21].

Da mesma forma, observou-se que expressões faciais podem ser espontâneas ou forçadas. Claramente no dia-a-dia, é mais comum que as pessoas apresentem expressões espontâneas ao invés de expressões forçadas, pois uma expressão forçada pode não transmitir a verdadeira emoção. Segundo Zhang *et al.* [6] expressões forçadas e espontâneas diferem em várias dimensões, incluindo a complexidade, tempo e intensidade. Du *et al.* [21] afirmam que os sistemas neurais envolvidos na produção de expressões forçadas, e espontâneas são diferentes, mas as AUs da configuração padrão de EFCs são as mesmas.

Assim, sob os estudos mencionados acima, propõe-se realizar um método capaz de buscar e reconhecer automaticamente EFCs em imagens 3D, nos ambientes de captura: forçado e espontâneo, obtendo desempenho estado-da-arte. A ideia principal consiste em encontrar as expressões compostas, utilizando a configuração padrão de AUs definida em [3, 21], para posteriormente estudar como se produzem as EFCs nos dois ambientes de captura, e como são reconhecidas automaticamente. Para tal fim, consideram-se duas bases de dados públicas com imagens 3D: Bosphorus [5] (ambiente forçado) e BP4D-Spontanous [6] (ambiente espontâneo). O reconhecimento de EFCs realiza-se por meio da identificação de AUs, por conta disso, um dos objetivos para o prosseguimento desta pesquisa resulta em explorar detetores de AUs em 3D já existentes, que serão base para desenvolver novos detetores.

O restante desta dissertação de mestrado está organizado da seguinte maneira: no Capítulo 2, apresentam-se termos chave relativos a expressões faciais compostas, e a detecção de AUs em imagens 3D; o Capítulo 3, contempla os detectores de AUs em imagens 3D que posteriormente serão aplicados para o reconhecimento de EFCs. Assim, detalham-se: várias abordagens relativas na literatura, as bases de dados utilizadas nos experimentos, o método implementado para os detectores de AUs, os experimentos executados, e finalmente os resultados obtidos; O aporte da pesquisa será relatado no Capítulo 4 especificando: o método proposto para reconhecer EFCs em imagens 3D, os experimentos realizados, os resultados, as observações do proceder do método, e por último, uma comparação de comportamento no reconhecimento de EFCs nos ambientes forçados e espontâneos. Finalmente no Capítulo 5 são debatidas as conclusões do trabalho. Nota-se que os Capítulos 3 e 4, apresentam seções independentes de

método, experimentos, e resultados, já que foram implementados dois métodos diferentes, cada um deles com seus próprios experimentos e resultados, e acredita-se que unir as seções pode gerar no leitor confusão para relacionar as partes correspondentes.

1.1 Objetivos e Metas

O objetivo geral da presente pesquisa é desenvolver um método para o reconhecimento de expressões faciais compostas em imagens 3D.

Os objetivos específicos são a) realizar uma revisão bibliográfica sobre expressões faciais com maior foco nas EFCs; b) especificar e desenvolver um método automático para o reconhecimento de EFCs; c) aplicar o método desenvolvido em bases de dados com imagens 3D, capturadas nos ambientes forçados e espontâneos; d) analisar os dados e métricas obtidos nos experimentos executados.

As metas deste trabalhos são: a) aportar conhecimento na área de reconhecimento de expressões faciais em visão computacional; b) desenvolver um método automático estado-da-arte para reconhecimento de EFCs em imagens 3D; c) diferenciar como são produzidas as EFCs nos ambientes de captura forçados e espontâneos.

2 Fundamentos Teóricos

No presente Capítulo serão abordados alguns termos chaves para o entendimento do estudo de expressões faciais. Dessa maneira, descreve-se a sistematização padrão de expressões faciais de emoção, e uma classificação das mesmas. Adicionalmente, são explicadas algumas técnicas úteis para o processamento e classificação de imagens, que serão aproveitadas para alcançar o objetivo da presente pesquisa; o reconhecimento automático de EFCs em imagens 3D.

2.1 Sistema de Codificação de Ação Facial - FACS

O Sistema de Codificação de Ação Facial (FACS) foi definido por Ekman e Friesen em [12]. Anteriormente, já em 1970, Birdwhistell em [25], afirmou que as expressões faciais podem ser observadas como uma outra linguagem que possui as mesmas unidades e organização que qualquer uma das línguas faladas. Portanto, o FACS foi definido como um sistema que analisa a produção das diferentes categorias de expressões faciais. O sistema FACS contribui para uma representação clara e compacta da ativação muscular nas expressões faciais por meio das Unidades de Ação (AUs), que são os movimentos, explicitamente menores visualmente, dos músculos faciais individuais ou grupais. Contudo, Ekman e Keltner em [10] afirmaram que nem todos os movimentos dos músculos da face representam expressões de alguma emoção; ações faciais não são exclusivamente dedicadas a essa finalidade. Yudim em [26], afirma que Ekman e Frisen reconhecem a necessidade de codificar todas as configurações faciais possíveis, definindo assim, um próprio "alfabeto" para escrever "palavras", que possibilitem uma descrição de todas as configurações faciais que os humanos possam representar.

O FACS tem progredido através de três versões: a versão inicial (FACS 1978), uma atualização (FACS 1992), e uma nova edição (FACS 2002) sendo que na última versão são especificadas nove unidades de ação na parte superior da face e 18 na parte inferior. Adicionalmente, 14 movimentos de cabeça, nove posições e movimentos dos olhos, cinco unidades de ação de músculos que se movem simultaneamente, nove descritores de ação, nove comportamentos brutos e cinco códigos visíveis. Além disso, na versão de 2002 do FACS são definidos cinco níveis de intensidades (A, B, C, D, E). Sendo "A" a menor intensidade e "E" a maior. Pode ser subjetivo definir diretrizes para codificar intensidades nas AUs e requer muito esforço para estabelecer e manter níveis aceitáveis de relatividade; a importância de codificar intensidades vai depender da natureza do estudo das AUs. [14].

2.2 Expressões Faciais de Emoção

As expressões faciais representam um componente muito importante na comunicação verbal e não-verbal humana, pois são uma manifestação visível do estado afetivo, da atividade cognitiva, da intenção, da personalidade e da psicopatologia de uma pessoa. Assim, entendê-las

torna-se em um papel importante na comunicação social e outras áreas como: psicoterapia, educação, animação gráfica, etc [5]. Caso o processo de entender expressões faciais falhe, qualquer outro processo envolvido na comunicação será afetado, evitando que o verdadeiro sentido da emoção a ser transmitida seja corretamente interpretado [8].

Martinez e Valstar em [27] mencionam que os humanos percebem e interpretam as expressões faciais em termos de modelos mentais de emoção, estados afetivos, sinais sociais ou indicadores de saúde. Por causa disso, surgiram muitas teorias para codificar, representar e interpretar expressões faciais, inclusive em Visão Computacional. Na presente pesquisa, foram considerados dois tipos de expressões faciais: básicas e compostas, que serão brevemente descritas a seguir.

2.2.1 Expressões Faciais Básicas

Ekman e Friesen em [12] realizaram um estudo profundo sobre as expressões faciais, identificando seis delas, as quais são universais, e são classificadas como expressões faciais básicas. Estas expressões são: felicidade, tristeza, surpresa, medo, raiva e nojo. Atualmente, na literatura, vários estudos são focados neste tipo de expressões [18, 19, 20].

Segundo Du *et al.* em [21], entende-se que as expressões acima mencionadas são expressões faciais de emoções discretas e têm evoluído para ajudar aos seres humanos nas mudanças ambientais. Também citam que são uma resposta automática para eventos da mesma categoria de emoção, definidas por um simples componente de emoção (Ex. Felicidade). Porém, o fato de ser chamadas de "básicas" não significa que essas categorias sejam mais básicas que outras.

Yudin em [26] afirma que a invariância cultural ou universalidade das expressões faciais básicas tem implicações significativas de que pode haver uma base genética para a produção de expressões em humanos. Da mesma maneira, menciona que além dessas seis expressões básicas, Ekman e Friesen não encontraram outra correspondência entre configuração facial e intenção comunicativa que seja universal.

2.2.2 Expressões Faciais Compostas - EFCs

Du *et al.* em [3, 21], propõem uma análise de um outro tipo de expressões faciais, as quais são denominadas como Expressões Faciais Compostas (EFCs), por acreditar que nem todas as expressões faciais precisam estar associadas a uma única emoção, como acontece nas expressões faciais básicas, e que inclusive pode existir emoções contraditórias que sejam expressas pelos seres humanos em um determinado momento (ex. A expressão que é gerada por uma brincadeira que é divertida mas é ruim ao mesmo tempo).

Assim, nota-se, no primeiro trabalho proposto por Du *et al.* [3], são estudadas as configurações de AUs para 15 EFCs em uma base de dados com imagens 2D capturadas em ambiente controlado. Além disso, as imagens apresentam expressões faciais forçadas. No segundo trabalho de Du *et al.* [21], é apresentado o estudo de EFCs acrescentando duas expressões compostas e fazendo uma avaliação da configuração de AUs para expressões espontâneas, complementado dessa forma o trabalho anterior. O foco desse trabalho [21] é apresentar as EFCs desde o mais básico até aplicações médicas. Além disso, são identificadas mais duas categorias de EFCs: felicidade com medo e felicidade com tristeza. Também é realizada uma análise de existência de EFCs em imagens 2D com expressões espontâneas, para saber se são produzidas pelas mesmas AUs que as expressões compostas capturadas no laboratório. Foram utilizadas imagens 2D da *web*, obtidas por meio da busca com palavras chave (ex.: felicidade,

felicidade com nojo, felicidade e nojo, etc.). Quando uma imagem com expressão composta foi qualitativamente encontrada, a mesma foi quantitativamente avaliada e as AUs correspondentes foram manualmente anotadas. Como conclusão, os autores relatam que o padrão de ativação de AUs para expressões espontâneas é idêntico ao padrão observado no laboratório.

2.3 Padrões Binários Locais de profundidade - LDBP

Baseados nos padrões binários locais (LBP) definidos por Ojala *et al.* em [4] como descritores de textura, Sandbach *et al.* em [7] definem os padrões binários locais de profundidade (LDBP). Estes descritores são aplicados em mapas de profundidade, explorando assim, uma representação 2D da informação 3D. Nota-se que o custo de processo dos mesmos é igual ao de processar imagens 2D ao invés de imagens 3D. Na presente pesquisa (e no *baseline*) são utilizados para a detecção de AUs em imagens 3D.

Assim, define-se para cada pixel de um mapa de profundidade, uma vizinhança circular por P pontos ao redor de cada pixel de profundidade em um raio r. O valor do pixel central é subtraído do valor dos pixels da vizinhança, atribuindo 0 se o resultado for negativo, e 1 caso contrário. Posteriormente, unidades binárias são dispostas no sentido horário, definindo o valor binário que finalmente será convertido em decimal para o pixel central. Um exemplo do processo pode ser visto na figura 2.1.

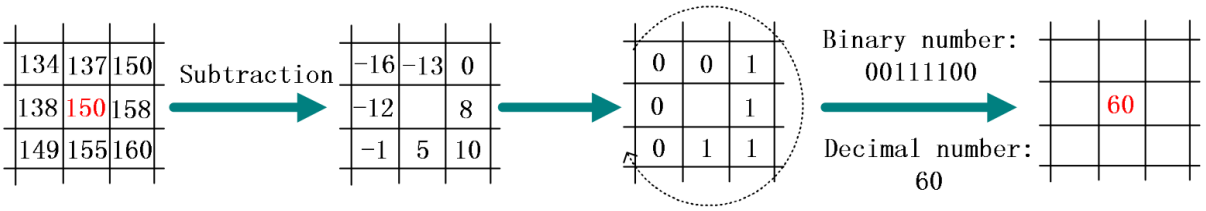


Figura 2.1: Exemplo do operador LBP, extraída de [4]

Considerando I^D como o mapa de profundidade, aplica-se LDBP para o pixel central $I^D(x_c, y_c)$, com uma vizinhança de P pixels $I^D(x_p, y_p)$ para $p = 0, ..., P\text{-}1$, e formalmente define-se:

$$LDBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(I^D(x_p, y_p) - I^D(x_c, y_c)) \tag{2.1}$$

$$s(v) = \begin{cases} 1 & v \geq 0 \\ 0 & v < 0 \end{cases} \tag{2.2}$$

2.4 Mapas de profundidade

Os mapas de profundidade representam uma projeção de imagens 3D, porém, apresentam características das imagens 2D. Tais mapas são considerados imagens 2.5D, pois são imagens 2D, mas conservam informação de profundidade, a qual é representada pela intensidade dos pixels. Conforme Vretos *et. al.* em [28], os mapas de profundidade são uma representação 2D de cenas de imagens 3D, que têm sido amplamente aplicados em análise facial [29, 7, 30, 31], nos quais a intensidade de cada pixel codifica a profundidade relativa do plano de referência.

No método desenvolvido, os vértices de cada imagem das bases com imagens 3D utilizadas foram projetados no plano X-Y. Cada vértice está representado nas coordenadas "x",

"y", "z" . Assim, a projeção das coordenas "x", "y" realizou-se no plano X-Y nos mapas de profundidade com resolução 300x300. A coordenada "z" foi aplicada para definir a profundidade em tons de cinza, sendo que tons mais claros representam os pontos em "z" mais próximos à câmera. Um exemplo de mapa de profundidade exibe-se na figura 2.2.

Figura 2.2: Exemplo de mapa de profundidade obtido a partir da projeção da imagem 3D do sujeito bs000 da base Bosphorus executando a AU34

2.5 Máquinas de suporte vetorial (SVM) binárias

De acordo com [32], as Máquinas de vetores de suporte (SVMs) são métodos de aprendizagem supervisionado, aplicados na classificação linear e regressão. Estes métodos são baseados na optimização da margem delimitada com os de vetores de suporte. A margem descreve uma região do espaço de características na qual não existem amostras, e o plano intermédio dessa região, que proporciona maior amplitude, é a solução do SVM. Na Figura 2.3, observa-se um exemplo de SVM, no qual são classificadas duas classes (verde e vermelha) pela margem (região amarela), delimitada pelos vetores de suporte (linhas tracejadas), sendo hiperplano intermédio representado como a linha continua no meio dessa área.

Inicialmente, considera-se um conjunto M com dados de entrada $\{x_1, ..., x_M\}$ e etiquetas $\{y_1, ..., y_M\}$, sendo $\{x_i \in \mathbb{R}^m\}$ e $y_i \in \{-1, 1\}$, SVM tem por objetivo encontrar uma função f, maximizando a distância entre o hiperplano ótimo e os vetores de suporte.

A referida maximização se alcança pela equação:

$$f(x) = w^T x + \omega_0 = 0 \tag{2.3}$$

Onde w é um vetor perpendicular ao hiperplano e ω_0 é uma variável para maximizar a margem do hiperplano.

Como o objetivo principal é encontrar uma margem maior, é necessário atender os hiperplanos paralelos ao hiperplano ótimo mais próximos aos vetores de suporte das classes, contidos em:

$$h(x)^+ = w^T x + \omega_0 = 1 \tag{2.4}$$

$$h(x)^- = w^T x + \omega_0 = -1 \tag{2.5}$$

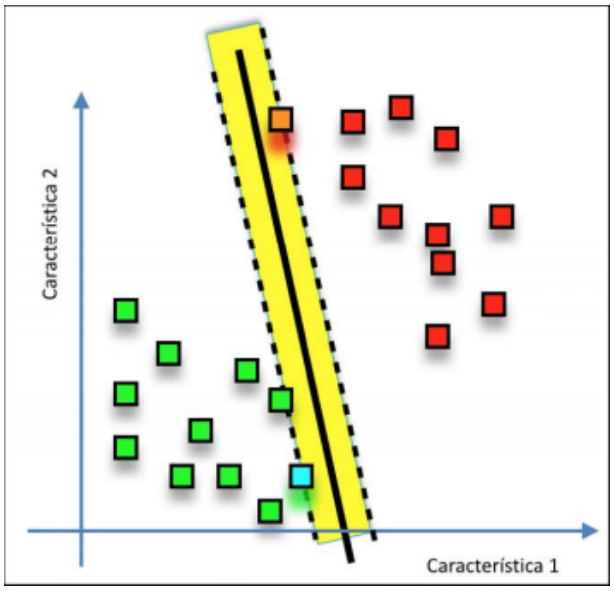


Figura 2.3: Classificador SVM para dois tipos de características (pontos de cor verde e vermelha), separadas pela margem (região amarela), delimitada pelos vetores de suporte (linhas tracejadas), sendo o hiperplano de solução a reta no meio dessa região. Os pontos laranja e azul são exemplos de características que caem sobre o limite da região margem

Dado que o conjunto de dados de aprendizado é linearmente separável, os hiperplanos 2.4 e 2.5 podem ser selecionados, maximizando a distância entre eles, de tal maneira que não hajam pontos nesse intervalo. Essa distância é representada como: $\frac{2}{|\omega|}$. Dessa forma, o valor de $|\omega|$ deve ser minimizado, garantido que cada vetor x de entrada seja da forma:

$$\begin{cases} w^T x + \omega_0 \geq 1 & \text{se } y = 1 \\ w^T x + \omega_0 \leq -1 & \text{se } y = -1 \end{cases} \tag{2.6}$$

Assim, resulta útil comprovar que maximizar a margem pode ser equivalente a minimizar o problema inverso, que se denomina como SVM primal, e se expressa da seguinte maneira:

$$\begin{aligned} \text{Minimizar} \quad & g(w) = \frac{1}{2} \parallel w \parallel^2 \\ \text{Dependendo de} \quad & \parallel w \parallel^2 = w^T w \\ \text{Considerando} \quad & y_i(w^T x_i + \omega_0) \geq 1 \quad i = 1, 2, 3, ..., M \end{aligned} \tag{2.7}$$

Sendo que a solução para o problema 2.7, é a minimização da equação de Lagrange adaptada as condições anteriores, respeito as variáveis w e ω_0, e posteriormente maximizando em relação aos multiplicadores de Lagrange, como se representa a seguir:

$$\max_{\alpha}(\min_{w,\omega_0}(L(w, \omega_0, \alpha)) \tag{2.8}$$

A equação de Lagrange referida é representada como:

$$L(w, \omega_0, \alpha) = \frac{1}{2} w^T w - \sum_i (\alpha_i [y_i (w^T x_i + \omega_o) - 1]) \tag{2.9}$$

Nota-se que a equação 2.9, introduz os α_i, os quais representam os multiplicadores de Lagrange. Sendo estes importantes, pois definem quais amostras constituem os vetores de suporte, considerando $\alpha_i > 0$

Dese modo, a minimização da equação 2.9, implica realizar derivadas parciais ao respeito das variáveis w e ω_0, igualadas a zero, resultando em:

$$w = \sum_i \alpha_i y_i x_i \tag{2.10}$$

$$\omega_0 = y_i - w^T x_i \tag{2.11}$$

Para por fim obter uma nova função solução, a qual é conhecida como SVM dual, e se representa como:

$$\begin{aligned} \text{Maximizar } \ d(\alpha) &= -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i \\ \text{Sujeito a } \quad \sum_i \alpha_i y_i &= 0 \ \ para \ \ cada \ \ \alpha_i \geq 0 \end{aligned} \tag{2.12}$$

Dando como resultado o classificador SVM, como se detalha a seguir:

$$svm(x) = sgn(\sum_i y_i \alpha_i x^T x_i + \omega_o) \tag{2.13}$$

No caso em que os dados não sejam linearmente separáveis, pode-se definir um *kernel* (ou núcleo), que assuma a forma: $k(x, z) = \phi(x)^T \phi(z)$, e que seja integrado ao SVM dual 2.12, gerando as seguintes mudanças:

$$\begin{aligned} \text{Maximizar } \ d(\alpha) &= -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j) + \sum_i \alpha_i \\ \omega_i &= y_i - \sum_j y_i \alpha_i k(x_i, x_j) \\ svm(x) &= sgn(\sum_i y_i \alpha_i k(x_i, x_j) + \omega_o) \end{aligned} \tag{2.14}$$

Na implementação do método de detecção de AUs da seção 3.1 é empregando o núcleo de interseção de histogramas, o qual é estândar, e está definido como:

$$k(x, z) = \sum_{i=1}^{n} \min(x_i, z_i) \tag{2.15}$$

3 Detecção de Unidades de Ação (AUs) em imagens 3D

Para poder desenvolver um método para reconhecer EFCs, que é um dos principais objetivos da pesquisa, considera-se a exposição do Liu *et al.* em [33], onde mencionam que métodos de reconhecimento de expressões faciais podem ser classificados aproximadamente em dois grupos: baseados em unidades de ação (aqueles que consideram as AUs componentes para reconhecer a expressão) e baseados em aparência (aqueles que consideram a imagem inteira da face para encontrar a expressão). Consequentemente, decidiu-se optar por métodos baseados na detecção de AUs para o desenvolvimento da proposta, visto que, atualmente, não existem bases com imagens 3D orientadas a esse tipo de expressões. Então, para poder reconhecer EFCs em imagens 3D, a principio as AUs devem ser detetadas, para posteriormente, compará-las com a configuração padrão de AUs de EFCs presente em [21], e assim poder determinar a expressão facial composta, como acontece no procedimento realizado em [34].

Como o método para reconhecer EFCs baseia-se na detecção de AUs, entende-se que é muito importante desenvolver um sub-método que permita detectar os movimentos musculares faciais em imagens 3D. Portanto, inicialmente realiza-se o levantamento de algumas abordagens na literatura para detecção de AUs em imagens 3D, estas são apresentadas na seção 3.1. Também, são estudadas bases de dados 3D que possam ser uteis tanto na detecção de AUs quanto na análise e reconhecimento de EFCs, por conta disso, foram escolhidas as bases Bosphorus e BP4D-Spontâneous, as mesmas serão brevemente detalhadas na secção 3.2. Após isso, desenvolve-se uma implementação baseada no trabalho proposto em [7]. Esta proposta aplica-se como sub-método para detectar AUs, e está constituída pelos seguintes estados: pré-processamento (para diminuir ruídos), alinhamento (das nuvens de pontos com os pontos fiduciais providos nas bases), extração de características (por meio da aplicação de Padrões Binários Locais em mapas de profundidade (LDBP)), e treinamento de classificadores binários SVM. Na seção 3.3 descreve-se com maior detalhe cada componente do sub-método implementado como detector de AUs.

3.1 Trabalhos relacionados

Existem muitas abordagens para detecção de AUs, inclusive para 3D. Nota-se que trabalhos no estado da arte têm varias semelhas na implementação, como ser: a utilização variantes dos Padrões Binários Locais (LBP) [35] como descritores de textura, a aplicação de AdaBoost (ou GentleBoost) na seleção de características, a construção de vetores 1D de por meio de HOG, o emprego de SVM como classificadores, a execução de validação cruzada em 10-*fold* para a geração dos *datasets* de treinamento e teste, e inclusive a consideração de imagens da base Bosphorus. Estas metodologias para detectar AUs em imagens 3D, serão brevemente detalhadas a seguir:

- Savran *et al.* em [36] fazem um estudo comparativo entre espaços 2D e 3D para a detecção de AUs. O desafio a ser explorado consistiu em observar quais das modalidades (imagens 2D ou 3D) são melhores, vendo inclusive a sua complementaridade. Para tal fim, as bases Cohn-Kanade DFAT-504 [37] (2D), e Bosphorus (3D) foram utilizadas, sendo identificadas 19 AUs na primeira, e 25 na segunda. Então, imagens 3D foram mapeadas em 2D mediante curvaturas de superfície. Esta metodologia oferece informação fiel da geometria do rosto de maneira compacta. Logo, para a detecção das AUs nas imagens 2D (inclusive nas 3D mapeadas) foi empregado o método de Bartlett *et al.* [38], já que, o resultado da comparação não depende do desempenho das etapas intermediárias. Na fase de treinamento, foi aplicado AdaBoost para a seleção de características, e para a classificação: AdaBoost, linear SVM, e quatro classificadores Bayes. Todos os experimentos foram realizados utilizando validação cruzada em 10-*fold*. Os *datasets* foram compostos por qualquer imagem do rosto com a AU alvo, isolada ou em combinação com outras AUs, sendo esta tratada como uma amostra positiva dessa classe de AU, enquanto que todas as outras imagens que não envolveram a AU alvo foram aceitas como amostras negativas. Os melhores resultados foram obtidos com SVM linear (93.7% na Cohn-Kanade DFAT-504, 93.5% nas imagens 2D da Bosphorus, e 95.4% nas imagens 3D da Bosphorus). Os autores conseguiram determinar que, em geral, imagens 3D funcionam melhor para AUs que se encontrem na parte inferior do rosto, mas que imagens 2D têm melhor desempenho com AUs perto da região dos olhos, e que a combinação de características 2D e 3D se mostrou vantajosa à medida que o desempenho médio da detecção de AU aumentou (de 95.4% para 97.1%).

- Sandbach *et al.* em [29] propõem a técnica de Padrões Binários Locais Normalizados (LNBPs) para a detecção de AUs. LNBPs emprega as normais de polígonos triangulares que formam as malhas faciais 3D para codificar a forma em cada ponto. Assim, definiram dois descritores binários: $LNBP_{OA}$ (que calcula o produto escalar de duas normais) e LBP_{TA} (que calcula a diferença de dois ângulos das normais, o azimute e a elevação). Então, inicialmente imagens 3D foram alinhadas por meio de matrizes afins, para posteriormente com o plano x-y das malhas, definir os vetores de características pela aplicação dos descritores: 3DLBP (o qual é utilizado em reconhecimento facia [39]), $LNBP_{OA}$, LBP_{TA}, e pela combinação deles, por meio da utilização de histogramas. A seleção de atributos foi desenvolvida via Gentle-Boost (GB), e máquinas de suporte vetorial (SVMs) foram treinadas como classificadores. Os testes foram realizados na base Bosphorus, aplicando validação cruzada em 10-*fold*. Os resultados obtidos demonstraram valores em média similares para os três descritores, sendo levemente maior o alcançado pelo 3DLBP (95.21%). Porém, para muitos dos casos, os descritores LNBPs tiveram melhor desempenho em detectar AUs individualmente. Sem embargo, quando combinados 3DLBP com $LNBP_{OA}$, os resultados gerais (e alguns individuais) melhoraram (96.35%), devido à natureza complementar dos resultados de cada descritor.

- Sandbach *et al.* em [7] utilizam oito padrões binários para detecção de AUs em 3D transformando as imagens 3D em duas representações 2D que contêm informação da geometria facial: mapas de profundidade e Imagens de Projeção de Distância Azimutal (APDI). Posteriormente, aplicaram para cada representação 2D os descritores: LBP, fase de quantização local (LPQ), filtros Gabor e filtros monogênicos para extração de características. Depois disso, estas características foram concatenadas em histogramas e selecionadas por meio de GB, para finalmente serem treinadas em SVM. O método foi testado nas bases Bosphorus (por meio da validação cruzada em 10-*fold*), e D3DFACS

[40] (através de bancos de dados distintos). A finalidade foi comprovar efetividade quando comparado com o descritor 3DLBP. Assim, autores confluíram que a maioria de descritores tiveram melhor desempenho que o padrão 3DLBP, sendo o resultado mais alto alcançado pelo Local Depth Gabor Binary Patterns (LDGBPs) em ambas as bases (97.2% para a Bosphorus e 89.4% para a D3DFACS). Adicionalmente, observaram que foi mais difícil generalizar características baseadas em APDI para a base de dados D3DFACS, devido a variações de suavidade com a base Bosphorus.

- Bayramoglu *et al.* em [41] propõem um método que combina um descritor baseado em LBP e propriedades geométricas faciais, para a detecção das AUs. Para tal fim, definem os descritores: Padrões Binários Locais em 3D de Centro Simétrico (CS-3DLBP) e Geometria Baseada em Razões (RBG). O primeiro está baseado em região eficiente, onde valores dos ângulos entre pontos 3D opostos são comparados e as diferenças são consideradas limiares, enquanto que o segundo considerara relações de distancia, áreas, e ângulos nas faces 3D; concatenando 24 características em um vetor simples. Para testar o método utilizaram a base Bosphurus, considerando validação cruzada em 10-*fold*. Assim, as imagens depois de um pré-processamento, foram projetadas em mapas de profundidade, para posteriormente, serem aplicados os dois descritores, definindo as características, as quais foram classificadas por meio de Florestas Aleatórias. Para cada AU, avaliaram os descritores CS-3DLBP e RBG separadamente, mas também em combinação, concatenando os vetores de características. Resultados mostraram altos valores de detecção, sendo o melhor (97.7%) o alcançado pela combinação de ambos os descritores (CS-3DLBP + RBG). Autores concluíram que o método resultou ser melhor detectando AUs da parte inferior da face.

- Yudin *et al.* em [42] descrevem um novo uso de um *framework* geométrico para estender o conceito de normalização de dados ao domínio de superfícies, aplicando-o para a detecção de AUs em imagens 3D, na base Bosphorus. A normalização possui meios efetivos para reduzir a variação de observações de uma classe de fenômenos, diminuindo a dimensionalidade e o número de graus de liberdade necessários para desenvolver modelos matemáticos que identifiquem instancias dessas classes com êxito. Sendo assim, a geometria facial foi normalizada para reduzir toda a variabilidade da fisionomia individual, transformando o rosto em uma representação agnóstica de fisionomia. Então, dada uma face neutral de origem, uma com a AU objetivo, e uma deformada como modelo padrão, procurou-se a transferência de deformação para uma outra face análoga que expresse a mesma AU. Para tal finalidade, utilizou-se o algoritmo Registro não rígido (NNR). Adicionalmente, considerou-se uma suavização das malhas, que foi realizada por meio do algoritmo de Laplace-Beltrami. Os autores replicaram o método de [41], aplicando o *framework* de normalização na fase de pré-processamento, e alcançando melhorias no resultado de detecção para a maioria dos casos em até um 1.9%, apesar de algumas diferenças na implementação e nos resultados.

3.2 Bases de dados

A seguir, serão brevemente descritas as duas bases de dados utilizadas nesta investigação. As bases referidas contêm imagens 3D, e foram escolhidas pois se mostrou interessante o fato das imagens contidas nas mesmas serem capturadas em dois ambientes distintos: forçado, para a base Bosphorus [5], e espontâneo, para a base BP4D [6]. No presente trabalho, denomina-se

ambiente forçado quando a expressão capturada na imagem, obteve-se de um indivíduo instruído para interpretar aquela expressão. Enquanto ambiente espontâneo será quando a expressão capturou-se de uma maneira mais "natural" , quer dizer, a expressão provocou-se de alguma forma, por exemplo, contando uma piada ao sujeito.

3.2.1 Base de dados - Bosphorus

A base de dados para análise facial em 3D "Bosphorus"[5], representa uma base com múltiplas expressões e poses. Os dados 3D foram capturados com o dispositivo digitalizador, o Mega Capturos II 3D [43], esse aparelho aplica luz–estruturada para capturar imagens em três dimensões. As imagens foram obtidas a partir de um grupo de 105 sujeitos, com diferentes características (60 homens e 45 mulheres, homens com barba e/ou bigode, pessoas entre 25 e 35 anos de idade, de e diferentes grupos étnicos), que apresentaram várias poses (13 em *yaw*, *pitch* e rotações cruzadas), expressões forçadas (seis expressões faciais básicas e 28 AUs) e condições de oclusão (mãos, cabelo, barba e óculos). Apesar das expressões serem capturadas de maneira forçada, para que estas pareçam mais naturais, foram utilizados atores e atrizes de teatros profissionais, óperas e conservatórios.

No estado da arte, esta é a primeira base de dados pública do tipo. Existem outras bases de dados 3D [44, 45, 46, 47, 48, 49, 50, 51, 6], entretanto, são orientadas ao reconhecimento facial e ocasionalmente análise de expressão 3D. Além disso, observou-se que essas bases contêm limites nas categorias de expressões, poucas poses da cabeça ou não apresentam oclusão. Em contrapartida, revisando na literatura verificou-se que existem vários trabalhos que utilizam a base de dados 3D Bosphorus para o reconhecimento de expressões faciais [29, 41, 7, 36].

Originalmente a base Bosphorus é composta por: imagens 2D, pontos de nuvens 3D e arquivos com anotação dos pontos fiduciais do rosto tanto em 2D quanto em 3D. Assim, para o desenvolvimento da presente pesquisa, se mostraram interessantes as imagens 3D da base que apresentam tanto expressões faciais básicas, como também subconjuntos de AUs. Deve-se destacar que essas imagens foram anotadas por especialistas em FACS. Consequentemente, do total de 4652 imagens de faces na base, selecionaram-se apenas 2900 para a detecção de AUs e posterior análise de EFCs, pois estas continham anotações de AUs e de pontos fiduciais. Exemplos de imagens de sujeitos da base Bosphorus podem ser vistos na Figura 3.1.

3.2.2 Base de dados BP4D-Spontaneus

A face humana é um objeto deformável que consiste em três dimensões, por isso é importante uma representação em 3D, assim, é apresentada a base de dados Binghamton-Pittsburgh 3D Dynamic Spontaneous Facial Expression Database (BP4D) [6]. A base foi construída a partir de vídeos 3D, capturados em um grupo de 41 jovens adultos (23 mulheres e 18 homens), a partir de oito tarefas (entrevistas, vídeos clipe, provas de susto, improvisação, provas de ameaça, provas de frio, provas de insulto e provas de odor), que não foram previamente anunciadas aos participantes, obtendo assim, as expressões faciais espontâneas.

No estado da arte é a primeira base de dados 3D pública com expressões faciais espontâneas. Há também o *dataset* 3D VT-KFER [52], capturado com o sensor Kinect 1.0, que considera expressões espontâneas, porém, está orientado ao reconhecimento de expressões faciais básicas + neutral, e não apresenta anotação de AUs. Existem outros trabalhos recentes para coletar, anotar e analisar expressões faciais espontâneas [53, 54, 55], porém, limitados a imagens 2D ou imagens térmicas.

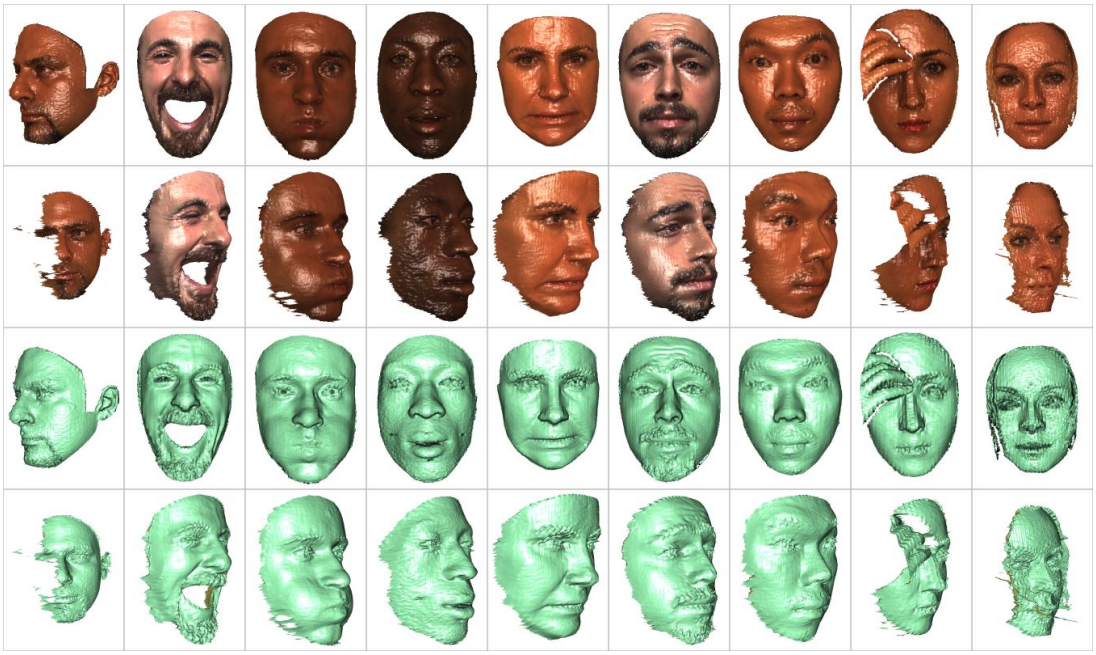


Figura 3.1: Imagens 3D de exemplo da base Bosphorus [5]

Composta por quadros, a base está direcionada às ações faciais seguindo o sistema FACS. Foram anotadas manualmente na base 27 AUs por dois especialistas certificados em FACS, para cada participante, focando-se nos 20 segundos (aproximadamente 500 quadros) na seção que tinha maior intensidade na expressão facial.

A base contém 328 arquivos de meta-data com a informação das AUs, considerando 146.847 quadros. Cada arquivo meta-data presente na base, essencialmente consiste em matrizes numéricas nas quais cada coluna corresponde a uma AU simples e cada linha a um quadro do vídeo. O valor de cada célula pode ser: "1", para existência, "0", para ausência ou "9", caso seja perdida a anotação ou a AU não tenha sido considerada. Na Figura 3.2 pode-se observar vários exemplos de sujeitos da base BP4D em representações de imagens 3D, 2D e anotação de AUs.

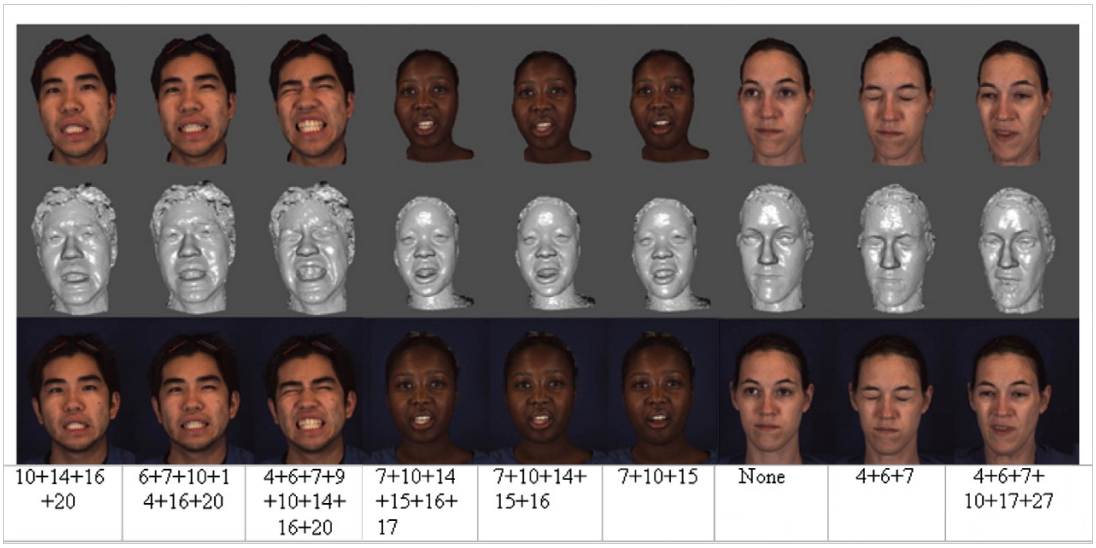


Figura 3.2: Imagens de exemplo da base BP4D-Spontaneous, com suas respetivas AUs [6]

3.3 Método

Após a revisão da literatura em 3.1, por fim escolhe-se como sub-método para detecção de AUs a aplicação de Padrões Binários Locais em mapas de profundidade (LDBP). Isto, pela simplicidade na sua implementação e porque os resultados dos autores originais em [7] são considerados muito bons em termos de Característica de Operação do Receptor (ROC) [56]. O método está representado na Figura 3.3, onde observa-se que cada imagem 3D primeiramente passa por um alinhamento, para depois ser projetada em mapa de profundidade, no qual aplica-se LBP, e extrai-se o vetor de características por meio de histogramas orientados a gradientes (HOG), posteriormente SVM é utilizado como classificador, e assim as AUs são detectadas. Para validar este sub-método foram escolhidas duas bases de dados com imagens 3D: Bosphorus e BP4D, já que contêm imagens capturadas em ambientes forçados e espontâneos, respetivamente. Nota-se que a base Bosphorus foi utilizada nos experimentos dos autores originais dos detectores de AUs escolhidos.

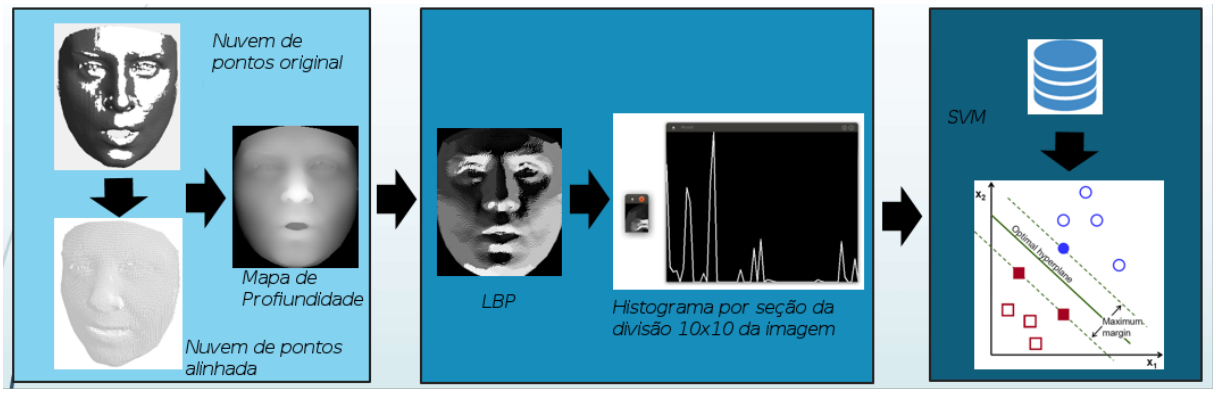


Figura 3.3: Esquema de funcionamento dos detectores de AUs. Neste diagrama pode-se observar que: dada a imagem 3D, esta é alinhada e projetada em um mapa de profundidade, ao qual aplica-se o descritor de textura LBP, para posteriormente obter o vetor 1D,para cada imagem, por meio de HOG, e treinar SVM como classificador

Inicialmente, examinou-se os dados providos nas duas bases de dados, determinando que os formatos delas se diferem - malhas são dispostas para a BP4D, enquanto que para a Bosphorus, existem apenas nuvens de pontos. Então, a primeira fase do processamento envolve triangulação das nuvens de pontos para criar malhas no caso da base Bosphorus, para depois disso, aplicar suavização Laplaciana em ambas bases. Como a base Bosphurus apresenta mais ruido, requer mais suavização. Este processo é realizado para garantir que ambas as bases estejam na mesma forma, a fim de comparação. Para a triangulação e a suavização utilizou-se a biblioteca *The Visualisation Toolkit* (VTK), para *Python*.

Como próxima etapa, as imagens 3D de ambas as bases foram alinhadas no plano X-Y por meio de matrizes afins. Estas matrizes permitem movimentos de rotação, traslação e escala dos pontos de nuvens, em relação aos pontos fiduciais selecionados como padrão de alinhamento em cada uma das bases. Para a base Bosphorus utilizou-se os pontos fiduciais do sujeito "bs000_LFAU_10_0", enquanto que para a base BP4D, foram considerados os pontos fiduciais do sujeito "F001/T1/2449", pois em ambos os casos, os sujeitos estão frontalmente alinhados. Para o alinhamento da base Bosphorus, foram selecionados dez pontos fiduciais marcados originalmente na base: quatro dos olhos, cinco do nariz e um do queixo, como pode ser visualizado na Figura 3.4. Para a base BP4D utilizaram-se 11 pontos, também marcados na base: quatro dos olhos, seis do nariz e um do queixo, representados na Figura 3.5.

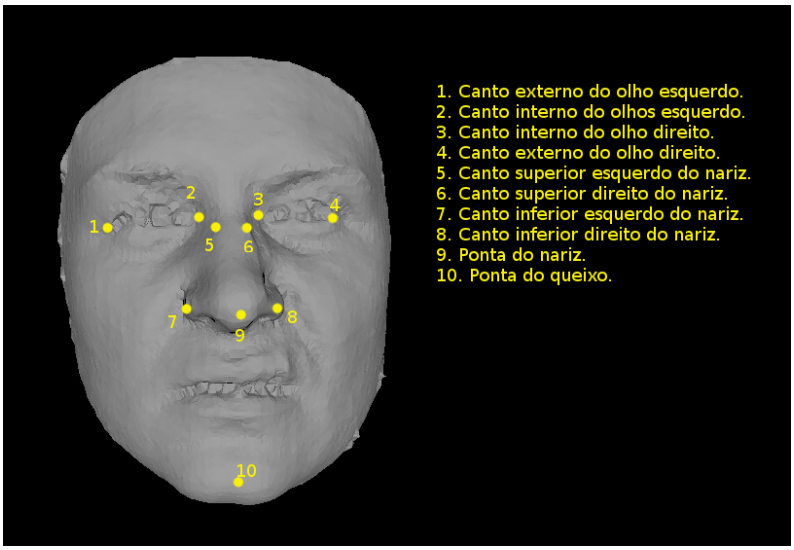


Figura 3.4: Pontos fiduciais selecionados para o alinhamento, do sujeito bs000 expressando a AU 10 da base Bosphorus

Depois disso, com as novas nuvens de pontos alinhadas, realiza-se a projeção das mesmas em uma representação 2D: os mapas de profundidade. Estes mapas facilitam o processamento por terem as características das imagens 2D; permitindo a aplicação de métodos provados nesse tipo de imagens, e reduzindo o custo computacional [36]. Cabe ressaltar que este tipo de representação não perde informação do plano 3D; esta se representa pela intensidade dos pixels na imagem. Deste modo, se procede a projeção das nuvens de pontos 3D a pixels no espaço 2D, convertendo os valores X-Y do vértice 3D para seus correspondentes X-Y na dimensão 2D do pixel, e considera-se o valor de z como a intensidade do pixel. Se houver vários pontos em um mesmo pixel, o valor da intensidade será a média dos valores da coordenada z dos mesmos. Os mapas resultantes têm uma dimensão de 120x120 pixels, para ambas bases.

Posteriormente, imagens projetadas devem ser redimensionadas para 300x300 pixels. Isso se dá por meio da aplicação da interpolação bicubica. No caso da base Bosphorus, esse procedimento foi realizado sem problemas. Em contrapartida, imagens da base BP4D tiveram que passar por um processo de segmentação para a obtenção do rosto isoladamente, pois as imagens originalmente contêm informação da cabeça inteira, cabelo, pescoço e inclusive parte do tronco, conforme se observa na Figura 3.6.

O processo de segmentação é baseado em [57], e está representado no diagrama da Figura 3.7. Consequentemente, aplica-se um filtro de média com janela 5x5, para depois utilizar a interpolação bicubica, como no caso das imagens da base Bosphorus. Em seguida, imagens passam por *Clustering* com *K-Means* com o valor de K igual a 3. Feito isso, emprega-se o método de segmentação do limiar binário para pixels com intensidades maiores a 148.

Finalizada a projeção e todos os processos conseguintes, obteve-se um total de 2900 mapas frontalmente alinhados para a base Bosphorus, das 2902 imagens com AUs anotadas, enquanto que para a BP4D foram selecionados aleatoriamente 10000 mapas frontalmente alinhados de um total de 146847 que possuem anotação de AUs. Exemplos de mapas corretamente projetados nota-se na Figura 3.8, a da esquerda para a base Bosphorus, e a da direita para a base BP4D.

Alguns mapas foram desconsiderados, em ambas as bases. No caso da Bosphorus, duas imagens não foram consideradas no restante do processo, pois uma foi deformada no alinhamento como se evidencia a esquerda da Figura 3.9, e a outra possuía tanto ruído (pontos afastados que

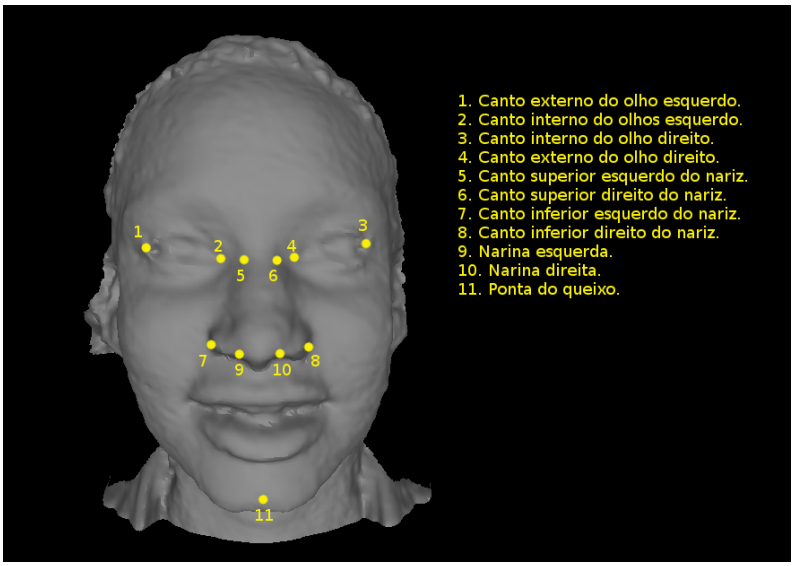


Figura 3.5: Pontos fiduciais selecionados para o alinhamento, do sujeito F001, tarefa 1, quadro 2449 da base BP4D

Figura 3.6: Mapa de profundidade projetado do sujeito F000 da base BP4D realizando a tarefa T1 para o quadro 2440

não faziam parte da face) que o pré-processamento não conseguiu eliminar. Para a base BP4D foram omitidas as imagens que tinham: grandes buracos, deformações, erros na segmentação, ou algum tipo oclusão. Na Figura 3.9 a direita, visualiza-se um exemplo de mapa da base BP4D com oclusão e falhas na segmentação.

Com a finalidade de melhorar a qualidade dos mapas de profundidade e reduzir fundo, realiza-se um pré-processamento. Desse modo, aplica-se nos mapas de ambas as bases o filtro de média com diâmetro de vizinhança igual a cinco. Após isso, o fundo é reduzido mediante a utilização de uma caixa delimitadora, assim, mapas são redimensionados para 220x300 pixels. Um exemplo dos mapas para cada uma das base, depois do pré-processamento, pode ser visualizado na Figura 3.10, na esquerda para a Bosphorus e na direita para a BP4D.

Em seguida, características são extraídas por meio da aplicação do descritor de textura LBP nos mapas de profundidade. Este procedimento se denomina LDBP [7]. Esse descritor consegue codificar para cada pixel a forma local ao redor da vizinhança, e raio com valores iguais a oito. Exemplos da aplicação do descritor LDBP em mapas de profundidade da base Bosphorus (esquerda) e da base BP4D (direita) são apresentados na Figura 3.11.

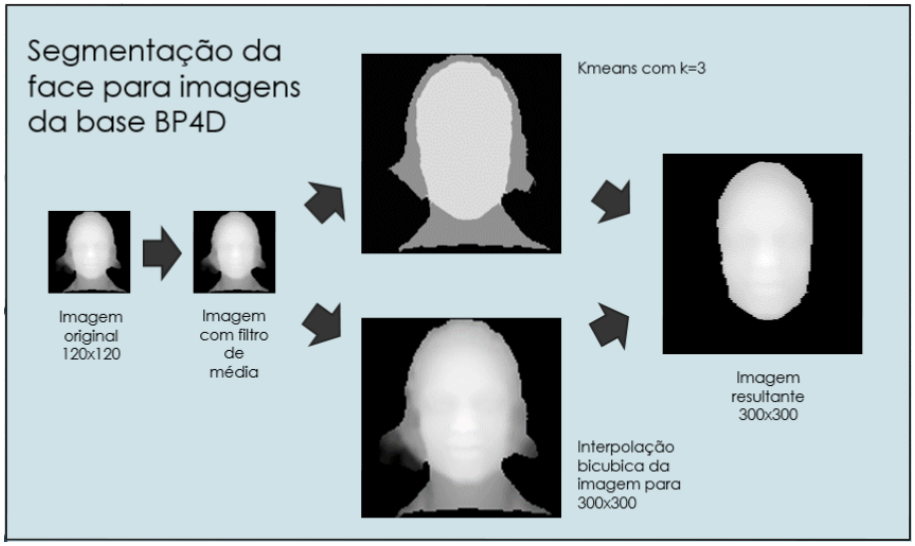


Figura 3.7: Método de segmentação aplicado em imagens da base BP4D para a separação da face do restante da imagem. Neste, observa-se como dado o mapa de profundidade, deve-se aplicar o filtro de média e a interpolação bicúbica, para posteriormente utilizar *K-means*, comparando amba as imagens resultantes, e assim segmentar o rosto do restante da cabeça

Depois da aplicação do descritor de textura LDBP, a imagem resultante divide-se em quadrados 10x10 (todos do mesmo tamanho), e definem-se os vetores de características, por meio de HOG. Dessa forma, histogramas normalizados de 60 *bins* são calculados dentro de cada uma das regiões da imagem, para que em seguida, sejam concatenados um atrás do outro, como se ilustra na Figura 3.12. Resultando em vetores 1D com uma longitude total de 6000 características.

Finalmente, os vetores de características são treinados em maquinas SVM binárias para a detecção de AUs. No caso, os classificadores empregaram um núcleo de interseção de histogramas, e utilizam validação cruzada em *10-fold*, como otimização de parâmetros. A validação é realizada dividindo os *datasets* em 10 partes iguais, selecionando nove para treinamento e uma para teste; maior detalhe desse procedimento será explicado em 3.4. Sendo assim, classificadores SVM são treinados para cada base com o fim da detecção das diferentes AUs. Os experimentos e resultados serão discutidos nas seções 3.4 e 3.5 respetivamente.

Ressalta-se a diferença do *baseline* e método implementado, na omissão do algoritmo de *GentleBoost* (GB) para seleção de características, pois particularmente para os experimentos executados nos detectores desenvolvidos, os resultados foram mais altos quando se desconsiderou GB, isto pode ser explicado pelo fato de que no método original utilizaram uma versão própria e modificada do GB estândar. Ainda, destaca-se que GB originalmente no *baseline* foi aplicado apenas para otimização do tempo de execução.

3.4 Experimentos nas bases de dados

Uma vez implementado o método descrito na seção 3.3, treinamentos e testes são executados com as imagens das bases Bosphorus e BP4D. Em ambos os casos, para o treinamento, aplica-se como classificador SVM, utilizando validação cruzada em *10-fold*. Consequentemente, os *datasets*, previamente selecionados (as 2900 imagens da base Bosphorus e as 10000 da base BP4D), são divididos aleatoriamente em dez partes iguais, sendo utilizadas nove delas para gerar conjuntos de treinamento balanceados, conforme realizado no *baseline*, para cada uma

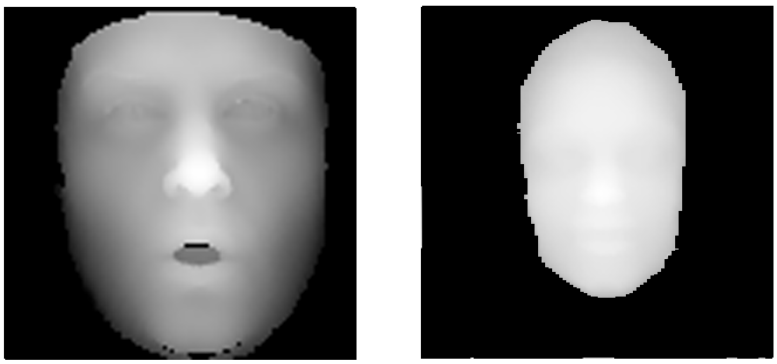
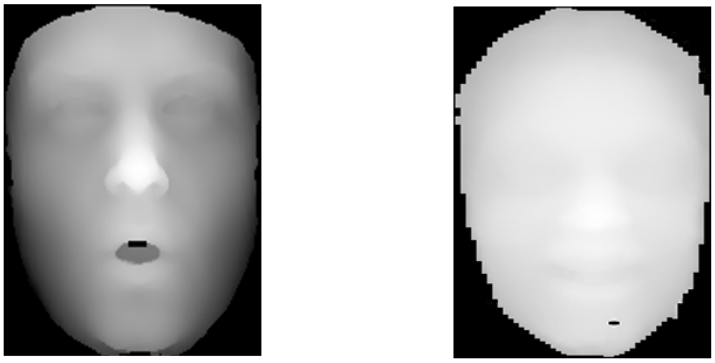


Figura 3.8: Mapas de profundidade frontalmente alinhados. (a) mapa obtido do sujeito bs000 da base Bosphorus expressando as AUs 22 e 25. (b) mapa projetado da malha 3D do sujeito F001 na tarefa T1 para o quadro 2440

Figura 3.9: Mapas de profundidade falhos. (a) mapa deformado obtido do sujeito bs017 da base Bosphorus com a expressão "Feliz". (b) mapa projetado da malha 3D do sujeito M015 na tarefa T3 para o quadro 168, com oclusão, falhas na segmentação e presença de buracos

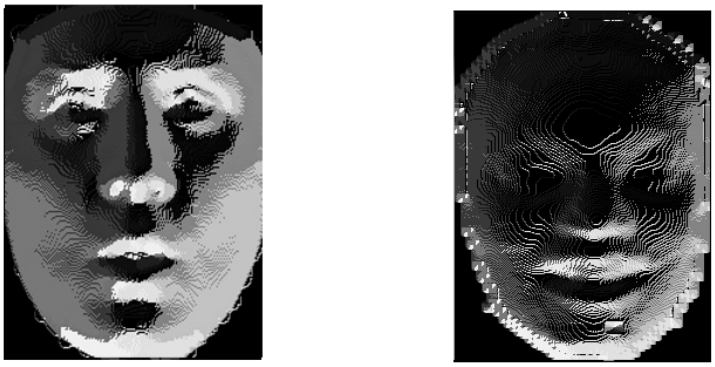
das AUs que serão detectadas. Cada conjunto é composto por todos os exemplos positivos que demostraram a ocorrência de cada AU, mais um número igual de exemplos negativos. Ressalta-se que somente nos conjuntos de treinamento do *dataset* da base Bosphorus, apenas são consideradas como exemplos positivos aquelas imagens com intensidades de AUs no intervalo de C até E. Não se realiza essa seleção no *dataset* da base BP4D, pois não existem anotações de intensidades para todas as AUs na base. Em seguida, com a parte restante da divisão dos *datasets*, constrói-se o conjunto de testes.

Seguindo o *baseline*, realizam-se experimentos para detectar as AUs: 1, 2, 4, 9, 10, 12, 12L, 12R, 14, 15, 16, 17, 18, 20, 22, 23, 24, 25, 26, 27, 28, 34, 43, e 44. Consequentemente, notou-se que para executar os experimentos do *baseline* na base BP4D, precisam ser anotadas as AUs 12L, 12R, 25, 26, 43, e 44, pois essa anotação não existe originalmente na base. Porém, realizou-se somente a anotação manual das AUs 25 e 26, pois como estão presentes em 11 das EFCs, devem ser detectadas para o reconhecimento dessas expressões. Em contrapartida, foram desconsideradas as AUs: 12L, 12R, 43, e 44, na base BP4D, principalmente porque a anotação manual ocupa bastante tempo (considerando a anotação de 10000 imagens) e não seria executada



(a) (b)

Figura 3.10: Mapas resultantes do pré-processamento para melhorar qualidade e diminuir fundo. (a) mapa do sujeito bs000 da base Bosphorus expressando as AUs 22 e 25. (b) mapa do sujeito F001 na tarefa T1 para o quadro 2470



(a) (b)

Figura 3.11: Mapas de profundidade com aplicação do descritor de textura LDBP. (a) sujeito bs000 da base Bosphorus com as ações faciais 22 e 25. (b) sujeito F001 na tarefa T1 para o quadro 2470

por especialistas em FACS. Além disso, essas AUs não aparecem nas EFCs, e sua detecção não será aplicada posteriormente.

Assim, pensando que a finalidade da detecção das AUs em imagens 3D reside em reconhecer EFCs nesse mesmo tipo de imagem, foi necessário estudar todas as AUs componentes dessa classificação de expressões. Consequentemente, observou-se que na configuração padrão de EFCs estão presentes as AUs: 1, 2, 4, 5, 6, 10, 12, 15, 17, 20, 25, e 26. O diagrama da Figura 4.1 apresenta as AUs comuns entre cada base e entre as consideradas no padrão de configuração de AUs das EFCs. Nesse diagrama, constata-se principalmente que: entre as duas bases e o padrão de configuração de AUs das EFCs, existem apenas oito AUs anotadas em comum (AUs: 1, 2, 4, 10, 12, 15, 17 e 20), entre o padrão e a base Bosphorus, há duas AUs (AUs: 25 e 26), e entre a base BP4D e o padrão, encontram-se apenas duas (AUs: 5 e 6). Percebeu-se que para encontrar EFCs na base Bosphorus é necessário detectar as AUs 5 e 6.

Por fim, executam-se os experimentos de detecção de AUs: como no *baseline*, e para o reconhecimento das EFCs, nas duas bases. Adverte-se que cada AU é tratada separadamente, isso quer dizer que, para os experimentos seguindo o *baseline*, foram desenvolvidos 24 detectores para a base Bosphorus e 20 para a base BP4D, enquanto que para os experimentos orientados

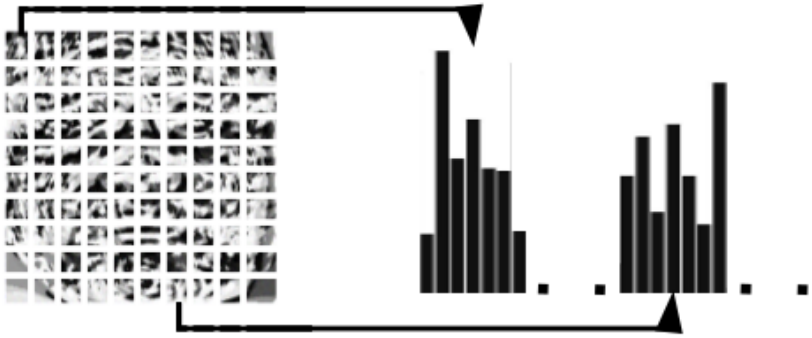


Figura 3.12: Construção dos vetores de características por meio da concatenação dos histogramas de cada sub-seção do mapa de profundidade com aplicação de LDBP. Figura obtida de [7]

ao reconhecimento de EFCs, reutilizaram-se 10 dos detectores do experimento anterior e desenvolveram-se apenas 2 detectores a mais (para as AUs 5 e 6), em ambas as bases. Resultados de acurácia serão debatidos na próxima secção 3.5.

3.5 Resultados na detecção de AUs

Finalizados os experimentos descritos na seção 3.4, obteve-se resultados interessantes a serem abordados. Primeiramente, detalha-se os resultados obtidos no experimento baseado no *baseline*: para os testes das 24 AUs procuradas na base Bosphorus, o valor da área sob a curva (AuC) em média foi de 79,4%, por outra parte, considerando os testes das 20 AUs na base BP4D, a o valor da AuC em média resultante atingiu os 86,2%. Em seguida, são apresentados os resultados do segundo experimento, o qual é orientado ao reconhecimento de EFCs: para os testes das 12 AUs detectas na base Bosphorus, o valor da AuC em média alcançou os 76,4%, enquanto que para os testes das mesmas 12 AUs identificadas na base BP4D, resultou em um valor da AuC de 86,8%. Destaca-se que todos esses resultados são próximos entre si (fato que lhes proporciona consistência) e que são considerados bons em termos de Característica de Operação do Receptor (ROC) [56].

Detalhando, exibem-se na Tabela 3.1 as porcentagens das AuCs para cada uma das 24 AUs da base Bosphorus e das 20 da base BP4D. Dessa forma, evidencia-se que nos resultados da base Bosphorus, seis AUs não superaram os 75,0% de reconhecimento (denotadas com "*"), 12 pertencem ao intervalo de 75%-90% (denotadas com "**") e seis são superiores a 90% (denotadas com "***"). Por outra parte, observa-se que, para base BP4D, apenas uma AU não superou os 75% de reconhecimento, 13 AUs pertencem ao intervalo 75%-90%, e seis superaram os 90% de reconhecimento. Nota-se que, foram assinaladas com "X" aquelas células onde a AU não foi considerada na detecção.

Enquanto na Tabela 3.2, apresentam-se os resultados individuais das 12 AUs consideradas na configuração padrão das EFCs, para as bases Bosphorus e BP4D. Realizando a mesma análise que na Tabela 3.1, afirma-se que, nos resultados da base Bosphorus, quatro AUs não superaram os 75,0% de reconhecimento (denotada com *), sete pertencem ao intervalo de 75%-90% (denotadas com **) e uma superou os 90% de acerto (denotada com ***). Considerando a base BP4D, todas superam os 75,0%, nove estão no intervalo 75%-90% e três conseguiram alcançar valores de reconhecimento superiores a 90%.

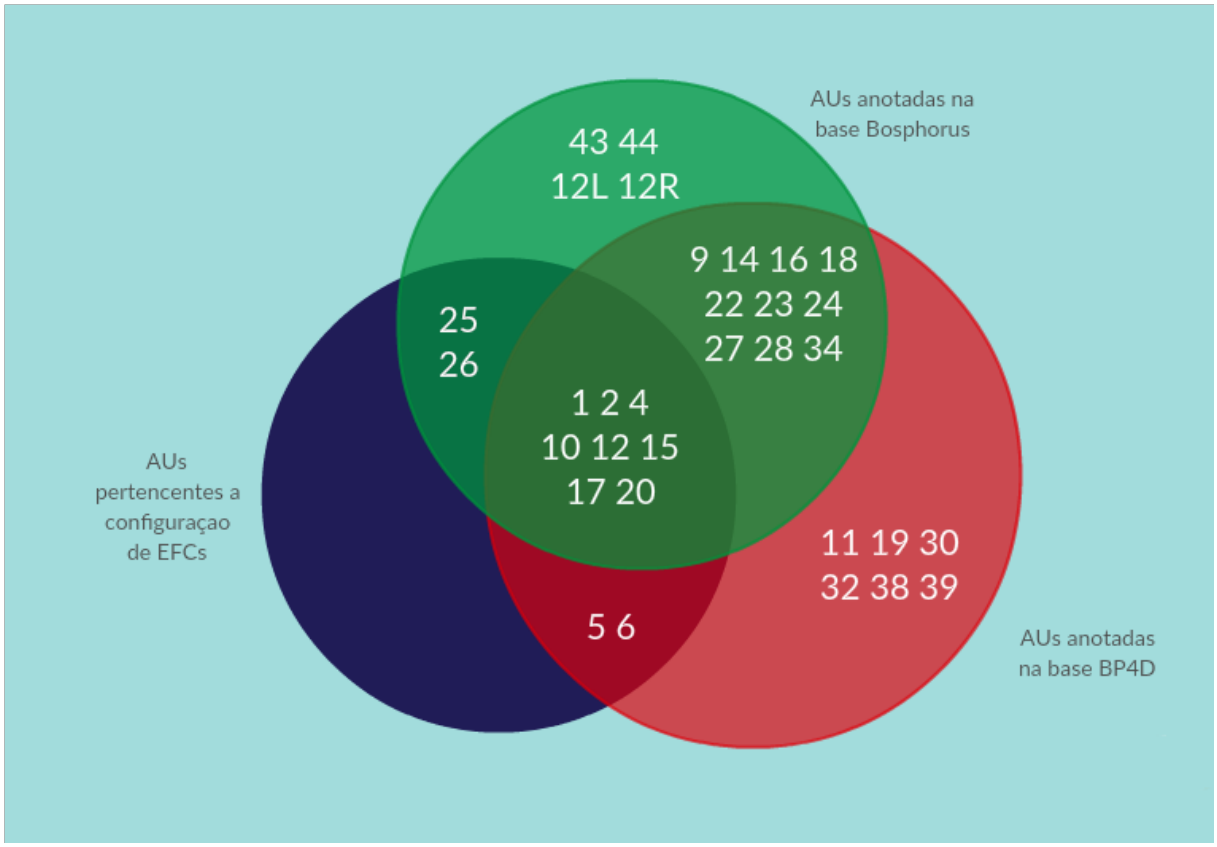


Figura 3.13: Diagrama de comparação de AUs pertencentes ao padrão de AUs de EFCs (azul) e as anotadas nas bases Bosporus (verde) e BP4D (vermelho)

Complementando as observações dos experimentos, torna-se interessante analisar o comportamento das duas bases selecionadas. Pode-se afirmar que a base BP4D obteve um melhor desempenho, tanto no experimento baseado em [7], quanto no experimento orientado na busca de expressões compostas. Isso pode-se constatar no gráfico da Figura 3.14 (O gráfico apenas considera as AUs em comum nas duas bases, representando os valores AuC de cada base). A base Bosphorus teve um melhor comportamento apenas em detectar as AUs 9, e 22. Na maioria dos casos, a variação foi pequena, porém, para o caso da AU 22, evidencia-se a maior diferença, um 30.7%, isso pode ser explicado pelo fato de existirem poucos casos de testes positivos com essa AU para a base BP4D (dois de 1000 na BP4D contra 12 de 2900 na Bosphorus), ou seja, não acertar em um caso da base BP4D compromete muito mais o valor AuC do que não acertar em um caso na Bosphorus. Em contrapartida, a base BP4D teve uma melhor detecção nas outras 18 AUs.

Sobre os resultados do experimento efetuado para detectar AUs ocorrentes na definição de EFCs, reconhece-se que ambas bases tiveram valores AuC muito próximos, como é apresentado no gráfico da Figura 3.15. Da mesma forma que no primeiro experimento, o valor AuC da detecção das AUs entre as bases não se afastou muito, o que já era esperado, dado que as únicas variações são: a inclusão de duas AUs que não foram consideradas no experimento anterior, e que foram avaliados menos casos de estudo. Nesse segundo experimento, a base BP4D teve uma detecção melhor em todos os casos das AUs consideradas. Além disso, pode-se verificar que as duas AUs introduzidas tiveram o mesmo comportamento das já analisadas anteriormente. Por último, revelou-se que a quantidade de AUs para detectar não gera mudanças drásticas nos valores AuC em média. Finalmente, estas AUs vão permitir realizar o reconhecimento das diferentes

Tabela 3.1: Resultados individuais de valores AuC em porcentagem (%) na detecção de AUs nas bases Bosphorus e BP4D, segundo os experimentos do *baseline*. Denotou-se com "*"quando valores são inferiores a 75%, "**" se caem entre 75%-90% e "***" caso superem os 90%. Assinalaram-se com "X" aquelas células onde a AU não foi considerada na detecção

AU	**% Bosphorus**	**% BP4D**
1	75,0 **	84,7 **
2	67,2 *	83,9 **
4	76,9 **	85,1 **
43	64,1 *	X
44	71,2 *	X
9	95,0 ***	88,7 **
10	76,3 **	87,8 **
12	82,2 **	92,9 ***
12L	94,8 ***	X
12R	92,4 ***	X
14	76,7 **	81,7 **
15	68,4 *	84,1 **
16	79,5 **	88,5 **
17	76,8 **	79,0 **
18	83,0 **	91,1 ***
20	75,8 **	90,9 ***
22	87,5 **	56,8 *
23	56,2 *	79,8 **
24	75,9 **	90,0 **
25	92,4 ***	95,9 ***
26	72,3 *	84,1 **
27	92,6 ***	95,5 ***
28	81,1 **	86,5 **
34	92,7 ***	97,7 ***
σ	**79,4**	**86,2**

EFCs, já que fazem parte da configuração padrão que define esta nova categoria de expressões faciais.

Adicionalmente, são apresentados na tabela 3.3, medidas de avaliação de matriz de confusão, para a detecção de todas as AUs consideradas para a base Bosphorus. Mostrando valores de acurácia, sensibilidade, especifidade, valor preditivo positivo, e valor preditivo negativo. Nota-se o valor ideal é de 100.0% para todas as medidas consideradas, menos para o valor preditivo negativo, que é 0.0%. Dessa maneira, observa-se que em geral o método obteve bons valores de acurácia, apresentado a pior para a AU 23 e a melhor para a AU 34, os outros valores estão próximos aos 80% como esperado. Além disso, o método apresentou valores de sensibilidade variáveis, que vão desde 41.2% (AU 43) até 100% (AUs 9, 12L, e 12R), demonstrando que alcançou-se reconhecer a maior parte dos casos positivos nas diferentes AUs detectadas. Também, por meio da especifidade, percebe-se que uma boa performance para detectar casos negativos, sendo o melhor de 100% (AU 25) e o pior 68.0% (AU 23). Similarmente, o valor preditivo positivo logrado expõe que houve vários casos com baixos valores devido a uma grande influencia nos casos que foram falsamente reconhecidos, sendo o pior 4.3% (AU 23) e o

Tabela 3.2: Resultados individuais de AuC em porcentagem (%) na detecção de AUs nas bases Bosphorus e BP4D, considerando as AUs da configuração padrão das EFCs. Denotou-se com "*" quando valores são inferiores a 75%, "**" se caem entre 75%-90% e "***" caso superem os 90%

AU	**%Bosphorus**	**%BP4D**
1	75,0 *	84,7 **
2	67,2 *	83,9 **
4	76,9 **	85,1 **
5	73,7 *	85,0 **
6	79,6 **	88,7 **
10	76,3 **	87,8 **
12	82,2 **	92,9 ***
15	68,4 *	84,1 **
17	76,8 **	79,0 **
20	75,8 **	90,9 ***
25	92,4 ***	95,9 ***
26	72,3 **	84,1 **
σ	**76,4**	**86,3**

melhor 100.0% (AU 25). Finalmente, o valor preditivo negativo apresenta-se como complemento para a especifidade do método. Por meio desses valores pode-se afirmar uma grande dificuldade em detectar a AU 23, e que resulta mais fácil detectar as AUs 34 e 25.

Da mesma forma, na tabela 3.4 apresentam-se essas mesmas medidas de avaliação de matrizes de confusão, sob as mesmas considerações, mas para a base B4PD. Assim, verifica-se que o método alcançou bons valores de acurácia, sendo o melhor 95.3% (AU 34) e o pior 63.6%(AU 22). Sobre a sensibilidade, afirma-se que teve bom desempenho, sendo o melhor 100% (AUs 18, 20, e 34) e o pior 50.0% (AU 22). A especifidade reflete que os casos negativos foram corretamente detectados na maioria das AUs, porém apresentando o valor menor 63.6% (AU 22) e o maior 95.3% (AU 34). Conjuntamente, o valor preditivo positivo revelou baixos valores para muitas das AUs, manifestando índices altos de casos que foram falsamente reconhecidos, resultando o mais desafiante em 0,3% (AU 22), mas em contrapartida a porcentagem mais alta em 95.0% (AU 25). Aqui o valor preditivo negativo também é complemento da especifidade, colaborando a comprovar os seus valores. Em geral, a base teve a AU 22 como a mais desafiante, mas excelente detecção para a AU 25 e 34.

Finalmente, a maneira de sintetizar o trabalho de detecção desenvolvido, ressaltam-se os seguintes pontos importantes observados a partir dos resultados dos experimentos realizados para a detecção de AUs em imagens 3D nas bases Bosphorus e BP4D:

- Apesar dos resultados no primeiro experimento diferirem dos valores no *baseline*, evidencia-se que são próximos. Estas variações podem ser explicadas pela seleção aleatória dos *datasets* (conforme realizado no *baseline*), mas também pelo fato de existirem algumas diferenças na implementação de ambos os métodos.

- A base BP4D teve resultados levemente superiores, acredita-se que isso pode ser justificado pelo fato de ter selecionado um *dataset* maior (2900 imagens para a base Bosphorus e 10000 para a base BP4D), porém, a diferença não foi tão representativa na maioria dos casos.

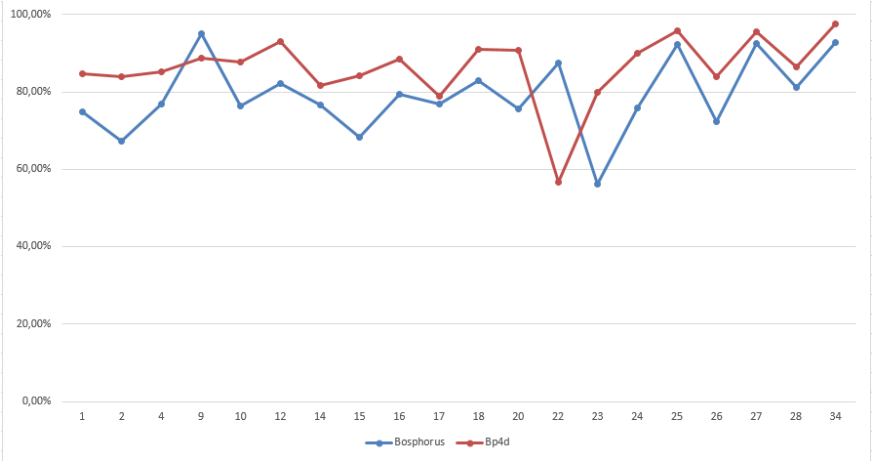


Figura 3.14: Gráfico com os valores AuC individuais das AUs detectadas nas bases Bosphorus (linha azul) e BP4D (linha vermelha) seguindo o *baseline* [7]. No eixo vertical, são representados os valores AuC em porcentagem, enquanto no eixo horizontal, enumeram-se as AUs

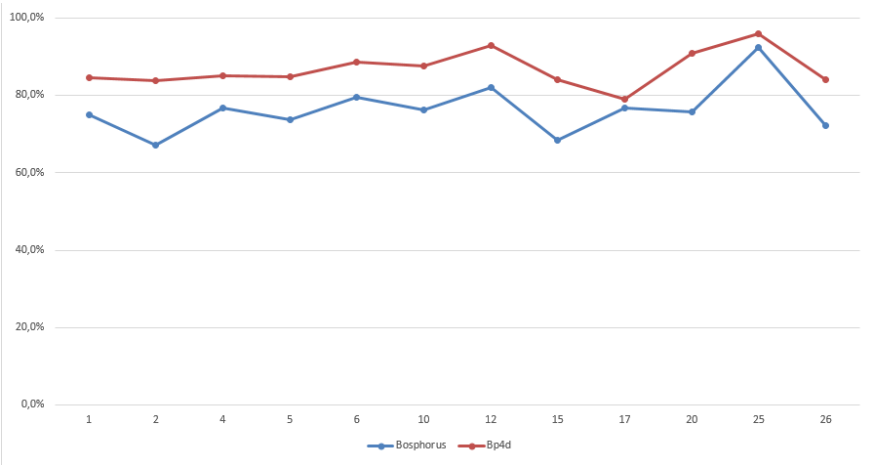


Figura 3.15: Gráfico com os valores AuC individuais das AUs detectadas nas bases Bosphorus (linha azul) e BP4D (linha vermelha) para o posterior reconhecimento de EFCs. No eixo vertical, são representados os valores AuC em porcentagem, enquanto no eixo horizontal, enumeram-se as AUs

- Contudo, afirma-se que a base BP4D teve um desempenho melhor do que a base Bosphorus na detecção da maioria das AUs, já que apesar de não apresentar muita diferença entre os valores das AuCs e das próprias acurácias, outros valores como a sensibilidade e o valor preditivo positivo demonstram melhores porcentagens para a BP4D.

- É interessante notar que as duas bases (Bosphorus e BP4D) tiveram um comportamento similar nos experimentos realizados, em geral os resultados individuais do valor da AuC se aproximaram aos 80%. Representando assim, que o método é bom para os testes elaborados.

- Acredita-se que as diferenças do método desenvolvido com o *baseline* [7] se devem a vários motivos, dente eles pode-se citar: a dificuldade de replicar a distribuição de exemplos na validação cruzada, as potenciais alterações na implementação própria, e até na escolha de bibliotecas aplicadas.

Tabela 3.3: Medidas básicas de avaliação da matriz de confusão da detecção de AUs para base Bosphorus. Sendo representadas nas colunas as porcentagens de acurácia, sensibilidade, especifidade, valor preditivo, e valor preditivo negativo. E as linhas, as diferentes AUs consideradas

	Acurácia	**Sensibilidade**	**Especifidade**	**Valor preditivo positivo**	**Valor preditivo negativo**
1	84,1	63,2	87,3	42,9	12,7
2	83,4	45,7	88,6	35,6	11,4
4	86,2	64,7	89,1	44,0	10,9
43	81,7	41,2	87,1	29,8	12,9
44	71,7	50,4	91,9	85,5	8,1
5	80,3	64,9	82,6	35,3	17,4
6	83,8	74,2	84,9	37,1	15,1
9	90,7	100,0	90,0	42,6	10,0
10	78,6	73,3	79,2	28,9	20,8
12	89,7	70,4	94,1	73,1	5,9
12L	89,7	100,0	89,5	11,8	10,5
12R	89,7	100,0	84,8	14,0	15,2
14	85,2	69,6	83,9	27,1	16,1
15	78,6	57,1	79,7	12,5	20,3
16	78,3	81,0	78,1	22,4	21,9
17	80,3	72,4	81,2	30,0	18,8
18	87,2	77,8	88,2	40,4	11,8
20	79,7	71,4	80,1	15,4	19,9
22	91,4	83,3	91,7	30,3	8,3
23	67,2	44,4	68,0	4,3	32,0
24	83,4	66,7	85,2	31,6	14,8
25	94,5	84,8	100,0	100,0	0,0
26	80,3	60,0	84,6	44,8	15,4
27	94,8	90,0	95,2	58,1	4,8
28	84,1	77,8	84,3	13,7	15,7
34	96,2	88,9	96,4	44,4	3,6

- O fato de realizar a seleção dos *datasets* de treinamento e de teste de maneira aleatória provocou desequilibro para alguns casos, separando uma quantidade imagens insuficiente para treinamento, o que afetou na acurácia na detecção de AUs em imagens de teste.
- Quando existiam poucas ocorrências de alguma AU positiva nos testes, os resultados da acurácia foram muito afetados, caso não se detectasse alguma imagem com a presença da AU procurada. Por exemplo, para a AU 22, na base BP4D, apenas ocorreram dois casos positivos, o fato de não ter sido detectado um deles implicou em uma queda significativa na acurácia do detector.
- Finalmente, acredita-se que com os detectores desenvolvidos, é possível realizar uma nova aplicação: o método de reconhecimento de EFCs em imagens 3D. O qual será analisado no Capitulo 4.

Tabela 3.4: Medidas básicas de avaliação da matriz de confusão da detecção de AUs para base Bosphorus. Sendo representadas nas colunas as porcentagens de acurácia, sensibilidade, especifidade, valor preditivo, e valor preditivo negativo. E as linhas, as diferentes AUs consideradas

	Acurácia	**Sensibilidade**	**Especifidade**	**Valor preditivo positivo**	**Valor preditivo negativo**
1	84,3	85,4	84,0	59,1	16,0
2	84,6	82,9	84,9	47,1	15,1
4	84,8	85,6	84,6	56,3	15,4
5	78,6	91,7	78,3	9,4	21,7
6	88,7	89,4	88,0	88,2	12,0
9	85,7	92,1	85,3	29,6	14,7
10	87,6	87,1	88,4	92,2	11,6
12	92,6	90,9	95,0	96,2	5,0
14	81,9	79,0	84,3	80,9	15,7
15	83,7	84,8	83,4	54,7	16,6
16	89,7	87,2	89,8	25,8	10,2
17	79,0	78,8	79,1	66,0	20,9
18	82,2	100,0	82,1	2,2	17,9
20	82,1	100,0	81,7	10,9	18,3
22	63,6	50,0	63,6	0,3	36,4
23	79,9	79,7	80,0	47,8	20,0
24	88,5	92,1	87,9	55,4	12,1
25	95,9	96,7	95,1	95,0	4,9
26	87,0	80,2	87,9	47,8	12,1
27	91,0	100,0	90,9	7,2	9,1
28	87,3	85,7	87,3	12,7	12,7
34	95,3	100,0	95,3	2,1	4,7

4 Reconhecimento automático de expressões faciais compostas em imagens 3D

O estudo das expressões faciais humanas começou há mais de um século, quando Darwin apresentou o seu trabalho *"As expressões de emoções no homem e nos animais"* [58]. Nessa pesquisa, Darwin afirmava que não seria possível entender as expressões de emoção humana sem primeiro entender as expressões dos animais. Com o passar do tempo, o estudo de expressões faciais tem evoluído, vários autores seguiram esta linha, entre os mais reconhecidos estão Paul Ekman e Wallace Friesen, sendo a definição do sistema FACS uma de suas maiores colaborações. Além disso, definiram seis expressões básicas, que Ekman e Keltne em [10], demostraram que são universais.

Entretanto, para Du *et. al* em [3, 21], existem muitas expressões de emoções mais do que as seis básicas, por causa disso, definem 17 expressões faciais compostas. Que são expressões construídas pelas expressões básicas e que têm sentido, pois representam dois estados de emoção que podem ser expressados em um mesmo momento. Apesar disso, afirmaram que não são todas as combinações de expressões básicas resultam em uma expressão composta que possa ser gerada na face e que possua sentido em sua interpretação.

Sendo assim, foi realizado um levantamento na literatura, e notou-se que, ultimamente, tem aumentado o interesse em estudar EFCs. Existem trabalhos em reconhecimento de EFCs em imagens 2D [23] e em agentes virtuais 3D [59], além disso, anotação automática desse tipo de expressões em imagens 2D [34], e inclusive uma base de dados 2D que as contempla [60]. Em vista disso, propõe-se um método de reconhecimento de EFCs, com o diferencial de ser aplicado em imagens 3D de humanos reais, e também considerando os ambientes de captura forçado e espontâneo, com o intuito de comparar o seu comportamento.

4.1 Método

Para poder reconhecer EFCs em imagens 3D, foi desenvolvido um método automático, o mesmo consegue identificar esse tipo de expressões por meio da detecção das AUs presentes nas imagens, representando asssim uma nova aplicação de detectores de AUs, que no caso são considerados aqueles desenvolvidos em 3.3, e ponderando a configuração padrão de AUs definida pelos autores originais em [3, 21]. Nota-se que se teve a hipótese de que EFCs podem ser aproximadas (quando algumas das AUs foram detectadas), mas descartou-se essa ideia, já que para enlaçar as imagens consideradas como "verdadeiras"(as definidas pelas AUs anotadas por especialistas em FACS) com as detectadas existia muita confusão, considerando a grande quantidade de combinações possíveis de AUs para cada EFC nas diferentes imagens. Portanto, para este caso, optou-se por definir que cada expressão é reconhecida apenas quando são encontradas todas suas AUs componentes.

Dessa maneira, a seguir descreve-se o referido método para poder reconhecer EFCs em imagens 3D. O mesmo é apresentado na Figura 4.1. Assim, dada uma imagem 3D como entrada, esta passa por um pré-processamento, para em seguida detectar as AUs presentes para posteriormente compará-las com as AUs definidas na configuração padrão de EFCs, e dessa maneira, determinar a expressão composta correspondente, caso exista.

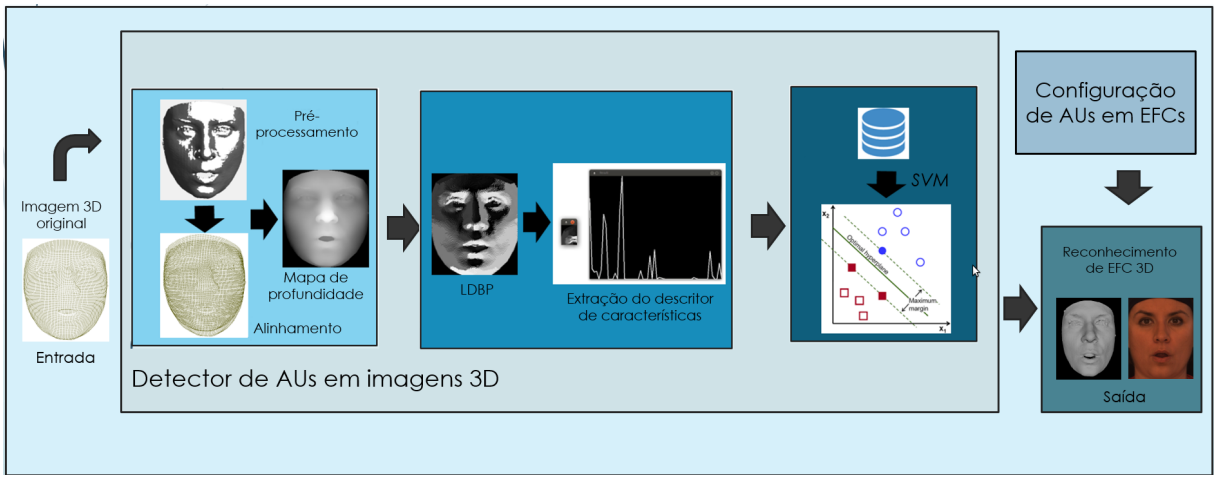


Figura 4.1: Esquema do método de reconhecimento de EFCs em imagens 3D. Neste, observa-se que: dada uma imagem 3D, está passa pelos detectores de AUs (e todo o processo que os implica e foi detalhado no Capítulo 3), para comparar a AUs encontradas com as AUs da configuração padrão de EFCs, e assim reconhecer a EFCs correspondente, caso exista.

Uma vez terminada a implementação do sub-método para detecção de AUs em imagens 3D em 3.3, realiza-se o reconhecimento automático de EFCs nesse tipo de imagem. Para tal fim, empregam-se as mesmas bases de dados 3D utilizadas anteriormente na detecção de AUs (Bosphorus e BP4D). Assim, para poder determinar essa nova classificação de expressões nas bases selecionadas, foi elaborado um método que considera a configuração padrão de AUs de EFCs. Então, o referido método é aplicado para conseguir o reconhecimento de EFCs, executando o seguinte de experimento de duas maneiras: considerando as AUs anotadas originalmente nas bases (para comprovar existência das EFCs nas bases), e utilizando os detectores de AUs desenvolvidos (para reconhecer EFCs de maneira automática).

Inicialmente, antes da aplicação do método de reconhecimento de EFCs nas bases, torna-se importante realizar as seguintes observações sobre casos de expressões compostas a considerar:

- O método desenvolvido não tem a capacidade de diferenciar três EFCs: "tristeza com nojo", "horror" e "ódio", pois estas apresentam a mesma configuração de unidades de ação facial muscular (AUs 4 e 10), portanto, essas expressões são consideradas um "caso especial" .

- Logicamente nem todas as imagens apresentaram EFCs, dado que as bases utilizadas nos experimentos não são orientadas a EFC. Consequentemente, esse tipo de imagem, é nomeado como "caso desconhecido". Nota-se que este caso, nem sempre é igual a expressão neutra, imagens podem apresentar AUs, mas que em combinação não representam nenhuma expressão conhecida.

4.2 Experimentos

Com o método descrito na seção 4.1, realizou-se um experimento, com a finalidade de reconhecer EFCs em imagens 3D. Dessa maneira, agruparam-se as EFCs pelas suas AUs componentes, com a finalidade de diminuir buscas desnecessárias. O objetivo consistiu em determinar diretamente uma das EFCs, encontrando todas as suas AUs na imagem. Um exemplo desse tipo de funcionamento seria: caso em uma imagem de entrada estiverem presentes as AUs 1, 2, 5 e 25, se diz que se reconheceu a EFC de "impressão".

Este experimento executou-se inicialmente no total dos *datasets* de ambas as bases, com o fim de observar a possibilidade de reconhecer EFCs em imagens 3D, por meio das AUs anotadas, resultados serão apresentados em 4.3.1. Após isso, com os *datasets* de testes das duas bases, utilizados anteriormente na detecção de AUs em 3.4, efetivou-e o reconhecimento de EFCs por meio da detecção de AUs. Aqui, ponderou-se como acerto, cada imagem que apresentou a mesma EFC reconhecida tanto pelas AUs marcadas e quanto pelas detectadas. Em 4.3.2 debatem-se os resultados obtidos.

4.3 Resultados do reconhecimento de EFCs em imagens 3D

A maneira de apresentar os resultados obtidos pela execução do método de uma forma mais organizada, esta seção se divide em duas partes: na sub-seção 4.3.1 são apresentados os resultados da aplicação do método, considerando as AUs anotadas nas bases; na sub-seção 4.3.2, são apresentados os resultados do reconhecimento automático pela detecção de AUs.

4.3.1 Resultados considerando AUs anotadas nas bases

Uma vez executado o método, inicialmente examinou-se o reconhecimento de EFCs por meio das AUs anotadas, no total dos *datasets* de imagens, nas duas bases (2900 para a Bosphorus, e 10000 para a BP4D). Da base Bosphorus, observou-se que não são produzidos todos os casos estudados, apenas nove ocorreram nela: (0) "caso desconhecido", (2) "felicidade com nojo", (4) "tristeza com raiva", (5) "tristeza com supressa", (6) "caso especial", (7) "medo com raiva", (10) "raiva com supressa", (11) "medo com nojo", e (13) "impressão". Além disso, notou-se que efetivamente na base BP4D todos os casos de estudo existem: (0) "caso desconhecido", (1) "felicidade com supressa", (2) "felicidade com nojo", (3) "tristeza com medo" (4) "tristeza com raiva", (5) "tristeza com supressa", (6) "caso especial", (7) "medo com raiva", (8) "medo com supressa", (9) "medo com nojo", (10) "raiva com supressa", (11) "medo com nojo", (12) "nojo com supressa", (13) "impressão", (14) "felicidade com medo", e (15) "felicidade com tristeza".

Contudo, entende-se que EFCs podem ser reconhecidas automaticamente em imagens 3D. As Figuras 4.2 e 4.3, apresentam exemplos de EFCs encontradas pelo método desenvolvido, para as bases Bosphorus e BP4D, respectivamente. Cada exemplo de EFC está composto por uma imagem 3D acompanhada com uma imagem 2D correspondente a direita da mesma. Dessa forma na Figura 4.2, representam-se as expressões "felicidade com nojo" e "tristeza com raiva" na primeira fila, "tristeza com nojo" e "medo com raiva" na segunda, "raiva com supressa" e "raiva com nojo"na terceira, e "impressão" na ultima fila.

No entanto, na Figura 4.3, mostram-se exemplos dos 15 casos estudados (14 EFCs mais "caso especial"), apresentando na primeira fila as expressões: "felicidade com supressa", "felicidade com nojo" e "tristeza com medo". Na segunda: "tristeza com raiva", "tristeza com supressa" e "caso especial". Na terceira: "medo com raiva", "medo com supressa" e "medo com

Figura 4.2: Expressões faciais compostas encontradas automaticamente na base Bosphorus, toda imagem 3D está acompanhada da imagem 2D equivalente. Sendo na primeira fila as expressões: felicidade com nojo e tristeza com raiva. Na segunda fila: tristeza com nojo e medo com raiva. Na terceira fila: raiva com supressa e raiva com nojo. Na quarta fila: impressão

nojo". Na quarta: "raiva com supressa", "raiva com nojo" e "nojo com supressa". Na ultima: "impressão", "felicidade com medo" e "felicidade com tristeza".

Finalmente, estudou-se as porcentagens de ocorrência de EFCs nas duas bases. Estas são apresentadas na Tabela 4.1, e revelam claramente que a BP4D (ambiente espontâneo) teve um melhor desempenho, isto principalmente pelo fato de terem sido gerados todos os casos de expressões faciais compostas estudados, e por existirem muitos outros casos de imagens com as expressões. Ainda, nota-se que, em ambos ambientes, existe uma proporção maior de imagens que não apresentaram nenhuma EFC, isto pode ser explicado pelo fato que as bases utilizadas não foram desenvolvidas com a finalidade de estudar EFCs, porém, esse caso se destacou muito mais no ambiente forçado.

4.3.2 Resultados pela detecção automática de AUs

A seguir, expõem-se os resultados da aplicação do método detalhado em 4.1, mas utilizando os detectores de AUs implementados em 3.3. Nota-se que este experimento foi executado em ambas as bases, mas apenas naquela porção de imagens que foi empregada para os testes na detecção de AUs (290 imagens para a base Bosphorus e 1000 para a BP4D). Todavia, considerou-se como caso de acerto àquelas imagens que apesentaram a mesma EFC reconhecida pelas AUs anotadas na base e pelas detectadas automaticamente.

Então, inicialmente são considerados apenas os casos de EFCs no *dataset* de teste da base Bosphorus. Dos nove casos encontrados pelas AUs anotadas na base, existem ocorrência

Figura 4.3: Expressões faciais compostas encontradas automaticamente na base BP4D, toda imagem 3D está acompanhada da imagem 2D equivalente. Sendo na primeira fila as expressões: felicidade com supressa, felicidade com nojo, tristeza com medo. Na segunda fila: tristeza com raiva, tristeza com supressa, tristeza com nojo. Na terceira fila: medo com raiva, medo com supressa, medo com nojo. Na quarta fila: raiva com supressa, raiva com nojo, nojo com supressa. Na quinta fila: impressão, felicidade com medo, felicidade com tristeza

de oito deles no *dataset*, assim, realizou-se o reconhecimento de EFCs pela detecção de AUs, obtendo uma acurácia média de 84.83%. Na matriz de confusão da Figura 4.4, é possível observar valores de sensibilidade para cada EFC, assim, percebem-se excelentes resultados em (2) e (10), além disso, resultados muito bons em (0). Nos outros casos não se alcançaram resultados tão altos, por conta de que não foram identificadas as AUs das expressões correspondentes, ou porque foram detectadas algumas AUs que não estavam presentes na EFC procurada. Nota-se que em geral cada EFC encontrada apresentou poucos casos para analisar, consequentemente, variações bruscas nas sensibilidades são produzidas por casos onde não existiam muitas imagens com a EFC buscada. Além disso, houve confusão de expressões em várias das imagens, sendo o caso mais confundido aquele onde se marcou como (0) pelas AUs anotadas, mas que pelas AUs detectadas como (4), (5) ou (11), o curioso dos casos (4) e (5) é que não foram reconhecidos por apenas uma AU que foi marcada na base de dados, mas não foi detectada. No caso (11) todas as AUs foram identificadas originalmente na base, e também detectadas, porém, detectaram-se AUs adicionais que evitaram classificá-lo corretamente. É curioso que a confusão se deu entre expressões que AUs em comum, havendo casos onde apesar que foram detectadas todas as AUs

Tabela 4.1: Porcentagens (sobre 100%) de EFCs geradas nas bases Bosphorus e BP4D

Nro	**Categoria de expressão**	**Bosphorus**	**BP4D**
0	Caso desconhecido	98.58	58.83
1	Felicidade com surpresa	0.00	3.20
2	Felicidade com nojo	0.07	29.15
3	Tristeza com medo	0.00	0.03
4	Tristeza com raiva	0.07	0.57
5	Tristeza com surpresa	0.07	0.75
6	Caso especial	0.17	1.26
7	Medo com raiva	0.14	0.31
8	Medo com surpresa	0.00	0.01
9	Medo com nojo	0.00	0.13
10	Raiva com surpresa	0.07	0.42
11	Raiva com nojo	0.07	3.14
12	Nojo com surpresa	0.00	0.08
13	Impressão	0.76	0.09
14	Felicidade com medo	0.00	1.79
15	Felicidade com tristeza	0.00	0.24

de uma EFC, a mesma não foi considerada como acerto, pois adicionalmente outras AUs foram detectadas.

Por outro lado, consideram-se os casos de EFCs no *dataset* de teste da base BP4D. Aqui, existem 11 dos 15 casos de EFCs encontrados pelas AUs anotadas. Além disso, executou-se o método de reconhecimento de EFCs, mas com as AUs detectadas. Isto, com a finalidade de comparar as expressões reconhecidas. Dessa comparação obteve-se uma acurácia média de 78.50%. Na matriz de confusão da Figura 4.5 é possível distinguir os melhores resultados de sensibilidade individual nos casos (0), (4), e (5) principalmente, outros resultados obtidos são aceitáveis, mas nota-se que para (6) o método foi menos efetivo. Percebe-se que gerou-se confusão entre expressões que compartilham várias AUs em comum. Ainda, existem casos onde foram detectadas todas suas AUs, mas também outras AUs, como no caso (11) onde foram detectas as AUs 4, 10 e 17, sendo que as AUs anotadas são 4 e 10.

A maneira de complementar os resultados exibidos, apresenta-se a tabela 4.2, na mesma podem ser vistas as porcentagens de acurácia, sensibilidade, especifidade, valor preditivo, valor preditivo negativo, e valor de área sob a curva, do reconhecimento automático de EFCs para a base Bosphorus. Assim, nota-se o seguinte:

- Os valores obtidos de acurácia demonstraram que o método apresenta boa capacidade de reconhecimento, sendo o caso mais desafiante (0) com um valor de 86.2%, e o melhor o a expressão (8) com 100.0%. Porém existem evidencias que a acurácia não é um valor muito confiável [61].

- Além disso, os valores da sensibilidade mostraram a capacidade de reconhecer casos positivos, porém, não foi possível calcular o referido valor para todas as EFCs, pois não apresentaram existência na base, são representados na tabela com o valor "SD", e poderiam ser considerados como casos criticos. Mas há casos interessantes com valores de 100.0%, como (2) e (10).

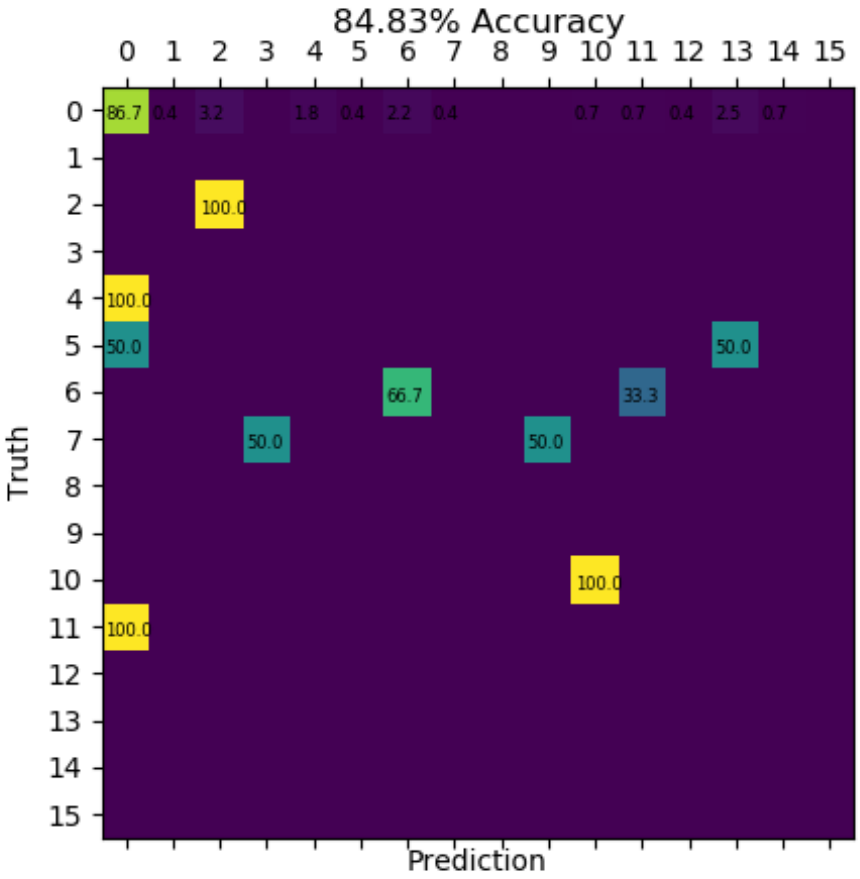


Figura 4.4: Matriz de confusão com as acurácias das EFCs reconhecidas na base Bosphorus. Sendo *"Truth"* as EFCs geradas na base pela anotação original de AUs, *"Prediction"* as EFCs reconhecidas pelas AUs detectadas. Os números representam: (0) "caso desconhecido", (1) "felicidade com supressa", (2) "felicidade com nojo", (3) "tristeza com medo", (4) "tristeza com raiva", (5) "tristeza com supressa", (6) "caso especial", (7) "medo com raiva", (8) "medo com supressa", (9) "medo com nojo", (10) "raiva com supressa", (11) "medo com nojo", (12) "nojo com supressa", (13) "impressão", (14) "felicidade com medo", e (15) "felicidade com tristeza"

- A especifidade reflete que o método foi capaz de reconhecer os casos negativos, com valores que vão desde 75.0% (0), até 100.0% (8).
- O valor preditivo positivo evidenciou a habilidade de classificar corretamente as EFCs, expondo como melhor o caso de (0), alcançando um 98.7%. Entretanto, o método não foi muito adequado para os outros casos, conseguindo valores baixos, ou não existindo oportunidade de obté-los.
- Também, o valor preditivo negativo representa um complemento para a especifidade, contendo valores que representam os casos que foram identificados erroneamente, sendo esses valores estudados principalmente para comprovar os valores de especifidade.
- Finalmente, os valores de AuC ilustram o desempenho do método, sendo para esta base de regular para ruim, pelos valores alcançados. Isto pode ser explicado pela pouca ou a nula existência de alguns dos casos de EFCs na base.

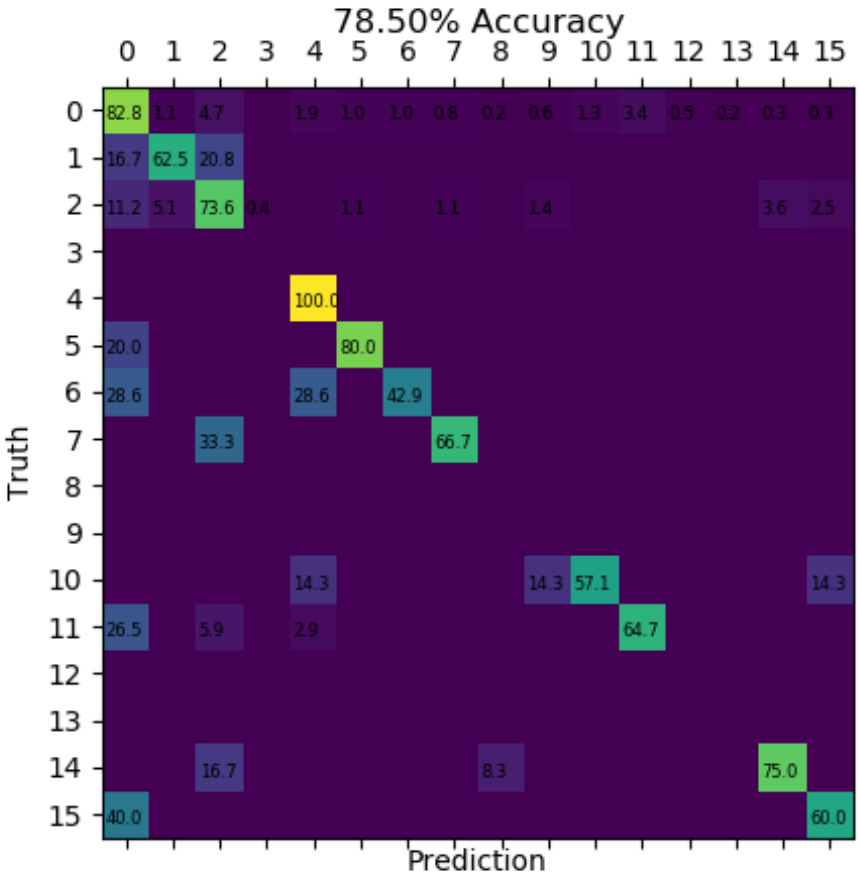


Figura 4.5: Matriz de confusão com as acurácias das EFCs reconhecidas na base BP4D. Sendo *"Truth"* as EFCs geradas na base pela anotação original de AUs, *"Prediction"* as EFCs reconhecidas pelas AUs detectadas. Os números representam: (0) "caso desconhecido", (1) "felicidade com supressa", (2) "felicidade com nojo", (3) "tristeza com medo", (4) "tristeza com raiva", (5) "tristeza com supressa", (6) "caso especial", (7) "medo com raiva", (8) "medo com supressa", (9) "medo com nojo", (10) "raiva com supressa", (11) "medo com nojo", (12) "nojo com supressa", (13) "impressão", (14) "felicidade com medo", e (15) "felicidade com tristeza"

- Embora existam bons resultados de acurácia, sensibilidade, especificidade, e valor preditivo negativo, Observa-se que ainda existem vários casos de falsos positivos e falsos negativos, o qual dificulta o reconhecimento de EFCs na base Bosphorus.

De igual forma, na tabela 4.3 são apresentadas as porcentagens de acurácia, sensibilidade, especifidade, valor preditivo, valor preditivo negativo, e valor de área sob a curva, do reconhecimento automático de EFCs, mas para a base BP4D. Desses valores é possível contatar o seguinte:

- Os valores de acurácia são altos, demonstrando um bom reconhecimento do método, com caso mais desafiante (0) com um valor de 84.4%, e os melhores (3) e (13) com 99.9%. Mas de igual forma que com a base Bosphorus, não é possível confiar nestes valores.

Tabela 4.2: Porcentagens de acurácia, sensibilidade, especifidade, valor preditivo, valor preditivo negativo, e valor de área sob a curva, de reconhecimento automático de EFCs para base Bosphorus para todas as EFCs. Os valores representados como "SD"aqueles onde não foi possível realizar o cálculo por falta de dados

	Acurácia	**Sensibilidade**	**Especifidade**	**Valor preditivo positivo**	**Valor preditivo negativo**	**AuC**
0	86.2	86.7	75.0	98.7	25.0	59.2
1	99.7	SD	99.7	0.0	0.3	52.1
2	96.9	100.0	96.9	18.2	3.1	60.6
3	99.7	SD	99.7	0.0	0.3	0.1
4	97.9	0.0	98.3	0.0	1.7	51.8
5	99.0	0.0	99.7	0.0	0.3	51.4
6	97.6	66.7	97.9	25.0	2.09	63.5
7	99.0	0.0	99.7	0.0	0.3	51.4
8	100.0	SD	100.0	SD	0.0	SD
9	99.7	SD	99.7	0.0	0.3	0.0
10	99.3	100.0	99.3	33.3	0.7	67.9
11	98.6	0.0	98.7	0.0	1.0	34.8
12	99.7	SD	99.7	0.0	0.3	52.1
13	97.2	SD	97.2	0.0	2.8	45.6
14	99.3	SD	99.3	0.0	0.7	52.1
15	100	SD	100.0	SD	0.0	SD

- A sensibilidade apresentada refrete casos com altos valores, com o (4) como o melhor com um 100.0%. Considera-se como casos mais desafiantes aqueles que não foram calculados (representados com "SD"), isto pela não de existência de casos na base.
- Da especifidade obtida, afirma-se que o método alcançou valores altos que superam os 87%. Aqui nota-se boa capacidade de reconhecer casos corretamente os casos negativos para todos os casos de EFCs considerados.
- O valor preditivo positivo mostrou um bom desempenho para os casos (0) e (2), 91.3% e 83.9%, respetivamente. Os outros valores são baixos devido os poucos casos de EFCs existentes.
- O valor preditivo negativo Também, foi considerado como complemento da especifidade, sendo utilizados para comprovar os valores da mesma.
- Por fim, os valores de AuC ilustram um desempenho melhor do que no caso da base Bosphorus, resultando muito bom para o caso (0) com um 96.0%, bom para (2), (6), (10) e (11) com valores entre 75.0a% e 90.0%. Nos outros casos o método não se mostrou muito ótimo.
- Assim , apesar que se consideram bastante bons os resultados logrados na acurácia, sensibilidade, especificidade, e valor preditivo negativo, também há vários casos de falsos positivos e falsos negativos, mas proporcionalmente menores do que na base Bosphorus, no entanto ainda existe dificuldade no reconhecimento de EFCs na base BP4D.

Tabela 4.3: Porcentagens de acurácia, sensibilidade, especifidade, valor preditivo, valor preditivo negativo, e valor de área sob a curva, de reconhecimento automático de EFCs para base BP4D para todas as EFCs. Os valores representados como "SD"aqueles onde não foi possível realizar o cálculo por falta de dados

	Acurácia	**Sensibilidade**	**Especifidade**	**Valor preditivo positivo**	**Valor preditivo negativo**	**AuC**
0	84.4	82.8	87.1	91.3	12.9	96.0
1	97.0	62.5	97.9	41.7	2.1	63.0
2	88.8	73.6	94.6	83.9	5.4	86.3
3	99.9	SD	99.9	SD	0.1	21.7
4	98.4	100	98.4	27.3	1.6	65.8
5	99.0	80.0	99.1	30.8	0.9	67.4
6	99.0	42.9	99.4	33.3	0.6	79.0
7	99.1	66.7	99.2	20.0	0.8	60.8
8	99.8	SD	98.8	SD	0.2	35.0
9	99.1	SD	99.1	SD	0.9	40.8
10	98.9	54.1	99.2	33.3	0.8	79.1
11	96.7	64.7	97.8	51.2	2.2	83.8
12	99.7	SD	99.7	SD	0.3	69.0
13	99.9	SD	99.9	SD	0.1	67.3
14	98.5	75.0	98.8	42.9	1.2	58.7
15	98.8	60.0	99.0	23.1	1.0	45.3

4.3.3 Observações

O método desenvolvido considera-se estado da arte, pois não existem outros estudos desse tipo de expressões (EFCs) em nenhuma das bases com imagens 3D, até o momento. Realizou-se a busca de EFCs em imagens 3D, pois pretendia-se aproveitar as vantagens que esse tipo de imagem possui. Aproveitou-se a não dependência tanto de pose, iluminação e mudanças na aparência facial que as imagens 3D de ambas bases oferecem. Não foram considerados casos de imagens com algum tipo de oclusão, inclusive foram eliminadas aquelas imagens da base BP4D que apresentaram essa característica (oclusão). Além disso, o método se mostrou útil em reconhecer varias das EFCs nas duas bases escolhidas. Contudo, os resultados demostraram melhor desempenho no reconhecimento de EFCs em ambientes espontêos.

Embora não exista estudos que permitam realizar comparações com o trabalho desenvolvido, encontrou-se alguns similares, que possibilitam analisar o comportamento dos resultados obtidos. Portanto, esses estudos detalham-se brevemente a seguir:

- Acredita-se que um dos trabalhos mais interessantes a considerar é o projeto implementado por Kastemaa em [59], já que os experimentos realizados também foram orientados ao reconhecimento de EFCs (sete casos) aplicando em um tipo de imagens 3D (agentes virtuais). Deve enfatizar-se no fato que agentes virtuais são apenas uma representação humana diferente de pessoas reais, e que o reconhecimento não foi realizado por um método automático, foram pessoas que observando a agente virtual, reconheceram as expressões produzidas por ela. Resultados representam quase a mesma realidade que os apresentados nas Figuras 4.4 e 4.5, com valores de sensibilidade entre 60% e 90%.

- Uma outra pesquisa pertinente a examinar, é o método desenvolvido por Liu *et. al* em [23], já que apesar dos experimentos serem processados com imagens 2D, de alguma forma, apresenta uma estrutura similar com o método desenvolvido: descritor de características LBP e classificadores SVM. Aqui nota-se que: imagens foram coletadas da Internet e o reconhecimento acontece em apenas nove EFCs. Nos resultados dessa pesquisa, bastante dificuldade em reconhecer as expressões compostas. Também, nota-se que são apresentados baixos valores de sensibilidade (em geral um 27%), valor preditivo positivo (28.2%), e acurácia (34.9%), considerando LBP como descritor de textura e SVM como classificador. Mas valores um pouco melhores com o classificador de média de classe mais próxima (NCMML) alcançando um sensibilidade de 30.1%, valor preditivo positivo de 29.2% e acurácia de 36.7%.

- Finalmente, explora-se a investigação de Du *et. al* em [3], autores originais que definiram as EFCs. O sentido da observação cai em estudar apenas aspectos de similaridade do comportamento no reconhecimento de EFCs. Então, verificou-se que da mesma forma que em [23], empregam imagens 2D, mas nesse caso, essas imagens foram capturadas em ambientes forçados. Além disso, realizaram o reconhecimento em 15 EFCs, pois as outras duas foram definidas posteriormente em [21]. Para tal fim, aplicaram como classificador SVM, mas na versão multi-classes. Os resultados são muito bons para a maioria das EFCs, não obstante, alguns valores de acurácia ainda não são muito altos. Os valores apresentados de sensibilidade vão desde 54% atá 93%.

Uma das dificuldades do método desenvolvido para reconhecimento de EFCs é de não poder diferenciar as expressões "tristeza com nojo", "ódio" e "horror". Na Figuras 4.6 e 4.7, apresentam-se alguns exemplos das expressões anteriormente mencionadas, para as bases Bosphorus e BP4D, respetivamente. Nas imagens componentes de ambas as Figuras, por observação simples, nota-se a diferença entre elas e consegue categorizá-las. Portanto, para poder classificar todas as imagens que pertençam ao denominado "caso especial"na secção 4.1, deve realizar-se uma análise visual. Torna-se interessante que em [34], essas expressões não compartilham a mesma configuração de AUs que em [3, 21], de fato, não são as únicas que apresentam variações, consequentemente, abre a opção de experimentar considerando essa outra padronização de EFCs pode resultar em um reconhecimento com valores de acurácia superiores que consigam uma separação das 17 expressões compostas.

Figura 4.6: Expressões faciais compostas não diferenciáveis pelo método desenvolvido da base Bosphorus, por ter a mesma configuração de AUs. Toda imagem 3D está acompanhada da imagem 2D equivalente. Sendo as expressões: "tristeza com nojo", "ódio" e "horror".

4.4 Ambientes forçados vs ambientes espontâneos

Como um dos fins da presente pesquisa é estudar o comportamento dos ambientes forçado é espontâneo, no que se refere ao reconhecimento de EFCs, inicialmente foram observadas

Figura 4.7: Expressões faciais compostas não diferenciáveis pelo método desenvolvido da base BP4D, por ter a mesma configuração de AUs. Toda imagem 3D está acompanhada da imagem 2D equivalente. Sendo as expressões: "tristeza com nojo", "ódio" e "horror".

bases de dados publicamente disponíveis, para poder utilizá-las, e determinar, por meio do método de reconhecimento automático implementado, quais dos dois ambientes representa de melhor forma as expressões compostas. Assim, revisando na literatura algumas das bases de dados orientadas a expressões faciais disponíveis [5, 6, 44, 40, 24], percebeu-se que as suas imagens foram capturadas em ambientes espontâneos ou forçados.

Na presente pesquisa, denomina-se ambiente forçado aquele onde a captura das imagens se realiza em espaços controlados. Cabe destacar que neste tipo de ambientes, sujeitos foram instruídos a realizar uma determinada expressão, por exemplo: no momento da captura, apresentou-se uma fotografia com uma expressão objetivo para um participante, este tentou imitá-la da maneira mais fiel possível. Por outro lado, se denominou ambiente espontâneo aquele onde as imagens foram obtidas de uma maneira mais "natural", em locais não controlados ou irrestritos. Nestes, os participantes expressaram as emoções que estavam sentindo no momento da captura, pela influência de algum estimulo, por exemplo, por um cheiro desagradável não esperado pelo indivíduo, que lhe provocou a expressão de nojo.

No entanto, na captura de expressões faciais, em muitos casos, utilizam-se atores profissionais, como na base Bosphorus. Isto com a finalidade de capturar emoções de maneira muito "natural", já que têm muito mais controle dos músculos faciais e conseguem expressar algumas emoções de maneira mais realista [62], porém, sempre existirá alguma diferença com as expressões espontâneas, além disso, não serão transmitidas as verdadeiras emoções do ator pelas expressões faciais "atuadas", portanto, esse tipo de captura considera-se forçado.

Simultaneamente, observou-se que existem vários autores [6, 63, 38, 64, 65, 66], interessados no estudo de expressões forçadas e espontâneas. Estes quais comprovaram que expressões forçadas e espontâneas diferem em comportamento de várias dimensões, incluindo complexidade, temporalidade, intensidade, e até na maneira de detectar as suas AUs. Não obstante, para Du *et al.* em [21], a configuração padrão de AUs nas EFCs é a mesma para os dois ambientes, apesar de que os sistemas neurais envolvidos na produção de expressões forçadas e espontâneas serem diferentes. Sob essa perspetiva, aplicou-se o mesmo método de estudo de EFCs em ambos os ambientes de captura.

Por fim, na presente pesquisa, confirma-se que ambientes forçados e espontâneos são diferentes e que é mais fácil encontrar EFCs em ambientes espontâneos, de acordo com resultados do método desenvolvido. Neste caso, a base BP4D teve um melhor comportamento em comparação com a base BP4D no reconhecimento de EFCs, já que apresentou todos os casos de expressões compostas estudados, e para cada caso, proporcionalmente mais imagens. Além disso, os resultados de reconhecimento obtidos de: acurácia, sensibilidade, especifidade, valor preditivo positivo, valore preditivo negativo, e AuC, representaram uma melhor realidade e com maior sentido, sendo mais próximos ao ideal (100.0% para os todos esses valores calculados, e 0.0% para o valor preditivo negativo) no ambiente espontâneo. Seria interessante poder experimentar em mais bases de ambos os ambientes, para poder consolidar esta afirmação, mas não existem

mais bases espontâneas com anotação de AUs e de pontos fiduciais disponíveis publicamente, até o momento.

5 Conclusões

No presente trabalho examinou-se o reconhecimento de EFCs em imagens 3D, com a finalidade de comparar o comportamento desse tipo de expressões nos ambientes de captura forçado e espontâneo. Para tal fim, desenvolveu-se um método que aplica detectores de AUs e que considera a configuração padrão de EFCs para decidir a expressão correspondente na imagem objetivo. Dessa maneira, realizou-se a busca das expressões compostas em duas bases de dados públicas com imagens 3D: Bosphorus e BP4D-Spontaneous, pertencentes os ambientes forçado e espontâneo, respetivamente. Resultados demonstram a possibilidade de encontrar essa moderna definição de expressões faciais em imagens 3D, sendo mais fácil o reconhecimento no ambiente espontâneo. Mas também, aproveitou-se algumas das características que as imagens 3D oferecem, como a não dependência de pose e de iluminação.

Inicialmente, realizou-se um levantamento sobre as expressões faciais; estudando sua importância, sua definição sistemática e duas de suas principais classificações. Além disso, observou-se que expressões faciais são consideradas cada vez mais em visão computacional. Entretanto, a maioria das pesquisas nessa área estão sendo focadas no reconhecimento de expressões faciais básicas em imagens 2D, capturadas em condições de laboratório. Porém, evidenciou-se que esse tipo de expressões não responde a muitos dos estados emocionais humanos, além disso, imagens de expressões faciais capturadas em condições controladas não representam as verdadeiras emoções transmitidas pelos indivíduos, finalmente, acredita-se que imagens 3D conseguem resolver algumas das dificuldades que imagens 2D apresentam.

Com base na revisão bibliográfica, iniciou-se o desenvolvimento de um método para análise de expressões faciais compostas. O mesmo considera representações de mais casos de estados emocionais (as expressões compostas), com a finalidade de o aplicar em imagens 3D, exatamente em duas bases de dados públicas (Bosphorus e BP4D) capturadas em dois ambientes diferentes (forçado e espontâneo).

Também, analisaram-se algumas opções de detectores de AUs em imagens 3D, já que o método desenvolvido para análise de EFC se baseia neles. Assim, escolheu-se como *baseline* os detetores que aplicam o descritor de textura LBP em mapas de profundidade para serem implementados por sua simplicidade na implementação e por apresentarem resultados convincentes na detecção de AUs. Logo, para avaliar a acurácia dos detectores desenvolvidos, experimentos com as bases Bosphorus e BP4D foram realizados. Concluindo que apesar dos resultados variarem com os obtidos no *baseline*, estes não apresentam muita divergência e fazem sentido comparados entre si. Portanto, os detectores de AUs seriam úteis na análise EFCs.

Por outro lado, desenvolveu-se um método que por meio da utilização da configuração padrão de EFCs permitiu realizar uma análise dessa moderna categoria de expressões faciais. O mesmo foi aplicado nas bases já utilizadas nos experimentos dos detectores de AUs (Bosphorus e BP4D). Notou-se que existem três EFCs que compartilham a mesma configuração de AUs, estas foram consideradas como "caso especial". Também, evidenciou-se que nem todas as imagens apresentaram EFCs, a maioria delas contêm algumas AUs que em combinação não representam

nenhuma das expressões conhecidas, estes casos se denominaram como "casos desconhecidos". Portanto, se analisaram esses dois casos, mais 14 EFCs.

Desse modo, primeiro foi estudada a produção de EFCs nas bases. Isto, ocorreu reconhecendo EFCs pelas AUs anotadas originalmente em ambas as bases por especialistas em FACS, comparando-as com as presentes no padrão de EFCs. Resultados demostraram maior efetividade na base BP4D, já que foram gerados todos os casos estudados, havendo proporcionalmente muitos mais exemplos do que na base Bosphorus, a qual tampouco apresentou todos os casos, apenas sete deles foram gerados nela.

Posteriormente, realizou-se o reconhecimento de EFCs, mas pelas AUs detectadas, em porções das bases Bosphorus e BP4D, utilizadas para testes nos detectores de AUs implementados em 3.3. Assim, para essa finalidade aplicaram-se esses detectores de AUs anteriormente desenvolvidos, contrapondo as AUs detectadas com as da configuração de EFCs, para reconhecer as expressões compostas. Resultados apresentaram acurácias em média altas para ambas as bases, porém, evidencia-se que BP4D foi mais efetiva, por reconhecer mais casos de EFCs.

Embora não seja possível comparar com outros trabalhos na literatura, realizou-se observações de alguns existentes [59, 23, 3] que fizeram métodos similares com a mesma finalidade (reconhecer EFCs), claro que com variações (como o emprego de imagens 2D, reconhecimento não automático, ou descritores de textura e classificadores diferentes). Nesses trabalhos notou-se uma mesma realidade; a dificuldade de reconhecer EFCs.

Por fim, ressalta-se alguns pontos importantes que a atual pesquisa evidenciou. Primeiro, foram aproveitadas algumas propriedades das imagens 3D, como ser que não apresentam problemas decorrentes de variações de pose, iluminação e de outras mudanças na aparência facial. Além disso, comprovou-se que EFCs existem em imagens 3D, isto foi fatível por meio de observação do padrão de AUs dessas expressões, originalmente anotadas nas bases utilizadas, mas também aplicando detecção de AUs. Finalmente, pelos experimentos realizados, notou-se que para a maioria dos casos, alcançou-se bons valores de acurácia, sensibilidade, especifidade, e valor preditivo negativo para ambas as bases, revelando que o método consegue realizar corretas predições, tanto positivas quanto negativas. Porém, com o valor preditivo positivo e AuC, percebe-se que ainda existe uma taxa grande de falsos positivos e falsos negativos.

Adicionalmente, comparando o comportamento de ambas as bases, destaca-se que o ambiente espontâneo alcançou demonstrar melhor desempenho, apresentando mais casos de EFCs e melhores resultados das medidas consideras, comparado ao ambiente forçado. Sendo o mais interessante o valor de AuC, já que permite concluir que em geral o ambiente espontâneo é mais ótimo.

O método desenvolvido mostrou-se útil no reconhecimento de EFCs, pois conseguiu identificá-las em duas bases de dados (Bosphorus e BP4D), no entanto, existem elementos que devem mudar para conseguir resultados melhores. Acredita-se que conseguindo maior acurácia nos detectores de AUs em imagens 3D, serão mais altos os valores de acurácia no reconhecimento de EFCs. Para tal fim, podem-se considerar outras opções para a detecção de AUs e inclusive aplicar redes neurais. Além disso, pode-se experimentar com métodos baseados em aparência (onde se utilizaria as imagens 3D como um todo para o reconhecimento das EFCs), ou poderia empregar características geométricas (propriedades que os pontos fiduciais das imagens 3D oferecem) como aconteceu em [41].

Referências Bibliográficas

[1] Albert Mehrabian. *Nonverbal communication*. Transaction Publishers, 1972.

[2] K.Peg. Facial action coding system. `http://www.paulekman.com/product-category/facs/`, 2016. Acessado em 01/08/2017.

[3] Shichuan Du and Aleix M Martinez. Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in clinical neuroscience*, 17(4):443, 2015.

[4] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[5] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Bosphorus database for 3d face analysis. In *European Workshop on Biometrics and Identity Management*, pages 47–56. Springer, 2008.

[6] Xing Zhang, Lijun Yin, Jeffrey F Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu, and Jeffrey M Girard. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10):692–706, 2014.

[7] G. Sandbach, S. Zafeiriou, and M. Pantic. Binary pattern analysis for 3d facial action unit detection. In *Proceedings of the British Machine Vision Conference, BMVC 2012, Surrey, UK*, pages 119.1–119.12, UK, September 2012. BMVA Press.

[8] Paul Ekman, Wallace Friesen, and JC Hager. Facial action coding system: The manual on cd-rom. instructor's guide. *Network Information Research Co, Salt Lake City*, 2002.

[9] Lisa A Parr and Bridget M Waller. Understanding chimpanzee facial expression: insights into the evolution of communication. *Social Cognitive and Affective Neuroscience*, 1(3):221–228, 2006.

[10] Paul Ekman and Dacher Keltner. Universal facial expressions of emotion: an old controversy and new findings. *Nonverbal Communi-cation: Where Nature Meets Culture*, pages 27–46, 1997.

[11] R. Jameel, A. Singhal, and A. Bansal. A comprehensive study on facial expressions recognition techniques. In *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, pages 478–483, Jan 2016.

[12] Wallace Friesen and Paul Ekman. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto*, 1978.

[13] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53. IEEE, 2000.

[14] Jeffrey F Cohn, Zara Ambadar, and Paul Ekman. Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, pages 203–221, 2007.

[15] JC Hager, Paul Ekman, and Wallace Friesen. Facial action coding system. salt lake city, ut: A human face. Technical report, ISBN 0-931835-01-1, 2002.

[16] A Freitas-Magalhães. *O código de Ekman: O cérebro, a face e a emoção*. Leya, 2015.

[17] A Freitas-Magalhães. *Facial Action Coding System-Manual de Codificação Científica da Face Humana*. Leya, 2016.

[18] Guillaume-Benjamin Duchenne and R Andrew Cuthbertson. *The mechanism of human facial expression*. Cambridge university press, 1990.

[19] Paul Ekman and Wallace V Friesen. *Pictures of facial affect*. consulting psychologists press, 1975.

[20] Carl-Herman Hjortsjö. *Man's face and mimic language*. Studen litteratur, 1969.

[21] Shichuan Du, Yong Tao, and Aleix M Martinez. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 111(15):E1454–E1462, 2014.

[22] Jyoti Kumari, R. Rajesh, and K.M. Pooja. Facial expression recognition: A survey. *Procedia Computer Science*, 58(Supplement C):486 – 491, 2015. Second International Symposium on Computer Vision and the Internet (VisionNet'15).

[23] Zhiwen Liu, Shan Li, and Weihong Deng. Recognizing compound emotional expression in real-world using metric learning method. In *Chinese Conference on Biometric Recognition*, pages 528–536. Springer, 2016.

[24] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Lijun Yin. Static and dynamic 3d facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012.

[25] Ray L Birdwhistell. *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press, 1970.

[26] Eric Robert Yudin and Research Thesis. Improving facial expression analysis via intrinsic normalization of surfaces improving facial expression analysis via intrinsic normalization of surfaces. In *Improving Facial Expression Analysis via Intrinsic Normalization of Surfaces*, 2015.

[27] Brais Martinez and Michel F Valstar. Advances, challenges, and opportunities in automatic facial expression recognition. In *Advances in Face Detection and Facial Image Analysis*, pages 63–100. Springer, 2016.

[28] Nicholas Vretos, Nikos Nikolaidis, and Ioannis Pitas. 3d facial expression recognition using zernike moments on depth images. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 773–776. IEEE, 2011.

[29] G. Sandbach, S. Zafeiriou, and M. Pantic. Local normal binary patterns for 3d facial action unit detection. In *2012 19th IEEE International Conference on Image Processing*, pages 1813–1816, Sept 2012.

[30] Stefano Berretti, Boulbaba Ben Amor, Mohamed Daoudi, and Alberto Del Bimbo. 3d facial expression recognition using sift descriptors of automatically detected keypoints. *The Visual Computer*, 27(11):1021–1036, 2011.

[31] T. W. Shen, H. Fu, J. Chen, W. K. Yu, C. Y. Lau, W. L. Lo, and Z. Chi. Facial expression recognition using depth map estimation of light field camera. In *2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–4, Aug 2016.

[32] Universitat Autònoma de Barcelona. Detección de objetos. `https://www.coursera.org/learn/deteccion-objetos/lecture/JXyER/l4-6-support-vector-machines-svm-desarrollo-matematico`, 2017. Acessado em 04/01/2018.

[33] Zhiwen Liu, Shan Li, and Weihong Deng. Boosting-poof: Boosting part based one vs one feature for facial expression recognition in the wild. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 967–972. IEEE, 2017.

[34] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5562–5570, 2016.

[35] Caifeng Shan, Shaogang Gong, and Peter W McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.

[36] A. Savran, B. Sankur, and M. T. Bilge. Facial action unit detection: 3d versus 2d modality. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 71–78, June 2010.

[37] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53. IEEE, 2000.

[38] Marian Stewart Bartlett, Gwen Littlewort, Mark G Frank, Claudia Lainscsek, Ian R Fasel, and Javier R Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of multimedia*, 1(6):22–35, 2006.

[39] Yonggang Huang, Yunhong Wang, and Tieniu Tan. Combining statistics of geometrical and correlative features for 3d face recognition. In *BMVC*, pages 879–888. Edinburgh, 2006.

[40] Darren Cosker, Eva Krumhuber, and Adrian Hilton. A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling. In *2011 International Conference on Computer Vision*, pages 2296–2303. IEEE, 2011.

[41] N. Bayramoglu, G. Zhao, and M. Pietikäinen. Cs-3dlbp and geometry based person independent 3d facial action unit detection. In *2013 International Conference on Biometrics (ICB)*, pages 1–6, June 2013.

[42] Eric Yudin, Aaron Wetzler, Matan Sela, and Ron Kimmel. Improving 3d facial action unit detection with intrinsic normalization. In S. Kurtek H. Drira and P. Turaga, editors, *Proceedings of the 1st International Workshop on DIFFerential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories (DIFF-CV 2015)*, pages 5.1–5.10. BMVA Press, September 2015.

[43] Inc. China SouVR Co. Inspeck 3d mega capturor ii digitizer. `http://www.en.souvr.com/product/200712/324.html/`, 2008. Acessado em 01/08/2017.

[44] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J Rosato. A 3d facial expression database for facial behavior research. In *7th international conference on automatic face and gesture recognition (FGR06)*, pages 211–216. IEEE, 2006.

[45] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek. Overview of the face recognition grand challenge. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 947–954 vol. 1, June 2005.

[46] Timothy C Faltemier, Kevin W Bowyer, and Patrick J Flynn. Using a multi-instance enrollment representation to improve 3d face recognition. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6. IEEE, 2007.

[47] Thomas Heseltine, Nick Pears, and Jim Austin. Three-dimensional face recognition using combinations of surface feature map subspace components. *Image and Vision Computing*, 26(3):382–396, 2008.

[48] Cheng Zhong, Zhenan Sun, and Tieniu Tan. Robust 3d face recognition using learned visual codebook. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007.

[49] Ana Belen Moreno and Armand Sánchez. Gavabdb: a 3d face database. In *Proc. 2nd COST275 Workshop on Biometrics on the Internet, Vigo (Spain)*, pages 75–80, 2004.

[50] Charles Beumier and Marc Acheroy. Face verification from 3d and grey level clues. *Pattern recognition letters*, 22(12):1321–1329, 2001.

[51] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3d dynamic facial expression database. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference On*, pages 1–6. IEEE, 2008.

[52] S. Aly, A. Trubanova, L. Abbott, S. White, and A. Youssef. Vt-kfer: A kinect-based rgbd+time dataset for spontaneous and non-spontaneous facial expression recognition. In *2015 International Conference on Biometrics (ICB)*, pages 90–97, May 2015.

[53] Shangfei Wang, Zhilei Liu, Siliang Lv, Yanpeng Lv, Guobing Wu, Peng Peng, Fei Chen, and Xufa Wang. A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Transactions on Multimedia*, 12(7):682–691, 2010.

[54] S Mohammad Mavadati, Mohammad H Mahoor, Kevin Bartlett, Philip Trinh, and Jeffrey F Cohn. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2):151–160, 2013.

[55] Mohammad H Mahoor, Steven Cadavid, Daniel S Messinger, and Jeffrey F Cohn. A framework for automated measurement of the intensity of non-posed facial action units. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 74–80. IEEE, 2009.

[56] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861 – 874, 2006. ROC Analysis in Pattern Recognition.

[57] Maurício Pamplona Segundo, Luciano Silva, Olga Regina Pereira Bellon, and Chauã C Queirolo. Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(5):1319–1330, 2010.

[58] Charles Darwin and Phillip Prodger. *The expression of the emotions in man and animals*. Oxford University Press, USA, 1872.

[59] Juho Kastemaa. Recognizing compound facial expressions of virtual characters in augmented reality, 2017.

[60] Kaili Zhao, Wen-Sheng Chu, Fernando De la Torre, Jeffrey F Cohn, and Honggang Zhang. Joint patch and multi-label learning for facial action unit detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2207–2216, 2015.

[61] Charles E Metz. Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier, 1978.

[62] Harald G Wallbott and Klaus R Scherer. Cues and channels in emotion recognition. *Journal of personality and social psychology*, 51(4):690, 1986.

[63] Michel F Valstar, Maja Pantic, Zara Ambadar, and Jeffrey F Cohn. Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 162–170. ACM, 2006.

[64] Shushi Namba, Shoko Makihara, Russell S Kabir, Makoto Miyatani, and Takashi Nakao. Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences. *Current Psychology*, pages 1–13, 2016.

[65] Sherin Aly, Andrea Trubanova, Lynn Abbott, Susan White, and Amira Youssef. Vt-kfer: A kinect-based rgbd+ time dataset for spontaneous and non-spontaneous facial expression recognition. In *Biometrics (ICB), 2015 International Conference on*, pages 90–97. IEEE, 2015.

[66] Zhiwen Liu, Shan Li, and Weihong Deng. Boosting-poof: Boosting part based one vs one feature for facial expression recognition in the wild. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 967–972. IEEE, 2017.