AFER – An Algorithm for Detection and Affective Analysis of Facial Expressions in Human Computer Interaction

AFER – Affective Facial Expression Recognition

Thallys, R. M., Costalat

Federal University of Para, Institute of Technology, thallys.costalat@itec.ufpa.br

Arissa, O., Niwa

Federal University of Para, Institute of Technology, arissaniwa3@gmail.com

Fabíola, P. O., Araújo

Federal University of Para, Institute of Technology, fpoliveira@ufpa.br

In this context of technological advancements and the Information Revolution, connecting humans and machines has become imperative. One of the biggest challenges nowadays is making computers capable of identifying human emotions and, perhaps, understanding them. Given the complexity of this task, several techniques need to be applied, fully exploring the interdisciplinarity involved in the process. This work aims to, through concepts and practices of affective computing and human-computer interaction, develop effective means to detect emotions through facial expressions, associating computer vision tools with psychological studies. Using the Face Recognition library implemented in Python, satisfactory results were obtained after parameter adjustments in the algorithm, achieving recognition of more than 80% in 5 of the 6 analyzed emotions.

**Additional Keywords and Phrases:** Emotion Detection, Affective Computing, Human-Computer Interaction, Computer Vision, Digital Image Processing

ACM Reference Format:

Arissa, O., Niwa, Thallys, R.M., Costalat, Fabíola, P. O., Araújo. 2021. AFER – An Algorithm for Detection and Affective Analysis of Facial Expressions in Human Computer Interaction: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock ’18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1. Introdução

Treinar máquinas para detectar e agir de acordo com o estado emocional do usuário, simulando reações que remetam a emoções humanas e tornem a interação homem-máquina tão natural quanto possível é um dos principais objetivos da Computação Afetiva. Neste ponto evidencia-se a forte ligação existente entre Computação Afetiva, Interação Humano-Computador e a Experiência do Usuário (UX).

A associação das três áreas de estudo supracitadas nos possibilita aprofundar conhecimentos e desenvolver técnicas mais robustas para uma adaptação gradativa e eficiente da tecnologia nos mais variados aspectos do cotidiano humano. Para tanto, é necessário que seja feita uma ponte entre a computação e estudos na área da psicologia com ênfase na detecção de expressões faciais e gestos, uma vez que a linguagem corporal é a principal forma de comunicação humana segundo estudos do antropólogo Ray Birdwhistell (1918—1994).

1. MATERIAIS E MÉTODOS

Nesta seção, serão apresentados os principais pontos acerca dos materiais utilizados para a produção do presente algoritmo, descrevendo-se as bibliotecas Face Recognition, utilizada na detecção de faces em uma imagem, assim como mapear os pontos da face (*landmarks*) presente na mesma, e OpenCV, responsável pela manipulação e captura dos *frames* envolvidos neste estudo. Ambas bibliotecas são utilizadas conjuntamente.

Além disso, é descrito a linguagem de programação em Python, utilizada neste estudo para o desenvolvimento de cálculos algébricos e tratamento de dados a partir de bibliotecas já presentes na linguagem, como Pandas, Numpy e CSV.

Por fim, é explicitada a metodologia desenvolvida para que fossem alcançados os objetivos deste artigo, atentando-se para conceitos de Computação Afetiva.

* 1. Materiais Utilizados
     1. Sistema FACS. XXXXXXXX XXXXXXXXXXXX XXXXXXXXXXX XXXXXXX XXX X X X XXXXXXXXX XXXXXXXXXXXXX XXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXX XXXXXXXXXXXXX XXXXXXXXXX XXXXX X X X XXXXXXXX XXXXXXXXXXXXXXXXXXXX X X X X XXXXXXXXXXXXXXXXXX X X X XXXXXXXXXXXXXXXXXXXXXXX X X XXXXX.
     2. OpenCV. OpenCV (Open Source Computer Vision Library) é uma biblioteca de software de visão computacional e aprendizado de máquina de uso livre. Tal ferramenta foi desenvolvida para fornecer uma infraestrutura comum para aplicativos de visão computacional. Possui interfaces C ++, Python, Java e MATLAB e suporta Windows, Linux, Android e Mac OS. A biblioteca OpenCV é focada principalmente em aplicativos de visão em tempo real e tira proveito das instruções MMX e SSE, quando disponíveis [1].

A biblioteca tem mais de 2.500 algoritmos otimizados, que incluem um conjunto abrangente de algoritmos de visão computacional e aprendizado de máquina clássicos e de última geração. Esses algoritmos podem ser usados para detectar e reconhecer rostos, identificar objetos, classificar ações humanas em vídeos, rastrear movimentos de câmera, rastrear objetos em movimento, reconhecer cenários e estabelecer marcadores para sobrepor os mesmos com realidade aumentada, etc. Além disso, é amplamente utilizada por empresas, grupos de pesquisa e órgãos governamentais.

* + 1. Face Recognition Library. Face Recognition é uma biblioteca opensource capaz de estimar a posição de até 68 pontos da face, chamados landmarks (Figura 1). Foi utilizada para implementar, em conjunto com a biblioteca OpenCV (Seção 2.1.2) e técnicas de programação em linguagem Python 3.7, um sistema de capaz de calcular a distância entre os landmarks presentes em imagem, a qual servirá de base para aferir mudanças na estrutura facial do usuário.

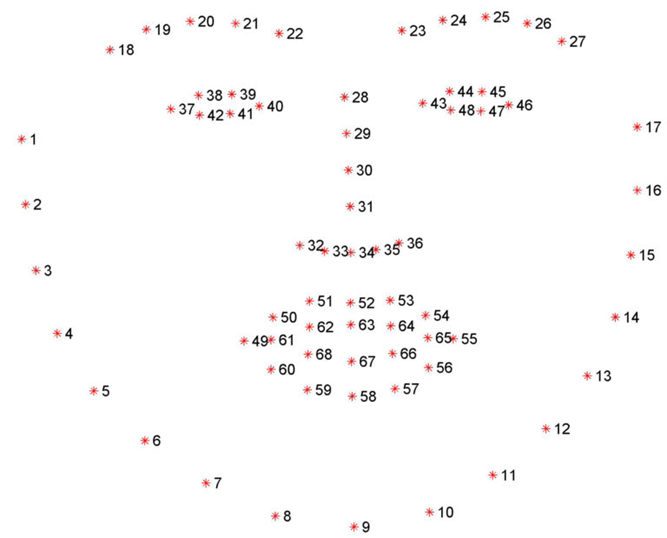


Figura 1: Adaptado de (ROSEBROCK, 2017) [4].

A biblioteca possibilita reconhecer e manipular dados das faces a partir de Python ou linha de comando. Foi desenvolvida utilizando técnicas de reconhecimento facial de última geração da dlib (biblioteca de software multiplataforma de uso geral escrita na linguagem de programação C ++) construído a partir de técnicas de *deep learning*. O modelo tem precisão de 99,38% no Labeled Faces in the Wild Benchmark [2], um banco de dados de fotos de rosto projetado para estudar o problema de reconhecimento irrestrito de rosto. O conjunto de dados contém mais de 13.000 imagens de rostos coletados da web. Cada rosto foi etiquetado com o nome da pessoa retratada. 1.680 das pessoas retratadas têm duas ou mais fotos distintas no conjunto de dados.

* + 1. Base de Dados Utilizada. A base de dados escolhida foi a “Cohn-Kanade”, documentada por Kanade, Cohn e Tian [3]. O banco é composto por frames de vídeos curtos gravados em ambiente controlado pela equipe de pesquisadores, partindo da expressão facial neutra até o ápice de determinada emoção. A base de dados também conta com informações sobre os participantes como, sexo, faixa etária e etnia, além de arquivos em texto contendo a emoção detectada em cada sequência de imagens e suas respectivas AUs, capturadas no momento ápice de cada expressão.

A parte disponível para download do banco de dados conta com imagens de 97 estudantes universitários entre 18 e 30 anos matriculados em aulas de introdução à psicologia da Universidade Carnegie Mellon, e foi previamente preparada por investigadores certificados do FACS.

Além disso, 65% dos participantes são do sexo feminino, 15% afro-americanos e 3% latinos ou asiáticos. As imagens foram capturadas a partir de duas câmeras Panasonic WV3230 conectadas a um gravador de vídeo S-VHS AG-7500 da mesma marca.

Os estudantes foram instruídos pelo condutor da pesquisa a demonstrar determinadas expressões a partir do modelo exibido como exemplo, tratando-se assim de uma base de dados com expressões faciais atuadas, na qual somente a última imagem de cada sequência, contendo o ápice da expressão alvo foi analisada de acordo com as AUs do FACS por investigadores profissionais e categorizada de acordo com as sete emoções disponíveis no banco.

Para este trabalho, foram separadas as sequências de imagens em dois grandes grupos: pessoas do sexo masculino e as pessoas do sexo feminino. Em seguida foram feitas subdivisões dentro de cada grande grupo, separando as sequências de imagens pertinentes a cada emoção, bem como aquelas não classificadas no banco, identificando os nomes dos arquivos que continham a primeira (neutra) e última imagem (ápice) de cada sequência.

* 1. Metodologia Aplicada

A ferramenta desenvolvida tem como objetivo micro expressões faciais para classificá-las de acordo com os estudos do especialista Paul Ekman. Neste artigo, foi desenvolvido o algoritmo que captura pontos da face, identificando AUs e associando-as às suas respectivas emoções a partir dos estudos da EMFACS (INSERIR REFERENÊNCIA).

*2.2.1 Captura de Imagens e Rastreamento das Características Faciais*. O algoritmo trata as imagens identificando os pontos principais do rosto (*facial landmarks*), utilizando a biblioteca Face Recognition capaz de retornar 68 pontos mapeados conforme a Figura 1. Óculos, barbas e demais características que ocultem parte da face devem ser evitados para melhor desempenho.

*2.2.2 Parametrização do Algoritmo.* As primeiras configurações da ferramenta foram feitas utilizando as imagens disponíveis do banco de dados CK, cuja classificação já era conhecida. O algoritmo utiliza a imagem inicial de cada sequência (contendo a expressão neutra) como referência e as compara com a última imagem da sequência (ápice da expressão) para identificar se houve mudanças na aparência da face ocasionadas pela movimentação muscular de cada AU.

Assim, foi necessário primeiramente associar os pares de pontos capturados pela biblioteca com as movimentações musculares previstas pelo FACS, conforme mostrado na tabela 1 a seguir.

Tabela 1: Relação entre AUs e pares pontos capturados pela biblioteca Face Recognition.

| AU | AU Name | Point Pairs Face Recognition | Figure 1 Reference |
| --- | --- | --- | --- |
| 1 | Inner Brow Raiser | Left Eyebrow[4] - Left Eye [3], Right Eyebrow[0] - Right Eye [0] | 22-40, 23-43 |
| 2 | Outer Brow Raiser | Left Eyebrow[1] - Left Eye [3], Right Eyebrow[3] - Right Eye [0] | 19-40, 26-43 |
| 4 | Brow Lowerer | Left Eyebrow[4] - Right Eyebrow[0] | 22-23 |
| 5 | Upper Lid Raiser | Left Eye [2] - Left Eye [4], Right Eye [1] - Right Eye [5] | 39-41, 44-48 |
| 6 | Cheek Raiser | Left Eye [2] - Left Eye [4], Right Eye [1] - Right Eye [5] | 39-41, 44-48 |
| 9 | Nose Wrinkler | Left Eyebrow[4] - Nose Tip[0], Right Eyebrow[0] - Nose Tip[4] | 22-32, 23-36 |
| 10 | Upper Lip Raiser | Top Lip[3] - Nose Tip[2] | 52-34 |
| 12 | Lip Corner Puller | Top Lip[0] - Top Lip[6] | 49-55 |
| 15 | Lip Corner Depressor | Nose Tip[0] - Top Lip[0], Nose Tip[4] - Top Lip[6] | 32-49, 36-55 |
| 17 | Chin Raiser | Bottom Lip[3] - Nose Tip[2] | 58-34 |
| 20 | Lip Stretcher | Bottom Lip[3] - Chin[8] | 58-9 |
| 23 | Lip Tightener | Top Lip[0] - Top Lip[6] | 49-55 |
| 24 | Lip Pressor | Top Lip[3] - Bottom Lip[3] | 52-58 |
| 25 | Lips Part | Top Lip[3] - Bottom Lip[3] | 63-67 |
| 26  27 | Jaw Drop  Mouth Stretch | Top Lip[3] - Bottom Lip[3]  Chin[8] - Nose Tip[2] | 9-34  9-34 |
|  |  |  |  |

O procedimento foi realizado de forma empírica, utilizando como referência a biblioteca animada de AUs disponibilizada [5] associada aos conhecimentos do supracitado sistema (EXPLICAR MELHOR). Tal processo pode ser observado na Figura 2 a seguir, com imagens do banco de dados utilizado neste artigo.

**Expressão Neutra Ápice da Expressão**

Figura 2: Variação de AUs entre dois frames. Autor, 2021.

Para identificar modificações na aparência do rosto foi utilizado um cálculo simples para medir a variação de cada par de pontos selecionado (Equação 1) para representar uma AU. Primeiramente, calcula-se a distância entre os pontos de cada par detectado na expressão neutra. Em seguida o procedimento é repetido para a imagem que contém o ápice da expressão, e, ao final, o resultado numérico obtido para o cálculo do modulo de cada par de pontos na expressão neutra é subtraído do modulo de cada par de pontos no ápice da expressão (Equação 2), possibilitando detectar se houve uma variação positiva (aumento da distância no ápice) ou negativa (diminuição da distância no ápice). Como mostrado em:

(1)

(2)

Na Equação 1, *xo* e *yo* representam as coordenadas do primeiro ponto de cada par de ponto e *x*1 e *y*1 as coordenadas do segundo ponto de cada par de ponto. A variação (Equação 2) representa os valores numéricos finais, os quais foram realizados os cálculos auxiliares (mínimo, máximo, média e mediana) para definir a faixa de valores a ser considerada como relevante para cada emoção.

Uma vez que a saída do algoritmo é o valor numérico referente à variação da distância entre dois pontos, é possível identificar se tais pontos se aproximam ou afastam no momento ápice da expressão, isto é, podem ser associados à impressão visual que se tem quando é vista tal expressão. A exemplo disso, tem-se as sobrancelhas unidas, onde calcula-se a distância entre pontos que representam a parte interna das sobrancelhas e se verifica se a variação diminui.

A partir das medições feitas, foram calculadas métricas estatísticas como mínimos, máximos, médias e medianas para cada uma das AUs capturadas. A partir desses números, foi definida uma faixa de valores dentro da qual estaria a expectativa de modificação de cada AU para cada emoção, baseando-se na relação de emoções com suas respectivas AUs presente na teoria de Paul Ekman [6] [7], explicitadas na Tabela 2.

Tabela 2: Associação das seis expressões faciais com suas AUs e combinações. Ekman et al. (2002)

| Emoção | Indicações Visuais Principais | Indicações Visuais Auxiliares |
| --- | --- | --- |
| Alegria | AU6, AU12 | AU25, AU26, AU16 |
| Tristeza | AU1, AU15, AU17 | AU4, AU7, AU25, AU26 |
| Desgosto | AU9, AU10 | AU17, AU25, AU26 |
| Surpresa | AU5, AU26, AU27, AU1+2 |  |
| Raiva  Medo | AU2, AU4, AU7, AU23, AU24  AU20, AU1+5, AU5+7 | AU17, AU16, AU25, AU26  AU4, AU5, AU7, AU26, AU25 |
|  |  |  |

Primeiramente, o maior número de pares de pontos possíveis para cada AU (Tabela 1) foi considerado em busca de uma maior acurácia nos resultados. Contudo, notou-se que tantos cálculos por segundo culminariam em um atraso na resposta do algoritmo, fator de pouca apreciação para cenários de análise em tempo real. Logo, uma nova associação foi criada, mais refinada e objetiva.

O mesmo foi feito posteriormente quanto ao número de AUs que estavam sendo consideradas para cada emoção, afim de evitar utilizar AUs cujas faixas de valores fossem muito semelhantes em mais de uma emoção. Assim, o refinamento final da utilização das AUs por emoção é exposto na Tabela 3.

Tabela 3: Refinamento Final da Associação das seis expressões faciais com suas AUs e combinações. Autor (2021).

| Emoção | AU por Emoção |
| --- | --- |
| Alegria | AU6, AU12 |
| Tristeza | A1, A15, A17 |
| Desgosto | AU10, AU25, AU9 |
| Surpresa | AU26, AU5 |
| Raiva  Medo | AU4, A24, AU26  AU1, AU26 |
|  |  |

*2.2.3 Algoritmo AFER.* O algoritmo retorna valores numéricos para cada par de pontos relacionados a AUs com valores negativos quando há uma diminuição na distância entre os pontos e valores positivos quando a distância aumenta. Os dados coletados são confrontados com o que se espera que ocorra em cada AU de acordo com a parametrização feita utilizando o banco de dados.

Assim, se a variação de determinadas AUs estiver dentro dos limites pré estabelecidos para cada emoção, o algoritmo irá aferir a emoção referente a tal variação de AUs. Caso contrário, será recebido do algoritmo o resultado de emoção neutra. Desse modo, pode-se observar a dinâmica no algoritmo exposto na Figura 3.

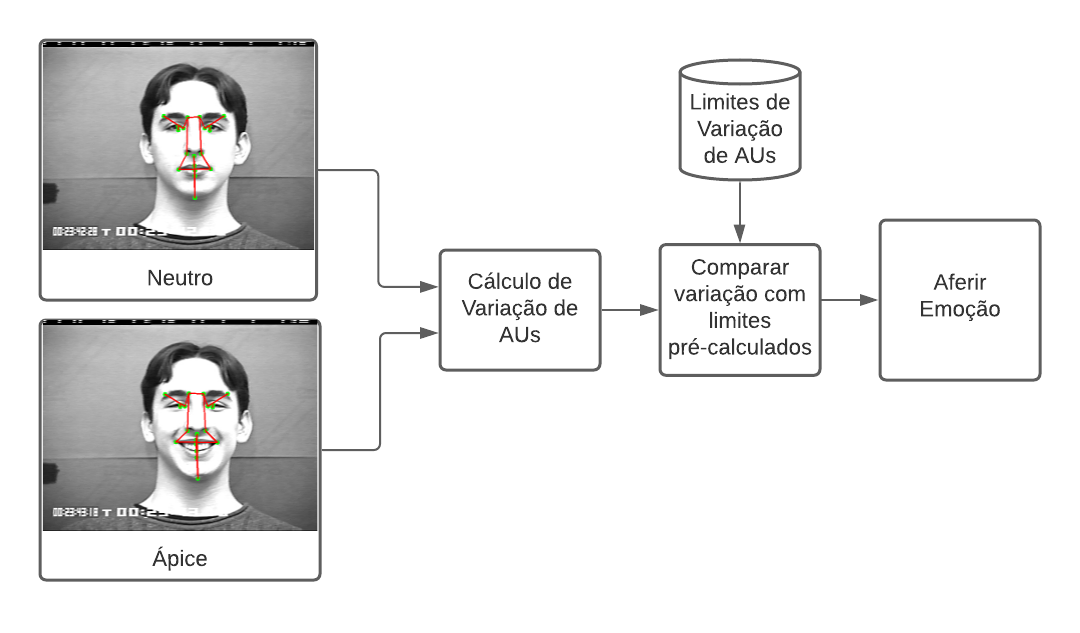


Figura 3: Algoritmo de Captura de Emoção – AFER. Autor, 2021.

1. Resultados

Para este estudo foram feitos dois tipos de testes de verificação de acurácia, sendo somente do gênero masculino. Primeiramente utilizou-se sequências já classificadas no banco utilizado, necessitando somente comparar os resultados esperados. Posteriormente, os testes foram efetuados utilizando sequências que não estavam classificadas no banco de dados, as quais foram submetidas a uma classificação manual pelos autores do presente trabalho.

**3.1 Dados Pré-Classificados**

Neste caso, para o grupo masculino, para um total de 105 sequências no banco, 84 registros de emoções foram aferidos pelo algoritmo, ou seja, uma taxa de captura de 80%, dos quais obteve-se um total de 75 acertos (89,20% de acurácia global) para este gênero. Vale salientar que as taxas de captura e de acerto para cada emoção também são expostas na Tabela 4, verificando-se que, neste caso tivemos uma taxa de acerto de 100% para a emoção “Felicidade”. Precisa fazer a análise das outras emoções. Porquê a taxa de acerto da tristeza foi tão baixa ?

Tabela 4: Resultados de testes utilizando imagens pré-classificadas no banco de dados masculino. Autor (2021).

| Emoção | Taxa de Captura (%) | Taxa de Acerto (%) |
| --- | --- | --- |
| Alegria | 90,91% | 100% |
| Raiva | 80% | 91,7% |
| Desgosto | 70% | 85,70% |
| Surpresa | 80% | 95,80% |
| Tristeza  Medo | 80%  75% | 50%  80,83% |
|  |  |  |

**3.2 Dados Classificados Manualmente**

Algumas imagens no banco de dados não estavam pré-classificadas. Após a classificação manual, considerando o grupo masculino, 53 sequências de imagens foram submetidas ao AFER e, 39 registros de emoções foram aferidos pelo algoritmo, ou seja, uma taxa de captura de 73,58%, dos quais obteve-se um total de 20 acertos (aproximadamente 51,3% de acurácia global) para este gênero. Vale salientar que as taxas de captura e de acerto para cada emoção também são expostas na Tabela 5 a seguir.

Na classificação manual realizada pelos pesquisadores, notou-se que aproximadamente 37,3% aparentavam ser tristeza, mas apresentam uma expressão caricata, talvez por imperícia em movimentar os músculos envolvidos na expressão de tristeza ou desconhecimento de tal expressão por parte dos sujeitos filmados.

Tabela 5: Resultados de testes utilizando imagens classificadas manualmente no banco de dados masculino. Autor (2021).

| Emoção | Taxa de Captura (%) | Taxa de Acerto (%) |
| --- | --- | --- |
| Alegria | 70% | 85,7% |
| Raiva | 50% | 100% |
| Desgosto | 83,33% | 40% |
| Surpresa | 11,11% | 100% |
| Tristeza  Medo | 50%  40% | 14,3%  0,0% |
|  |  |  |

A classificação manual foi feita visando utilizar os dados como conjunto de teste para o algoritmo e, coincidentemente, a emoção em questão foi a que resultou em um dos resultados mais baixos durante a classificação. Na primeira tentativa, foi obtida uma taxa de acerto global de 51,3% e ao descartar os exemplos que continham supostas expressões de tristeza a taxa de acerto subiu para 72%.

1. CONCLUSÃO

(Antes de fazer o apanhado geral sobre os resultados, precisa resgatar o objetivo geral do trabalho.)

Utilizando as técnicas aplicadas e previamente discutidas, o resultado obtido após refinamento de parâmetros pode ser considerado satisfatório, considerando também a taxa de captura calculada, alcançando acurácia de 100% na classificação de expressões de felicidade e acerto de mais de 85% de acerto em outras quatro emoções (tabela 5).

Conforme previsto pela teoria [7], a taxa de acerto nas detecções de expressões de tristeza ficou muito abaixo do desejável, uma vez que não há registrada, cientificamente, movimentações musculares definitivas o bastante para caracterizar tal expressão.

Há poucos exemplos de expressões faciais de tristeza classificadas no banco de dados utilizado [3] em comparação com as outras emoções, e, durante a classificação manual das sequências de imagem não classificadas, notou-se que aproximadamente 37,3% aparentam ser tristeza, mas apresentam uma expressão caricata, talvez por imperícia em movimentar os músculos envolvidos na expressão de tristeza ou desconhecimento de tal expressão por parte dos sujeitos filmados. (Está igual ao trecho dentro no meio do artigo)

As semelhanças entre as expressões de raiva e medo também geram frequente confusão na classificação de expressões do algoritmo, fator que deverá ser aprimorado em trabalhos futuros, bem como o refinamento responsável por melhor diferenciar a AU20 da AU12. (Precisa colocar como trabalhos futuros, inserir nos testes o gênero feminino, realizar testes em tempo real e utilizar aprendizado de máquina para a descoberta de padrões faciais para classificar as expressões faciais nas 7 emoções básicas)

REFERÊNCIAS

1. About OpenCV. 2021. em: <https://opencv.org/about/>
2. Adam Geitgey. 2020. face-recognition 1.3.0. Acesso em dezembro, 2020. Disponível em: https://pypi.org/project/face-recognition/
3. Takeo Kanade, Jeffrey F. Cohn and Yingli Tian. 2000. Comprehensive Database for Facial Expression Analysis, Dept. of Psychology and Robotics Institute, Carnegie Mellon University.
4. Adrian Rosebrock. 2017. Facial landmarks with dlib, OpenCV, and Python. Acesso em junho, 2021. Disponível em: https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/
5. Bryn Farnsworth. 2019. Facial Action Coding System (FACS) – A Visual Guidebook. Acesso em novembro, 2020. Disponível em: https://imotions.com/blog/facial-action-coding-system/
6. Ekman, P., Friesen, W. V. and Hager, J. C.. 2002. Facial Action Coding System – The Manual. Publicado por Research Nexus Division of Network Information Research Corporation. ISBN 0-931835-01-1.
7. Ekman, P. and Friesen, W.V.. 2003. A guide to recognizing emotions from facial expressions. Publicado por Malor Books. ISBN 9781883536367.