

AFER (AFFECTIVE FACIAL EXPRESSION RECOGNITION) - AN ALGORITHM FOR DETECTION AND AFFECTIVE ANALYSIS OF FACIAL EXPRESSIONS IN HUMAN COMPUTER INTERACTIONArisa Oliveira Niwa - UNIVERSIDADE FEDERAL DO PARÁ - Orcid: <https://orcid.org/0000-0002-4338-3174>Fabíola Pantoja De Oliveira Araújo - UNIVERSIDADE FEDERAL DO PARÁ - Orcid: <https://orcid.org/0000-0002-8485-5799>

Thallys Rubens Moreira Costalat -

Sara Das Mercês Silva - UNIVERSIDADE FEDERAL DO PARÁ - Orcid: <https://orcid.org/0000-0003-3436-0787>

This work presents the AFER algorithm to detect and classify emotions through facial expressions, developed by associating computer vision tools (Face Recognition and OpenCV) and psychological studies by Ekman et al. (2002). In this context of technological advancements and the Information Revolution, connecting humans and machines has become imperative. Affective Computing studies show that one of the biggest challenges nowadays is making computers capable of identifying human emotions and, perhaps, understanding them, seeking to strengthen the man-machine relationship. The algorithm captures points on the face, associates them in pairs to represent the muscle movement indicated by each AU, checks if there was movement in such pairs of points to identify whether or not there was visually perceptible change in facial features, to, finally, indicate which emotions was detected. Using computer vision libraries 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. AFER presents a different proposal from other researched algorithms and tools because it was developed based on FACS knowledge and enables the manipulation, treatment and storage of data obtained through the capture of face landmarks, clearly exposing the way in which such processing was conducted. It also encourages the active and inclusive participation of users whose interaction with the computer needs to go beyond the simple mouse, keyboard and monitor setting, especially in with user experience evaluation processes.

Keywords: Affective Computing, Emotions, Face Recognition, Human-Computer Interaction, Computer Vision, Emotion Detection

AFER (AFFECTIVE FACIAL EXPRESSION RECOGNITION) - ALGORITMO PARA DETECÇÃO E ANÁLISE AFETIVA DE EXPRESSÕES FACIAIS NA INTERAÇÃO HUMANO-COMPUTADOR

Este trabalho visa apresentar o algoritmo AFER para detectar e classificar emoções por meio de expressões faciais, desenvolvido associando ferramentas de visão computacional (Face Recognition e OpenCV) a estudos psicológicos (Ekman et al 2002). Nesse contexto de avanços tecnológicos e da Revolução da Informação, conectar humanos e máquinas tornou-se imperativo. Estudos da Computação Afetiva apontam que um dos maiores desafios da atualidade é fazer com que os computadores sejam capazes de identificar as emoções humanas e, quiçá, entendê-las, buscando estreitar a relação homem-máquina. O algoritmo captura pontos da face, associa-os em pares para representar a movimentação muscular indicada por cada AU, e verifica se houve movimentação em tais pares de pontos para identificar ou não movimentação visualmente perceptível nas características do rosto analisado, para, por fim, indicar qual emoção foi detectada. Utilizando bibliotecas de visão computacional implementadas em Python, resultados satisfatórios foram obtidos após ajustes de parâmetro no algoritmo, alcançando reconhecimento de mais de 80% em 5 das 6 emoções analisadas. O AFER apresenta uma proposta diferente dos outros algoritmos e ferramentas pesquisadas por ter sido desenvolvido com base nos conhecimentos do FACS e possibilitar a manipulação, tratamento e armazenamento de dados obtidos através da captura de pontos da face, expondo de forma clara a maneira como tais processamentos foram conduzidos. Além de fomentar a participação ativa e inclusiva de usuários cuja interação com o computador precisa ir além dos simplórios mouse, teclado e monitor, especialmente tratando dos processos de avaliação da experiência do usuário.

Palavras-chave: Computação Afetiva, Emoções, Reconhecimento Facial, Interação Humano-Computador, Visão Computacional, Detecção de Emoções

AFER - Um Algoritmo para Detecção e Análise Afetiva de Expressões Faciais na Interação Humano-Computador (*An Algorithm for Detection and Affective Analysis of Facial Expressions in Human Computer Interaction*)

AFER – *Affective Facial Expression Recognition*

1 INTRODUÇÃO

As ferramentas e aplicações computacionais permitem que as pessoas tenham diferentes experiências ao interagir com as máquinas. Nessa perspectiva, o processo de interação do usuário com a máquina, no qual o ser humano se comunica com a aplicação utilizando a interface, é estudado na área da ciência da computação denominada Interação Humano-Computador (Preece et al., 1994). A IHC estuda os princípios e métodos que possibilitam a construção de equipamentos tecnológicos para o uso humano (Mendes, 2017).

O estudo das emoções no campo computacional, está intrinsecamente ligado à IHC, pois objetiva analisar os usuários enquanto utilizam um software, em busca de um trabalho mais eficiente e uma experiência de usuário cada vez melhor (Picard, 1997). Sendo assim, os seres humanos externalizam o que estão sentindo por meio de manifestações emocionais identificadas por expressões faciais, elementos corporais e voz. No estudo das emoções, existem as consideradas expressões básicas: tristeza, raiva, surpresa, medo, nojo, felicidade e neutro (Ekman, 2003).

Nesse sentido, com base nos estudos do antropólogo Birdwhistell, sobre expressão facial, gestual, postura e movimentos corporais, estima-se que duas pessoas podem transmitir entre si até 10.000 bits de informação por segundo durante uma interação, tornando o canal visual uma fonte rica de informação sobre o estado emocional de um indivíduo. Fator que pode ser aplicado à análise de experiência do usuário a partir da detecção de expressões faciais, utilizando métodos que facilitem a obtenção de informações afetivas sobre os usuários visando o melhor desenvolvimento de hardwares e softwares responsivos a emoções.

O presente trabalho visa apresentar o algoritmo AFER visando detectar e classificar a emoção do usuário a fim de estreitar a relação entre máquina e humano utilizando técnicas de visão computacional associadas aos conhecimentos do FACS (*Facial Action Coding System*) (Ekman, 2002) para implementar um algoritmo que identifica emoções nas expressões faciais com base no mapeamento de pares de pontos associados às AUs (*Action Units*) do mencionado sistema.

1.2 TRABALHOS CORRELATOS

Os estudos no contexto de emoções e afetividade na área da Computação abordam o conceito de Computação Afetiva (CA) que, segundo Picard (1997), estuda como os computadores podem reconhecer, modelar e responder às emoções humanas e, dessa forma, como podem expressá-las através de uma interface ou interação computacional. Há pesquisas e estudos na literatura sobre a importância da interação humano-computador e como ela vem aumentando, exigindo formas de avaliar como o ser humano está reagindo a tal interação, assim como pesquisas sobre o impacto emocional no usuário e como é importante avaliar esses fatores.

No estudo de Costa et al. (2015), os pesquisadores apresentam as definições de experiência de usuário (UX) discutindo a relevância das emoções e afetividade na UX e IHC, sobre a perspectiva de jogos digitais, mostrando que estudar e entender a importância de tais aspectos durante a interação com jogos digitais é extremamente útil e necessário, pois, softwares inseridos na cultura digital com fins recreativos possuem grande influência na vida e no humor das pessoas.

Ainda dentro do contexto de jogos digitais, a pesquisa de Vale *et al.* (2020) avaliou o impacto emocional das limitações de usabilidade em jogos sérios utilizados no ensino de estudantes com necessidades especiais. Os autores destacam que, apesar das barreiras de interação com os jogos, os estudantes com deficiência intelectual externalizaram mais emoções positivas que negativas. Além disso, as emoções positivas estavam intimamente ligadas ao desempenho dos estudantes: quanto mais positiva era a emoção demonstrada pelo estudante, menor era o tempo em que finalizava as atividades do jogo. Já as emoções negativas não influenciaram diretamente no tempo dos estudantes.

A pesquisa de Medonça *et al.* (2018) traz um sistema que recebe informações (imagens e vídeo) através da interação do usuário com o computador e utiliza a técnica de reconhecimento facial e de expressões faciais geradas. Os autores definiram quatro emoções para serem identificadas: felicidade, nojo, raiva e surpresa. Além disso, aplicaram uma Rede Neural Convolutiva utilizando o conjunto de dados aumentado da base JAFFE, para a classificação das emoções. Os resultados alcançados foram relativamente positivos, um acerto de 62%, diante da complexidade de se classificar expressões faciais.

Os trabalhos aqui apresentados destacam a importância de analisar emocionalmente o usuário ao interagir com softwares. Além disso, dois trabalhos destacaram como as emoções influenciam na interação humano-computador. Um estudo apresentou também uma ferramenta de identificação e classificação de emoções por meio de expressão facial. O grande diferencial deste estudo em relação aos apresentados aqui é tornar possível detectar e classificar seis emoções do usuário através da captura, que pode ser feita via sensor de movimentos ou webcam, de pontos da face que possam ser manipulados e armazenados visando a futura criação de um banco de dados contendo a associação entre *facial landmarks* e as AUs do FACS. Este artigo está estruturado da seguinte forma: a Seção 2 apresenta a metodologia do desenvolvimento do sistema; na Seção 3 são apresentados os resultados e, por último, a conclusão desta pesquisa é mostrada na Seção 4.

2 METODOLOGIA

Nesta seção são descritos sucintamente os principais aspectos do FACS (*Facial Action Coding System*), utilizados para associar determinados movimentos de regiões da face humana a emoções específicas de forma sistemática.

Em seguida, serão apresentados os materiais utilizados para o desenvolvimento do presente algoritmo, descrevendo-se as bibliotecas OpenCV, responsável pela manipulação e captura dos *frames* envolvidos neste estudo, e *Face Recognition*, utilizada na detecção de faces em uma imagem, bem como no mapeamento dos pontos da face (*landmarks*). Ambas bibliotecas são utilizadas conjuntamente.

Além disso, é descrito o banco de dados *Cohn-Kanade* utilizado para efetuar os testes preliminares e a parametrização do algoritmo desenvolvido neste trabalho, dando detalhes da distribuição estatística de gênero e de amostras de emoções. Por fim, é explicitada a metodologia desenvolvida para que fossem alcançados os objetivos descritos na introdução.

2.1 Materiais Utilizados

2.1.1 Sistema FACS.

O Sistema de Codificação da Ação Facial (Facial Action Coding System - FACS) é baseado em observadores humanos, concebido por Ekman *et al.* (2002). Tornou-se um padrão no entendimento do comportamento facial, não somente nas pesquisas de ciência comportamental, mas também em outras áreas como análise computacional da face. O FACS é composto por 44 AUs (Action Units – Unidades de Ação relacionadas à atividade muscular facial), das quais 28 estão dispostas nas regiões da testa, sobrancelha, olhos, nariz, bochechas, boca e queixo, dentre as quais, com base na Tabela 1, foram selecionadas as listadas na Tabela 2 para o desenvolvimento do algoritmo de maneira a abranger as AUs consideradas como indicações principais de cada emoção tratada.

Sabendo que o FACS é apenas a catalogação e mapeamento dos músculos, também é necessário conhecer o Emotion FACS (EMFACS) capaz de traçar um paralelo entre as pontuações do FACS e seus respectivos estados emocionais com base nas emoções universais (Ferreira, 2017). Ao EMFACS cabe interpretar as características apontadas na análise facial baseada no FACS a partir de AUs e intensidades específicas a fim de associá-las às emoções. Sendo assim, seguem listadas na Tabela 1 algumas características (AUs) apontadas por Ekman *et al.* (2003) para a expressão facial proveniente de cada emoção.

Tabela 1: Associação das seis expressões faciais com suas AUs e combinações. Ekman et al. (2002)

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

Tabela 2: Refinamento Final da Associação das seis expressões faciais com suas AUs e combinações. Autor (2021).

Emoção	AU por Emoção
Felicidade	AU6, AU12
Tristeza	A1, A15, A17
Nojo	AU10, AU25, AU9
Surpresa	AU26, AU5
Raiva	AU4, A24, AU26
Medo	AU1, AU26

Para o algoritmo desenvolvido, foi necessário definir algumas restrições baseadas na teoria, como por exemplo focar na classificação de expressões de felicidade sem risada envolvida, uma vez que quando tal fenômeno ocorre há uma dificuldade maior em determinar se a expressão é ou não de felicidade. Mesmo na demonstração silenciosa de felicidade, o reconhecimento é simples (Ekman *et al.*, 2003), com exceção apenas para os casos de mistura entre a expressão de felicidade e de outras emoções. Os movimentos musculares necessários para sorrir são muito diferentes dos envolvidos nas expressões de emoções negativas (Ekman *et al.*, 2003), o que justifica o melhor desempenho do algoritmo na classificação desta emoção.

Ademais, optou-se por priorizar a análise de expressões de tristeza ao invés de sofrimento, uma vez que este é exibido de maneira mais óbvia no rosto que aquela. O sofrimento é mais prontamente compreendido, pois é usualmente acompanhado pelo choro. O mesmo não ocorre com a tristeza.que, em sua forma mais extrema, pode ser que não haja nenhum sinal no rosto além da perda de tônus muscular (Ekman *et al.*, 2003), o que não pode ser facilmente detectado por sensores ou câmeras.

2.1.2 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 (About OpenCV, 2021).

Outrossim, a biblioteca *OpenCV* é amplamente utilizada no meio científico para o desenvolvimento de ferramentas de captura de imagens em tempo real com aplicações em reconhecimento e detecção facial, como proposto por Khan et al. no desenvolvimento de um sistema de reconhecimento facial em tempo real através de uma câmera por meio de programação em *OpenCV*, *Haar Cascade*, *Eigenface*, *Fisher Face*, *LBPH* e *Python*.

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.

Neste trabalho a biblioteca foi utilizada para manipulação dos frames, por intermédio de suas funções intrínsecas, a fim de selecioná-los de acordo com as características para uma maior eficiência do algoritmo implementado.

2.1.3 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 capaz de calcular a distância entre os *landmarks* presentes em imagem, a qual servirá como base para aferir mudanças na estrutura facial do usuário.

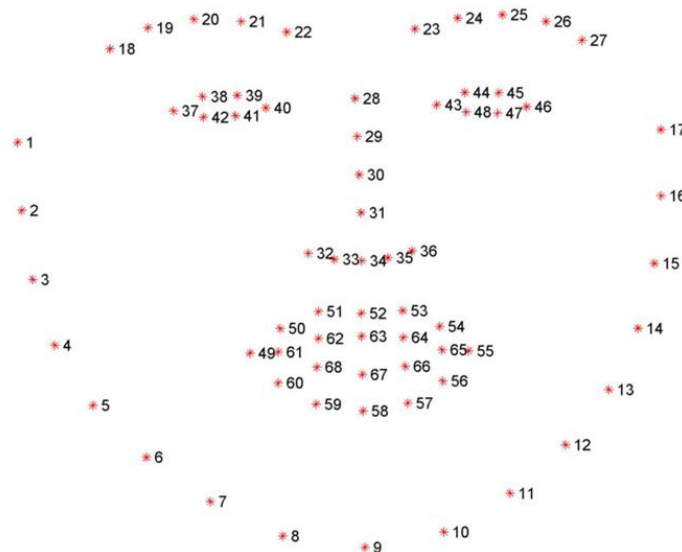


Figura 1: Adaptado de (ROSEBROCK, 2017).

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* (GEITGEY, 2020), um banco de dados de fotos de rosto projetado para estudar o problema de reconhecimento irrestrito de rosto, contendo um conjunto de dados com mais de 13.000 imagens de rostos coletados da *web*. Cada rosto foi etiquetado com o nome da pessoa retratada e, entre as pessoas retratadas, 1.680 têm duas ou mais fotos distintas no conjunto de dados.

2.1.4 Base de Dados Utilizada.

A base de dados escolhida para este trabalho foi a *Cohn-Kanade*, documentada por Kanade, Cohn e Tian (2000). 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.

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.

Outras bases de dados também foram estudadas a fim de verificar se eram adequadas ao propósito da presente pesquisa, como por exemplo a BAUM-1: BAHcesehir University Multimodal Face Database of Affective and Mental States (Zhalehpour *et al.*, 2017). Contudo, o banco de dados escolhido apresentou vantagens cruciais à decisão tendo sido gravado em ambiente controlado, separado em frames indicando as expressões neutras e ápices de expressão, imagens catalogadas por peritos FACS certificados e classificadas com base nas emoções universais, volume de dados considerável e documentação de fácil entendimento.

2.2 Metodologia Aplicada

O AFER tem como objetivo detectar expressões faciais para classificá-las de acordo com os estudos do especialista Paul Ekman. Para isso, o algoritmo captura pontos da face, associa-os em pares para representar a movimentação muscular indicada por cada AU, e verifica se houve movimentação em tais pares de pontos para identificar ou não movimentação visualmente perceptível nas características do rosto analisado, para, por fim, indicar qual emoção foi detectada. As seções seguintes detalham o desenvolvimento do algoritmo.

2.2.1 Rastreamento das Características Faciais.

O algoritmo trata as imagens identificando os pontos principais do rosto (*facial landmarks*) com base nas associações feitas, 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. Tal mapeamento possibilita o dimensionamento de uma faixa limitante de movimentação de AUs para cada uma das emoções capturadas pelo algoritmo.

Vale salientar que os pontos capturados não detectam atividade muscular, mas sim as alterações visuais na aparência do rosto analisado de acordo com o que se espera que ocorra como resultado da movimentação de cada AU. Assim, foi necessário primeiramente associar os

pares de pontos capturados pela biblioteca com as movimentações musculares previstas pelo FACS, relacionando-os com os pontos exibidos na Figura 1, conforme mostrado na Tabela 1.

Tabela 3: Relação entre AUs e pares pontos capturados pela biblioteca *Face Recognition*. (Autor, 2021).

AU	Nome da AU	Pares de Pontos (Face Recognition)	Referenciado à Figura 1
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	Jaw Drop	Top Lip[3] - Bottom Lip[3]	9-34
27	Mouth Stretch	Chin[8] - Nose Tip[2]	9-34

O procedimento foi realizado de forma empírica, utilizando como referência a biblioteca animada de AUs disponibilizada por Farnsworth (2019) associada aos conhecimentos do supracitado sistema, visando associar em pares os pontos mapeados pela biblioteca *Face Recognition* às impressões visuais de movimentação facial que se tem ao ver determinada AU sendo exibida no rosto do observado. Tal processo pode ser observado na Figura 2-a, representando a disposição das AUs na expressão neutra, e na Figura 2-b, representando a disposição das AUs no ápice da expressão “Felicidade”, a partir imagens do banco de dados utilizado neste trabalho.



Figura 2-a: Disposição de AUs durante a expressão “Neutra”. (Autor, 2021).
 Figura 2-b: Disposição de AUs durante o ápice expressão “Felicidade” (Autor, 2021).

2.2.2 Parametrização do Algoritmo.

As primeiras configurações da ferramenta foram feitas utilizando as imagens disponíveis do banco de dados CK, por se tratar de um banco com 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.

Para identificar modificações na aparência do rosto foi utilizado um cálculo simples para medir a variação de cada par de pontos selecionado para representar uma AU. Primeiramente, calcula-se a distância entre os pontos de cada par detectado na expressão neutra (Equação 1), como os cantos da boca (Figura 2-a). Em seguida o procedimento é repetido para a imagem que contém o ápice da expressão (Figura 2-b), e, ao final, o resultado numérico obtido pelo cálculo do módulo da distância de cada par de pontos na expressão neutra é subtraído do módulo da distância 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). As equações 1 e 2 são mostradas a seguir.

$$D = \sqrt{(x_1 - x_0)^2 - (y_1 - y_0)^2} \quad (1)$$

$$Variação = D_{ápice} - D_{neutro} \quad (2)$$

Na Equação 1, x_0 e y_0 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, a partir dos quais foram realizados os cálculos auxiliares (mínimo, máximo, média e mediana) para definir a faixa de valores a ser considerada como relevante, pelo algoritmo, 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 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 (2002) (Tabela 2).

Primeiramente, o maior número de pares de pontos possíveis para cada AU (Tabela 3) 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, retirando-se o primeiro par de pontos da AU6 (39-41).

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.

2.2.3 Algoritmo AFER.

Em suma, após os procedimentos percorridos na seção anterior, o algoritmo retorna valores numéricos para cada par de pontos relacionados a AUs: negativos, quando há uma diminuição na distância entre os pontos, ou positivos quando tal distância aumenta. Os dados coletados são confrontados com o que se espera que ocorra em cada AU de acordo com a parametrização (seção 2.2.2) feita utilizando o banco de dados.

Assim, se a variação de determinadas AUs estiver dentro dos limites pré-estabelecidos para cada emoção, o algoritmo acusará a emoção referente às variações aferidas, caso contrário, o algoritmo não aferirá emoção. Desse modo, pode-se observar a dinâmica no algoritmo exposto na Figura 3.

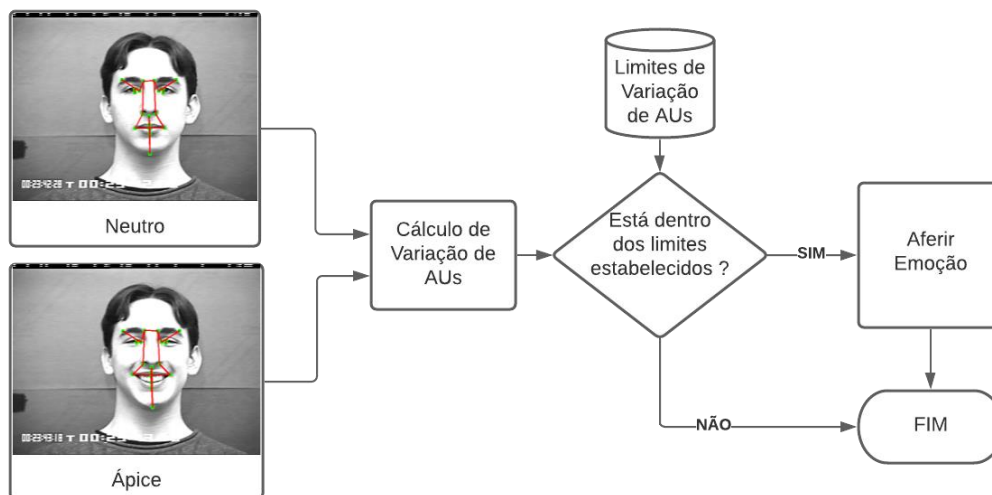


Figura 3: Algoritmo de Captura de Emoção – AFER. (Autor, 2021).

3 RESULTADOS

Para este estudo foram feitos dois tipos de testes de verificação de acurácia, inicialmente com o gênero masculino. Primeiramente utilizou-se sequências já classificadas na base de dados *Cohn-Kanade* (seção 2.1.4), necessitando somente comparar os resultados obtidos com os esperados. Posteriormente, os testes foram efetuados utilizando sequências que não estavam classificadas na base de dados, as quais foram submetidas a uma classificação manual pelos autores do presente trabalho.

3.1 Dados Pré-Classificados

Para os testes preliminares com o gênero masculino, a base de dados possui um total de 105 sequências de imagens, sendo que 84 delas foram identificadas pelo AFER. Isto é, 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 houve uma taxa de acerto de 100% para a emoção felicidade. O menor desempenho foi na detecção da emoção tristeza, fato já esperado tendo em vista que estudiosos da área alegam que essa emoção nem sempre pode ser percebida na face (Ekman *et al.*, 2003).

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

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

3.2 Dados Classificados Manualmente

Conforme relatado anteriormente, parte das imagens no banco de dados não estavam pré-classificadas. Sendo assim, 53 sequências de imagens foram classificadas manualmente pelos pesquisadores deste trabalho e submetidas ao AFER. Nesta classificação, notou-se que aproximadamente 37,3% aparentavam ser tristeza, mas apresentavam uma expressão caricata, talvez por imperícia em movimentar os músculos envolvidos na expressão de tristeza ou desconhecimento de tal expressão por parte das pessoas filmadas.

Após a execução do algoritmo de classificação, 39 registros de emoções foram feitos pelo algoritmo, indicando uma taxa de captura de 73,58%, dos quais obteve-se um total de 20 imagens classificadas corretamente, aproximadamente 51,3% de acurácia global. Vale salientar que as taxas de captura e de acerto para cada emoção também estão expostas na Tabela 5.

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

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

A classificação manual foi feita visando utilizar os dados como conjunto de teste para o algoritmo e, coincidentemente, a emoção tristeza foi a que resultou na mais baixa taxa de acerto, fato já esperado considerando os estudos de Ekman *et al.* (2003) apresentados na (seção 2.1.1). 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%.

A taxa de acerto baixíssima para expressões de medo pode ser associada à confusão do algoritmo com as AU12 e AU20, presentes na felicidade e no medo respectivamente, que movimentam regiões semelhantes do rosto, isto é, os cantos da boca. Fator este que pode ter sido agravado pela menor quantidade de sequências de imagem para essa expressão associada à baixa taxa de captura. Além de que a expressão de medo está muito ligada a questões de sobrevivência, não sendo comum pessoas conseguirem expressá-la, com êxito, de forma voluntária (Ekman *et al.*, 2003), portanto, os próprios peritos envolvidos na classificação de imagens do banco utilizado não as classificaram como expressões de medo.

4 CONCLUSÃO

No presente trabalho, foi apresentado o desenvolvimento do algoritmo AFER, capaz de detectar e classificar emoções a partir das expressões faciais, através do mapeamento de pontos e cálculos implementados na rotina criada. Para tanto, foi necessário relacionar os *landmarks* detectados pela biblioteca *Face Recognition* em pares para representar as movimentações musculares documentadas no FACS, além de realizar cálculos auxiliares, a partir dos dados obtidos por meio do banco de dados, para parametrizar as faixas de valores relevantes para cada emoção a serem consideradas pelo algoritmo.

Vale ressaltar que no atual estágio de desenvolvimento o algoritmo já é capaz de tratar imagens obtidas através de câmeras ou sensores de movimento, possibilitando a captura através de vídeos gravados e futura implementação para análise em tempo real.

Sendo assim, o AFER apresenta uma proposta diferente dos outros algoritmos e ferramentas pesquisadas por ter sido desenvolvido com base nos conhecimentos do FACS e possibilitar a manipulação, tratamento e armazenamento de dados obtidos através da captura de pontos da face, expondo de forma clara a maneira como tais processamentos foram conduzidos ao invés de exibir resultados a partir de algoritmos prontos já disponíveis no mercado.

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 80% em outras quatro emoções (Tabela 4).

Conforme previsto pela teoria (Ekman *et al.*, 2003), a taxa de acerto nas detecções de expressões de tristeza ficou muito abaixo do desejável, uma vez que não há registradas, cientificamente, movimentações musculares definitivas o bastante para caracterizar tal expressão.

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 presentes nas expressões de medo e de felicidade respectivamente. Ademais, pretende-se calibrar o algoritmo para realizar testes com imagens de pessoas com características femininas e utilizar aprendizado de máquina para a descoberta de padrões nas expressões faciais das emoções básicas.

REFERÊNCIAS

- [1] About OpenCV. 2021. em: <https://opencv.org/about/>
- [2] DA COSTA, A. F. et.al. Experiência de usuário e experiência de jogador: discussão sobre os conceitos e sua avaliação no projeto de jogos digitais Proceedings of XIV SBGames pages --10, 2015.
- [3] EKMAN, P. 2003. Emotions Revealed: Recognizing faces and feelings to improve, Times Books.
- [4] EKMAN, P., FRIESEN, W.V. 2003. A guide to recognizing emotions from facial expressions. Publicado por Malor Books. ISBN 9781883536367.
- [5] EKMAN, P., FRIESEN, W. V., 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.
- [6] FRANCO, M. S., RAPOSO, F., VALE, J. V., CARVALHO F., DEMAISON, A. 2020. SOUND DESIGN EM GAMES E UX: ESTUDO DE CASO DA RELAÇÃO ENTRE EFEITOS SONOROS E AÇÕES DE JOGO NO SUPER MARIO BROS, Ergodesign & HCI, volume 8, number 2, 2020, pages 1--26, doi 10.22570/ergodesignhci.v8i2.1460
- [7] FARNSWORTH, B. 2019. Facial Action Coding System (FACS) – A Visual Guidebook. Acesso em novembro, 2020. Disponível em: <https://imotions.com/blog/facial-action-coding-system/>
- [8] GEITGEY, A. 2020. face-recognition 1.3.0. Acesso em dezembro, 2020. Disponível em: <https://pypi.org/project/face-recognition/>
- [9] KANADE, T., COHN, J. F., TIAN, Y. 2000. Comprehensive Database for Facial Expression Analysis, Dept. of Psychology and Robotics Institute, Carnegie Mellon University.
- [10] MENDES, E. 2017. Interação Homem Computador, Universidade Virtual Africana.
- [11] MENDONÇA, T. D. B. 2018. Sistema de reconhecimento de expressões faciais para classificação de emoções de usuários em sistemas computacionais. 33 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade Federal do Ceará, Campus de Russas, Russas.
- [12] PICARD, R. W. 1997. Affective Computing, Mit Press, Cambridge, Ma, Usa.
- [13] PREECE, J., ROGERS, Y., SHARP, E., BENYON, D., HOLLAND, S., CAREY, T. 1994. Human Computer Interaction, Real World Research. Oxford UK: Blackwel.

- [14] ROSEBROCK, A. 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/>
- [15] ZHALEPOUR, S., ONDER, O., AKHTAR, Z., ERDEM, C. E. 2017. 'BAUM-1: A Spontaneous Audio-Visual Face Database of Affective and Mental States', IEEE Transactions on Affective Computing, Vol. 8, No.3.