

UNIVERSIDADE FEDERAL DO PAMPA

Alexia Bagesteiro Camargo

Detecção de Pontos em Faces para
Identificação de Emoções

Alegrete
2022

Alexia Bagesteiro Camargo

Detecção de Pontos em Faces para Identificação de Emoções

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal do Pampa como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Alessandro Bof de Oliveira

Coorientador: Prof. Dr. Anderson Priebe Ferrugem

Alegrete
2022

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

C172d Camargo, Alexia Bagesteiro
Detecção de pontos em faces para identificação de emoções /
Alexia Bagesteiro Camargo.
72 p.

Trabalho de Conclusão de Curso(Graduação)-- Universidade
Federal do Pampa, CIÊNCIA DA COMPUTAÇÃO, 2022.
"Orientação: Alessandro Bof de Oliveira".

1. Detector de Canto de Harris. 2. Detecção de Pontos de
Interesse. 3. Detecção de Emoções em Faces. I. Título.

Alexia Bagesteiro Camargo

Detecção de Pontos em Faces para Identificação de Emoções

Trabalho de Conclusão de Curso apresentado ao Curso de Ciência da Computação da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação.

Trabalho de Conclusão de Curso defendido e aprovado em 9 de março de 2022

Banca examinadora:

Prof. Dr. Alessandro Bof de Oliveira

Orientador

UNIPAMPA

Prof. Dr. Anderson Priebe Ferrugem

Coorientador

UFPEL

Prof. Dr. Mauricio Braga de Paula

UFPEL

Prof. Dra. Leticia Vieira Guimaraes

UERGS



Assinado eletronicamente por **Mauricio Braga de Paula, Usuário Externo**, em 09/03/2022, às 21:48, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **ALESSANDRO BOF DE OLIVEIRA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 09/03/2022, às 21:48, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Anderson Priebe Ferrugem, Usuário Externo**, em 09/03/2022, às 21:48, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



Assinado eletronicamente por **Leticia Vieira Guimarães, Usuário Externo**, em 09/03/2022, às 21:48, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.



A autenticidade deste documento pode ser conferida no site https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0749672** e o código CRC **BEAAE970**.

Este trabalho é dedicado às mulheres que,
contribuem para o campo da ciência e lutam diariamente pela igualdade.

AGRADECIMENTOS

Primeiramente, agradeço aos meus pais pelo incentivo e apoio desde o princípio desta trajetória, acreditando na minha capacidade e atuando como minha principal motivação para que eu nunca desista de lutar pelos meus objetivos.

Agradeço também a todos os professores por compartilharem seus conhecimentos e experiências que foram de extrema importância para minha formação, em especial ao prof. Dr. Alessandro Bof de Oliveira por ter aceitado o convite para orientar-me durante este Trabalho de Conclusão de Curso e ao prof. Dr. Anderson Priebe Ferrugem por sua colaboração como coorientador.

Para finalizar, agradeço aos familiares, colegas e amigos que estiveram presentes, proporcionando-me lembranças que levarei com carinho por toda a minha vida.

“We shall do a much better programming job,
provided we approach the task with a full appreciation of its tremendous difficulty,
provided that we stick to modest and elegant programming languages,
provided that we respect the intrinsic limitations of the human mind and approach the
task as Very Humble Programmers.”
(Edsger Dijkstra - 1930–2002).

RESUMO

As expressões faciais atuam como um meio de comunicação não verbal capaz de refletir o estado emocional do ser humano, e através desta compreensão podemos verificar a coerência entre o que é dito e o que é manifestado. Cada emoção básica transmitida por um indivíduo possui uma movimentação muscular facial, portanto, uma técnica chamada *Facially Expressed Emotion Labeling (FEEL)* costuma ser utilizada por pesquisadores, com o intuito de avaliar a capacidade das pessoas em compreender os sentimentos de outras por meio de suas expressões, consistindo na apresentação da imagem de um rosto em uma tela de computador, na sequência é apresentado esse mesmo rosto transmitindo uma das seis emoções consideradas básicas: alegria, tristeza, medo, raiva, nojo ou surpresa. Ao final, os participantes devem dizer qual sentimento foi expresso na imagem, porém nem todos os sentimentos são percebidos automaticamente com a mesma facilidade. Diante disso, há interesse na implementação de Interfaces Homem-Máquina (IHM) capazes de identificar a expressão de um indivíduo, auxiliando na resolução deste problema. Na área da robótica, já existem robôs capazes de identificar e replicar expressões faciais semelhantes as do ser humano com ajuda da Inteligência Artificial (IA). Algumas técnicas computacionais como algoritmos de detecção de canto são utilizadas para identificar pontos de interesse em imagens, dentre os mais conhecidos podemos mencionar o Detector de Canto de Harris, utilizado neste trabalho com o objetivo de analisar e identificar os melhores parâmetros para identificação de emoções em imagens de faces. Para realizar a análise proposta, serão utilizadas as bases de dados *The Karolinska Directed Emotional Faces (KDEF)*, *Japanese Female Facial Expression (JAFFE)* e CK+, juntamente com Rede Neural Convolutacional (RNC) e filtro de Gabor.

Palavras-chave: Detector de Canto de Harris. Detecção de Pontos de Interesse. Detecção de Emoções em Faces.

ABSTRACT

Facial expressions act as a non-verbal means of communication capable of reflecting the emotional state of the human being, and through this understanding we can verify the coherence between what is said and what is manifested. Each basic emotion transmitted by an individual has facial muscle movement, therefore, a technique called *Facially Expressed Emotion Labeling (FEEL)* is often used by researchers, in order to assess people's ability to understand the feelings of others through their expressions, consisting of the presentation of the image of a face on a computer screen, in the sequence that same face is presented transmitting one of the six emotions considered basic: joy, sadness, fear, anger, disgust or surprise. At the end, the participants must say which feeling was expressed in the image, but not all feelings are automatically perceived with the same ease. Therefore, there is interest in the implementation of *Human-Machine Interfaces (HMI)* capable of identifying the expression of an individual, helping to solve this problem. In the area of robotics, there are already robots capable of identifying and replicating facial expressions similar to those of humans with the help of *Artificial Intelligence (AI)*. Some computational techniques such as corner detection algorithms are used to identify points of interest in images, among the best known we can mention the *Harris Corner Detector*, used in this work with the aim of analyzing and identifying the best parameters for identifying emotions in faces images. To perform the proposed analysis, the databases *The Karolinska Directed Emotional Faces (KDEF)*, *Japanese Female Facial Expression (JAFFE)* and *CK+* will be used, along with *Convolutional Neural Network* and *Gabor filter*.

Key-words: Harris Corner Detector. Interest Point Detection. Emotion Detection in Faces.

LISTA DE FIGURAS

Figura 1 – Fluxo do processo.	31
Figura 2 – Imagens de amostra da base de dados JAFFE. Da esquerda para a direita: Raiva, nojo, medo, felicidade, tristeza, surpresa e neutro. . . .	35
Figura 3 – Exemplo de cada emoção do conjunto de estímulos KDEF.	36
Figura 4 – Exemplo das classes de expressão da base de dados CK+ 48.	37
Figura 5 – Enquadramento de algumas das unidades de ação definidas no FACS. .	38
Figura 6 – Expressões Faciais.	39
Figura 7 – Significados matemáticos dos autovalores λ_1 e λ_2	40
Figura 8 – Representação da função resposta R.	41
Figura 9 – Tipos de recursos do Classificador em Cascata de Haar.	42
Figura 10 – Estrutura em cascata para classificadores Haar.	42
Figura 11 – Exemplo de comando do <i>OpenCV</i>	42
Figura 12 – Função Harris e Gabor.	44
Figura 13 – Gerador de dados e preparação dos iteradores na base dados Jaffe. . .	45
Figura 14 – RNC e suas camadas.	46
Figura 15 – Construção do Modelo de Rede Neural Convolutacional (RNC).	47
Figura 16 – Estrutura geral do modelo.	48
Figura 17 – Exemplo de imagem com pontos detectados pelo Detector de Canto de Harris e com filtro de Gabor sobreposto.	51
Figura 18 – Gráfico da acurácia do modelo na base de dados JAFFE para $k=0.04$. .	52
Figura 19 – Gráfico da perda do modelo na base de dados JAFFE para $k=0.04$. .	53
Figura 20 – Avaliação de desempenho do reconhecimento de emoções na base de dados JAFFE para $k=0.04$	53
Figura 21 – Matriz de Confusão JAFFE para $k=0.04$	54
Figura 22 – Gráfico da acurácia do modelo na base de dados KDEF para $k=0.05$. .	55
Figura 23 – Gráfico da perda do modelo na base de dados KDEF para $k=0.05$. .	55
Figura 24 – Avaliação de desempenho do reconhecimento de emoções na base de dados KDEF para $k=0.05$	56
Figura 25 – Matriz de Confusão KDEF para $k=0.05$	56
Figura 26 – Gráfico da acurácia do modelo na base de dados CK+ 48 para $k=0.04$. .	57
Figura 27 – Gráfico da perda do modelo na base de dados CK+ 48 para $k=0.04$. .	58
Figura 28 – Avaliação de desempenho do reconhecimento de emoções na base de dados CK+ para $k=0.04$	58
Figura 29 – Matriz de Confusão CK+ 48 para $k=0.04$	59

LISTA DE TABELAS

Tabela 1 – Características das Bases de Dados Utilizadas	34
Tabela 2 – Quantidade de imagens por diretório	45
Tabela 3 – Métricas de Avaliação de Desempenho	49
Tabela 4 – Acurácias e Perdas JAFFE (treinamento e teste)	52
Tabela 5 – Acurácias e Perdas KDEF (treinamento e teste)	54
Tabela 6 – Acurácias e Perdas CK+ 48 (treinamento e teste)	57

LISTA DE SIGLAS

ASM Active Shape Model

FACS Facial Action Coding System

FAST Features from Accelerated Segment Test

HMM Hidden Markov Model

HOG Histogram of Oriented Gradients

IA Inteligência Artificial

ICA Independent Component Analysis

IHC Interação Humano-Computador

LDA Linear Discriminant Analysis

RAM Random Access Memory

RNC Rede Neural Convolutacional

SURF Speed-Up Robust Features

SUSAN Smallest Univalued Segment Assimilating Nucleus

SVM Support Vector Machine

UA Unidades de Ação

LISTA DE SÍMBOLOS

α	Constante alfa
λ	Autovalor
\mathcal{T}	Conjunto de trabalhos de TCC
Σ	Somatório

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Objetivos Específicos	26
1.2	Organização deste trabalho	26
2	TRABALHOS RELACIONADOS	27
3	DESENVOLVIMENTO	31
4	METODOLOGIA	33
4.1	Ambiente de Desenvolvimento	33
4.2	Bases de Dados Utilizadas	34
4.2.1	Japanese Female Facial Expression (JAFFE)	34
4.2.2	Karolinska Directed Emotional Faces (KDEF)	35
4.2.3	CK+ 48	36
4.3	Detector de Canto de Harris	37
4.4	Classificador em Cascata de Haar	41
4.5	Filtro de Gabor	43
4.6	Preparação dos Dados	44
4.7	Rede Neural Convolucional	45
4.8	Métricas de Avaliação de Desempenho	48
5	RESULTADOS	51
6	CONSIDERAÇÕES FINAIS	61
6.1	Conclusão	61
6.2	Trabalhos Futuros	61
	REFERÊNCIAS	63

1 INTRODUÇÃO

A detecção de emoções humanas a partir de expressões faciais é um processo utilizado em diversas áreas, como segurança, cibernética, processamento de imagens, robótica, realidade virtual, Interação Humano-Computador (IHC), psicologia, neurologia, marketing social, dentre outras, auxiliando na identificação de pistas de comunicação não-verbal, envolvendo o reconhecimento de padrões e regularidades em um conjunto de dados a ser analisado.

"A análise automatizada de emoções permite ter uma ideia aproximada dos sentimentos que uma pessoa está vivenciando, mesmo quando a própria pessoa não está completamente ciente deles." (ÁLVAREZ et al., 2018, p.1). De acordo com (HUSSAIN, 2013), o aprimoramento da comunicação entre o ser humano e o computador seria capaz de abrir uma variedade de possibilidades, como o desenvolvimento de dispositivos que alertam um motorista sonolento com o intuito de evitar acidentes de trânsito, tentam acalmar um cliente irritado, permitem a detecção da condição psicológica de um criminoso e podem determinar a agonia de um paciente, alertando o médico ou enfermeiro para ajudá-lo rapidamente em caso de emergência.

Grande parte dos estudos realizados nesta área são baseados na metodologia científica elaborada pelo psicólogo americano Paul Ekman e Wallace Friesen (EKMAN; FRIESEN, 1971), denominada *Facial Action Coding System (FACS)*, que possibilita a análise das emoções exibidas na face, diferenciando-as de acordo com o movimento da musculatura facial, por meio da codificação de *Unidades de Ação (UA)* (FRIESEN; EKMAN, 1978). Deste modo, existem pontos-chave faciais que são importantes durante a análise da emoção expressa através da face, que atualmente podem ser facilmente detectados por meio da tecnologia, com o auxílio de algoritmos de detecção de canto bastante conhecidos como *Detector de Canto de Harris*, *Smallest Univalve Segment Assimilating Nucleus (SUSAN)*, *Features from Accelerated Segment Test (FAST)*, *Moravec*, entre outros.

Dentro desse contexto, inúmeros trabalhos tem sido desenvolvidos aplicando técnicas de Inteligência Artificial (IA), *Deep Learning*, redes neurais e classificadores de Aprendizado de Máquina como máquinas de vetores de suporte, para realizar reconhecimento facial e identificar emoções universais básicas expressas pelo ser humano como alegria, tristeza, raiva, surpresa, nojo, medo e desprezo (LISSETTI; RUMELHART, 1998) (MA; KHORASANI, 2004), (RAJESH; NAVEENKUMAR, 2016), (MEHENDALE, 2020) e (ALI; KHATUN; TURZO, 2020).

O objetivo deste trabalho é analisar e identificar os melhores parâmetros para o detector de canto de Harris na identificação de emoções em imagens de faces.

1.1 Objetivos Específicos

- Para cada grupo de emoções, aplicar o detector de canto de Harris com um determinado parâmetro.
- Verificar os valores do parâmetro que apresentam melhor desempenho no reconhecimento de emoções.

1.2 Organização deste trabalho

Este trabalho apresenta-se organizado como segue. No capítulo 1 é apresentada a introdução do tema, bem como o objetivo e a proposta apresentada no trabalho. No capítulo 2 são apresentados os trabalhos relacionados, no capítulo 3 é apresentado o desenvolvimento, no capítulo 4 são apresentados os métodos utilizados. Os resultados e considerações finais são apresentados nos capítulos 5 e 6 respectivamente.

2 TRABALHOS RELACIONADOS

Este capítulo irá discutir de forma sucinta os trabalhos desenvolvidos sobre reconhecimento de expressão facial e avaliação de detectores de canto, bem como sua utilização para detectar pontos de interesse em faces para identificação de emoções. Os trabalhos estão organizados de acordo com o ano de publicação.

Antonini, Popovici e Thiran (ANTONINI; POPOVICI; THIRAN, 2003) propõem a representação da *Independent Component Analysis (ICA)* e *Support Vector Machine (SVM)* para extrair características faciais em um contexto de detecção e localização de rosto, com o objetivo de encontrar um espaço melhor para projetar os dados, a fim de construir diferentes classificadores de características faciais que sejam resistentes a variações de iluminação e condições desfavoráveis de aquisição de imagens utilizando o detector de canto de Harris como etapa de pré-filtragem. O método foi testado na base de dados BANCA em diferentes cenários, concluindo que o detector de canto pode ser ajustado para selecionar cantos suficientemente próximos das posições reais.

Utilizando *Active Apperarance Models (AAM)* para selecionar a melhor hipótese de detecção e localização de faces, Hamouz (HAMOUZ, 2004) desenvolveu um algoritmo com o objetivo de localizar rostos com precisão em imagens estáticas em tons de cinza, avaliando o método preconizado nas bases de dados XM2VTS, BANCA e BioID. Logo, comprovou-se que ao utilizar a verificação de hipóteses proposta, um aumento significativo no desempenho foi obtido em comparação com o desempenho dos detectores de características isoladamente.

Uma solução baseada em rede neural combinada com processamento de imagem para classificação das emoções universais foi proposta utilizando o algoritmo de detecção de canto de Harris. Widanagamaachchi (WIDANAGAMAACHCHI, 2009) concluiu que o sistema proposto baseado em rede neural é muito promissor e fornece melhor precisão no reconhecimento de emoções do que um classificador Bayesiano, obtendo uma taxa de 67,35% de acerto ao identificar a emoção correta e 50% de acerto mesmo quando as características são classificadas incorretamente.

Um novo esquema de seleção de marcos *Active Shape Model (ASM)* para melhorar o desempenho em aplicações de reconhecimento facial é proposto por (BEHAINE; SCHARCANSKI, 2012). O esquema seleciona pontos de referência robustos onde as características faciais relevantes são encontradas, atribuindo pesos maiores às suas características correspondentes na etapa de classificação da face. Os resultados apresentaram a melhoria alcançada pelo método proposto em comparação com o ASM e outros métodos. Logo, foi concluído que o desempenho do ASM pode ser melhorado utilizando a proposta do ASM aprimorado, que obteve 95,33% de acurácia na base de dados Essex Face.

Zhao (ZHAO, 2012) apresenta uma proposta de reconhecimento da emoção humana a partir de informações da linguagem corporal da cabeça para reconhecer os estados emocionais e cognitivos utilizando técnicas de *Soft Computing*, com o objetivo de investi-

gar novos métodos de IHC. Posteriormente, todas essas informações são combinadas por técnicas de *Soft Computing* para inferir os estados emocionais implícitos, utilizando projeção integral de várias etapas para segmentação de imagem, filtro de Gabor para detecção de características, e por fim, um algoritmo de aprendizagem de máquina *SVM* é aplicado para classificar expressões faciais em seis categorias básicas de emoção: felicidade, tristeza, medo, raiva, surpresa e nojo.

Uma solução para identificação de emoções mediante expressões faciais em imagens de webcam foi desenvolvida por Rosa e Bueno (ROSA; BUENO, 2013), aplicando técnicas de aprendizado de máquina utilizando Análise de Componentes Principais (PCA) e realizando experimentos nas bases de dados CAFE, JAFFE e FACES. Logo, foi desenvolvida a interface *web* utilizando a linguagem JAVA/JSP para a captura da imagem da face do usuário por meio de *webcam*, utilizando a tecnologia Flash. Após a aquisição da imagem, foi utilizado o algoritmo Viola-Jones para detecção de face e características de *Haar*, aplicando o filtro de Sobel sobre a imagem já equalizada. Constatou-se a superioridade de desempenho do algoritmo com a base CAFE diante das emoções de alegria (100%), tristeza (12%), raiva (37%), neutro (50%) e as quatro juntas (50%).

O trabalho de Ramos (RAMOS, 2013) apresenta um novo método para cálculo da pose da cabeça em imagens monoculares. Foram comparadas as posições das características faciais específicas com as de múltiplas instâncias do modelo da face em 3D, obtendo os pontos característicos nas imagens de entrada por meio do algoritmo *ASM*, utilizando a base de dados Pointing04. Os resultados apresentados mostraram-se promissores, estimando erros pequenos em relação aos outros métodos propostos para imagens monoculares.

Com o objetivo de detectar pontos característicos como olhos, nariz e boca de imagem térmica facial em diferentes condições como variação de iluminação, rotação e escala, Bhowmik, Shil e Saha (BHOWMIK; SHIL; SAHA, 2013) propuseram a extração de pontos de característica em face térmica utilizando o detector de canto de Harris. Foi utilizado o conjunto de imagens da base de dados *IRIS (Imaging, Robotics and Intelligent System)* com resolução de 320x240 pixels no banco de dados *OCTBVS (Object Tracking and Classification Beyond Visible Spectrum)*, comprovando que o detector de Harris é capaz de detectar um bom número de pontos de interesse sobre os olhos e região do nariz, mesmo se as imagens forem rotacionadas em 30 graus ou reduzidas com um fator de escala de 0.5, também apresentando resultados satisfatórios na imagem com mudança na variação de iluminação.

Kesarwani, Choudhary e Misra (KESARWANI; CHOUDHARY; MISRA, 2013) propõem um algoritmo adaptativo para extrair pontos de características faciais automaticamente, como cantos de sobrancelhas, cantos de olhos, narinas, ponta do nariz e cantos da boca em faces de visão frontal, com base no algoritmo de Harris, utilizando o detector de rosto Viola-Jones para detectar a localização do rosto e recortar essa região em

uma imagem. O método foi testado em um grande banco de dados de face frontal BioID no MATLAB com diferentes iluminações, expressões e condições de iluminação. Os resultados experimentais alcançaram uma média de taxa de sucesso de 96,15%.

Planejando avaliar a aceitação pública de produtos com base em sua marca, analisando a expressão facial do cliente, Chaugule et al. (CHAUGULE et al., 2016) realizou uma avaliação de produto com base na detecção de expressão facial otimizada. A abordagem proposta pega um vídeo em tempo real ou vídeo gravado como entrada e extrai quadros em intervalos regulares. Posteriormente, é realizada a detecção de rosto utilizando o algoritmo Viola-Jones, seguido da extração de pontos característicos com o auxílio de um algoritmo de Harris modificado, expressando em gráficos uma crítica positiva ou negativa para o produto. O algoritmo provou ser significativamente mais rápido e quase preciso para a aplicação necessária, reduzindo a complexidade de tempo para detecção de pontos de canto de $O(n^2 \log(MAX))$ para $O(n^2)$.

Para realização da tarefa de reconhecimento de expressão facial em sequências de vídeo, Kamarol et al. (KAMAROL et al., 2016) propõem uma nova abordagem para a extração de traços faciais com base na aparência adotando um método baseado em bloco. Neste trabalho foi utilizado o detector de rosto Viola-Jones e um *Spatio-Temporal Texture Map (STTM)* que utiliza a informação espaço-temporal extraída da função de canto de Harris tridimensional. Foi comprovado que o método proposto mostra um desempenho superior em comparação com as abordagens de última geração, com uma taxa média de reconhecimento de 95,37%, 98,56% e 84,52% nas bases de dados CK+, CASME II e AFEW com menor custo computacional.

Utilizando uma combinação da detecção de Viola-Jones com o algoritmo de detecção de cantos de Harris, *Haar-like features* e um sub-algoritmo denominado *Iterative Best Fit Algorithm*, Gondhi et al. (GONDHI et al., 2017) desenvolveram um algoritmo eficiente de detecção de pontos de referência facial para encontrar um ponto de referência explorando sua semelhança e configuração geométrica. O método foi testado na base de dados JAFEE, obtendo uma taxa de precisão é de 93%.

Verma e Parashar (VERMA; PARASHAR, 2018) apresentam uma avaliação de algoritmos de detecção de canto para modelagem de emoções humanas. Foram avaliados os algoritmos SUSAN, Harris, Moravec e FAST, juntamente com sua proposta de abordagem de força bruta utilizando Curvas de Bézier, seguindo as etapas de segmentação da cor da pele e extração de características faciais. A comparação foi baseada em quanto tempo cada algoritmo levou para detectar os cantos nas características faciais do rosto humano frontal em imagens 2D que exibiam quatro emoções (neutro, feliz, triste e chocado), concluindo que a abordagem proposta foi mais eficiente em termos de custo computacional em comparação com outros algoritmos para a finalidade da aplicação e para o conjunto de imagens consideradas.

Uma implementação e um estudo completo do detector de cantos Harris é apre-

sentado por Sánchez, Monzón e Nuez (SÁNCHEZ; MONZÓN; NUEZ, 2018). Foi implementado um algoritmo de supressão não máxima genérico que permite selecionar as características proeminentes na imagem, comparando o uso de interpolação quadrática e quártica para obter pontos de interesse com precisão de subpixel, concluindo que a abordagem quadrática é muito mais rápida e simples de implementar, e por fim, identificando a necessidade de implementação de uma técnica de convolução gaussiana mais rápida ou substituição por um filtro de caixa, com a finalidade de melhorar a velocidade do método.

Rashmi, Saini e Bagchi (RASHMI; SAINI; BAGCHI, 2018) propõem um novo modelo de reconhecimento de emoção facial humana utilizando o detector de canto de Harris e *SVM*, adotando o método proposto por Viola-Jones. A abordagem utiliza uma janela deslizante para pesquisar os pontos dos cantos, examinando as alterações de gradiente ao deslizar pela área, resultando na detecção de características e classificação de expressões faciais em cinco emoções básicas, sendo elas: neutro, felicidade, ansiedade, nojo e medo. O algoritmo proposto e implementado apresentou maior rapidez e resultados muito melhores para a extração de emoções do rosto humano.

O trabalho de Priya (PRIYA, 2018) apresenta uma proposta que pode ser utilizada para determinar os pontos-chave dos lábios em rostos. Foi desenvolvido um algoritmo que detecta a região da boca utilizando o método de Viola-Jones implementado no MATLAB, com o auxílio de alguns métodos como equalização de histograma, método de Otsu, ajuste de elipse e detecção de pontos de interesse utilizando o detector de cantos de Harris. Durante a análise de precisão das expressões de surpresa, raiva, felicidade e neutro em diferentes classificadores, a técnica do classificador logístico apresentou o melhor resultado com 95% de precisão.

Aplicando pré-processamento de imagens para redução de ruído e o algoritmo de Viola-Jones para extração dos rostos, foi elaborado um *framework* para reconhecimento de expressão facial utilizando algoritmos de extração de características feitas à mão e codificação de recurso supervisionado, comparando o desempenho de diferentes métodos de extração de características como *Speed-Up Robust Features (SURF)*, *FAST*, *Histogram of Oriented Gradients (HOG)* e detecção de cantos de Harris. O método proposto por Noor et al. (NOOR et al., 2020) foi avaliado na base de dados CK+ e a análise comparativa dos resultados obtidos mostrou que *HOG* apresenta a maior precisão de 90,79%.

3 DESENVOLVIMENTO

Neste TCC é documentada uma proposta de detecção de pontos em faces para identificação de emoções utilizando o detector de canto de Harris.

Para a realização dos testes apresentados no trabalho foram escolhidas três bases de dados, a *Karolinska Directed Emotional Faces (KDEF)*, a *Japanese Female Facial Expression (JAFFE)* e a *Extended Cohn-Kanade (CK+)*, das quais foram selecionadas imagens frontais de faces humanas expressando emoções distintas como raiva, medo, nojo, felicidade, tristeza, surpresa e neutro.

O detector de canto de Harris foi utilizado para extrair pontos de características faciais relevantes presentes nas imagens, como por exemplo cantos das sobrancelhas, da boca, dos olhos, narinas e ponta do nariz. Durante este processo, o parâmetro k foi variado de 0 até 0.2, com o objetivo de encontrarmos o valor de melhor desempenho na identificação de emoções em cada uma das bases de dados.

Posteriormente, foi aplicado o filtro de Gabor para ressaltar as informações nas regiões dos pontos detectados pelo detector de Harris. Essa informação foi combinada com a imagem original. Em seguida, foi utilizada uma Rede Neural Convolutiva e efetuada a classificação. O fluxo do processo realizado é mostrado na Figura 1.

Figura 1 – Fluxo do processo.



4 METODOLOGIA

Este capítulo apresenta os aspectos metodológicos a serem seguidos para o desenvolvimento do trabalho, que tem como objetivo aplicar o detector de canto de Harris, presente na biblioteca de processamento de imagens denominada *scikit-image*¹, com o parâmetro k utilizando a linguagem de programação Python 3², a fim de analisar e identificar os melhores valores de k para o detector de canto de Harris na identificação de emoções em imagens de faces.

A seção 4.2 descreve brevemente as bases de dados escolhidas, a seção 4.3 apresenta o Detector de Canto de Harris, a seção 4.4 apresenta o Classificador em Cascata de Haar, a seção 4.5 apresenta o Filtro de Gabor, a seção 4.6 aborda a preparação dos dados, a seção 4.7 descreve a Rede Neural Convolutiva (RNC), e por fim, as métricas de avaliação de desempenho utilizadas são apresentadas na seção 4.8.

4.1 Ambiente de Desenvolvimento

O ambiente de desenvolvimento utilizado foi o *Google Colaboratory*³, que permite a escrita e a execução de código Python arbitrário por meio do navegador e é especialmente adequado para aprendizado de máquina, análise de dados e educação. Mais tecnicamente, o *Colab* é um serviço de notebook Jupyter hospedado que não requer configuração para uso, enquanto fornece acesso gratuito a recursos de computação (RESEARCH, 2017). O ambiente disponibiliza as seguintes especificações técnicas:

- 12 GB de Random Access Memory (RAM).
- 107 GB de espaço em disco.
- Limite máximo de 12 horas por sessão.
- Linguagem de programação: Python 3.
- Bibliotecas e pacotes utilizados:
 - *scikit-image*: Coleção de algoritmos para processamento de imagens (WALT, 2021).
 - *TensorFlow*: Biblioteca de código aberto para aprendizado de máquina (TENSORFLOW, 2021).⁴
 - *Keras*: API de *deep learning* escrita em Python, executada na plataforma de aprendizado de máquina *TensorFlow* (KERAS, 2020).⁵

¹ <<https://scikit-image.org/>>

² <<https://www.python.org/>>

³ <https://colab.research.google.com/>

⁴ <https://www.tensorflow.org/>

⁵ <https://keras.io/>

- *OpenCV*: Biblioteca de software de visão computacional e aprendizado de máquina de código aberto (OPENCV, 2022).⁶
- *Numpy*: Pacote fundamental para computação científica em Python, utilizado para trabalhar com arranjos, listas e matrizes de forma mais eficiente (WALT; COLBERT; VAROQUAUX, 2011).⁷
- *scikit-learn*: Biblioteca da linguagem Python desenvolvida especificamente para aplicação prática de aprendizado de máquina. Dispõe de ferramentas simples e eficientes para análise preditiva de dados (PEDREGOSA et al., 2011)⁸.
- *matplotlib*: Biblioteca abrangente para criar visualizações estáticas, animadas e interativas em Python (MATPLOTLIB, 2021).⁹

4.2 Bases de Dados Utilizadas

Para a realização dos testes realizados neste trabalho, foram utilizadas três bases de dados, sendo elas: *Japanese Female Facial Expression (JAFFE)*, *Karolinska Directed Emotional Faces (KDEF)* e *CK+ 48*, cujas principais características são apresentadas na Tabela 1.

Tabela 1 – Características das Bases de Dados Utilizadas.

Nome da base de dados	Quantidade de imagens	Resolução	Formato
JAFFE	213	256x256	Tiff
KDEF	4900	562x762	JPEG
CK+ 48	981	48x48	PNG

As demais informações são descritas detalhadamente nas subseções seguintes.

4.2.1 Japanese Female Facial Expression (JAFFE)

O dataset *Japanese Female Facial Expression (JAFFE)*¹⁰ consiste em 213 imagens de 10 mulheres japonesas exibindo 6 expressões faciais diferentes e uma neutra, conforme mostrado na Figura 2. As imagens possuem escala de cinza de 8 bits no formato Tiff e resolução 256x256 pixels, além de uma média de avaliações semânticas em cada expressão facial avaliada por 60 espectadores japoneses.

⁶ <https://opencv.org/>

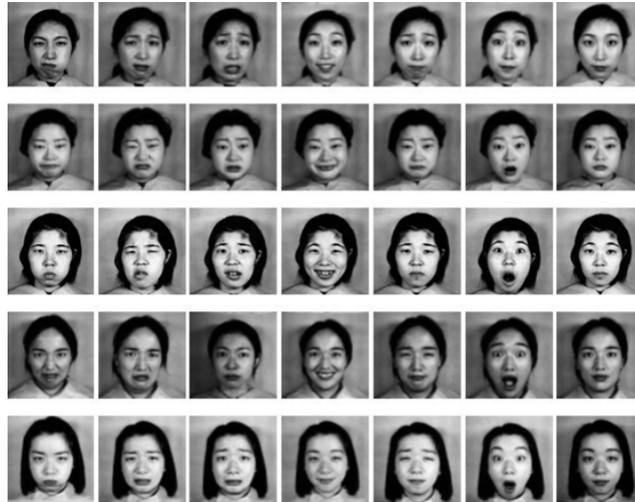
⁷ <https://numpy.org/>

⁸ <https://scikit-learn.org/>

⁹ <https://matplotlib.org/>

¹⁰ <<https://zenodo.org/record/3451524>>

Figura 2 – Imagens de amostra da base de dados JAFFE. Da esquerda para a direita: Raiva, nojo, medo, felicidade, tristeza, surpresa e neutro.



Fonte – Ji e Idrissi (2009, p. 6)

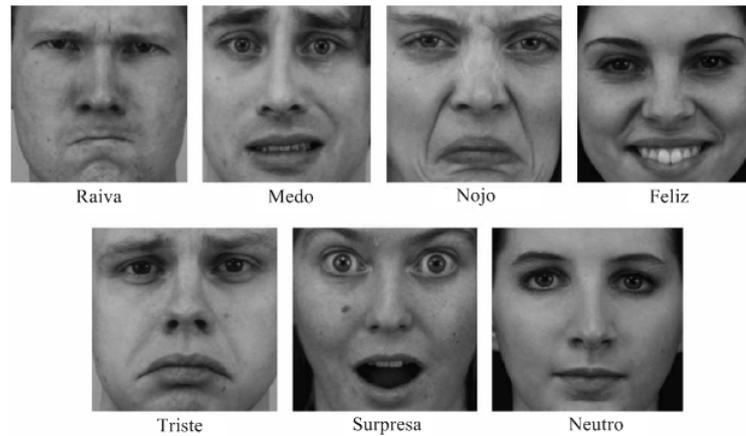
De acordo com o processo descrito por (LYONS; KAMACHI; GYOBA, 2020), cada expressadora tirou fotos de si mesma enquanto olhava através de uma folha de plástico semi-reflexiva em direção à câmera, com o cabelo amarrado longe do rosto para expor todas as zonas expressivas do rosto. Foram posicionadas luzes de tungstênio para iluminar o rosto de maneira uniforme e uma caixa delimitou a região entre a câmera e a folha de plástico para reduzir o reflexo posterior. Em seguida, as imagens foram impressas como fotografias monocromáticas e digitalizadas usando um scanner de mesa.

4.2.2 Karolinska Directed Emotional Faces (KDEF)

O dataset *Karolinska Directed Emotional Faces (KDEF)*¹¹ contém um conjunto de 4900 imagens de expressões faciais humanas de 70 indivíduos (35 mulheres e 35 homens) com resolução de 562x762 pixels no formato JPEG vistas de 5 ângulos, exibindo 7 expressões emocionais diferentes (raiva, medo, nojo, felicidade, tristeza, surpresa e neutro) conforme mostra a Figura 3.

¹¹ <<https://www.kdef.se/>>

Figura 3 – Exemplo de cada emoção do conjunto de estímulos KDEF.



Fonte – Goeleven et al. (2008, p. 5)

Todos os indivíduos eram amadores treinados com idade entre 20 e 30 anos, usando camisetas especiais cinza. Para participação na sessão de fotos, barbas, bigodes, brincos, óculos e maquiagem visível foram os critérios de exclusão. Todos os participantes foram instruídos a tentar evocar a emoção a ser expressa e tornar a expressão forte e clara (GOELEVEN et al., 2008). Eles estavam sentados a uma distância de aproximadamente três metros da câmera, a distância absoluta foi adaptada para cada sujeito ajustando-se a posição da câmera até que os olhos e a boca do sujeito estivessem em posições verticais e horizontais específicas e pré-definidas na tela da grade da câmera, enquanto as luzes foram definidas para lançar uma luz indireta suave distribuída uniformemente em ambos os lados do rosto (LUNDQVIST D.; FLYKT; ÖHMAN, 1998).

4.2.3 CK+ 48

O dataset CK+ 48¹² é uma versão recortada da base de dados CK+ (LUCEY et al., 2010) que consiste em 981 imagens de indivíduos expressando 7 emoções diferentes com resolução 48x48 pixels no formato PNG, conforme mostrado na Figura 4.

A base de dados original contém 593 sequências de vídeo de um total de 123 sujeitos diferentes, variando de 18 a 50 anos de idade com uma diversidade de gêneros e etnias. Cada vídeo mostra uma mudança facial da expressão neutra para uma expressão de pico direcionada, gravada a 30 quadros por segundo. Desses vídeos, 327 são rotulados com uma das 7 classes de expressão: raiva, desprezo, nojo, medo, felicidade, tristeza e surpresa (PAPERSWITHCODE, 2018).

¹² <<https://www.kaggle.com/shawon10/ckplus>>

Figura 4 – Exemplo das classes de expressão da base de dados CK+ 48.



Fonte – Adaptado de Shawon (2019)

4.3 Detector de Canto de Harris

De acordo com (HAN; CHEN; MENG, 2015) o canto é uma das características mais relevantes da imagem e sua detecção desempenha um papel importante na tecnologia de processamento digital de imagens, simplificando a quantidade de informações de processamento de dados, melhorando significativamente a velocidade de processamento de imagem e a precisão de correspondência.

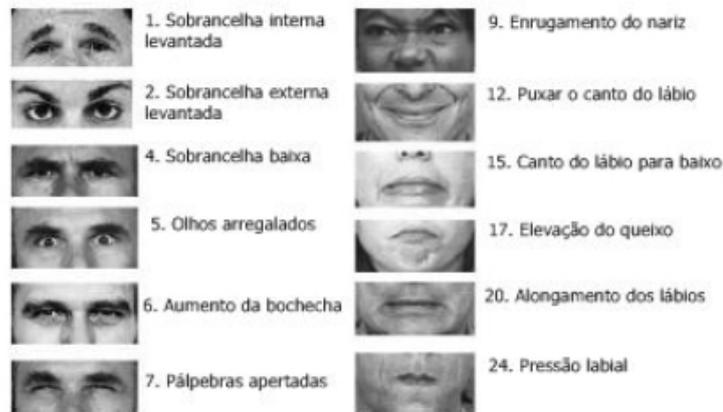
Diversos possíveis classificadores tem sido empregados para o reconhecimento de expressões faciais de maneira precisa, como Redes Neurais Artificiais (RNAs), *Support Vector Machine*, *Hidden Markov Model (HMM)*, Classificador Bayesiano, *Linear Discriminant Analysis (LDA)*, entre outros (GHIMIRE et al., 2017). Diante das diferentes técnicas de detecção de pontos de interesse propostas e avaliadas (SCHMID; MOHR; BAUCKHAGE, 2000), o detector de canto de Harris (HARRIS; STEPHENS, 1988), introduzido pela primeira vez por Chris Harris e Mike Stephens em 1988, ainda é um dos mais utilizados devido à baixa complexidade numérica e invariância ao deslocamento da imagem, rotação e variação de iluminação (ARDIZZONE; CASCIA; MORANA, 2009).

Para detectar pontos de interesse em faces para identificação de emoções, neste trabalho foi utilizado o detector de canto de Harris, que consiste em calcular o determinante da matriz de derivadas de segunda ordem da intensidade da imagem subtraindo o traço da própria matriz, sendo comumente utilizado em aplicações de visão computacional para extrair cantos e inferir recursos de uma imagem (DERPANIS, 2004). A definição de canto de Harris estava fundamentada em lugares na imagem onde a matriz de autocorrelação possui dois autovalores altos. Bons cantos para Shi e Tomasi (SHI et al., 1994), resultam na relação dos dois autovalores, desde que o menor dos dois autovalores seja maior do que

um limite mínimo.

Durante o processo, serão analisadas as *UA* definidas no *FACS*, como sobrancelha levantada interna, sobrancelha levantada externa, sobrancelha abaixada, olhos arregalados, bochecha levantada, pálpebra apertada, nariz enrugado, puxar o canto do lábio, canto do lábio para baixo, elevação do queixo, alongamento dos lábios, pressão labial, entre outros, conforme mostrado na Figura 5.

Figura 5 – Enquadramento de algumas das unidades de ação definidas no FACS.

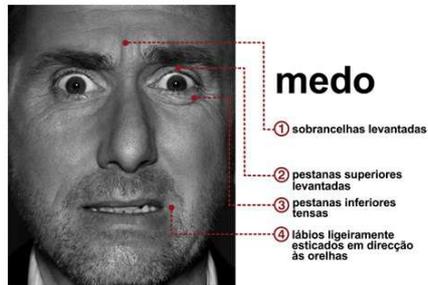


Fonte – Wu et al. (2009, p. 5)

Estas *UA* são responsáveis por auxiliar-nos na identificação das expressões faciais, e por meio dos pontos de interesse que serão detectados pelo detector de canto de Harris, seremos capazes de indentificar a emoção expressa na imagem, como nos exemplos apresentados na Figura 6.

Figura 6 – Expressões Faciais.

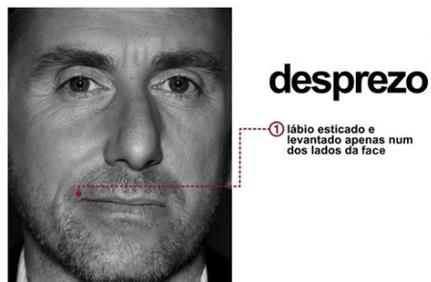
(a) Expressão de Medo



(b) Expressão de Alegria



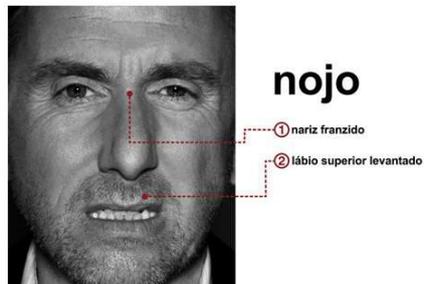
(c) Expressão de Desprezo



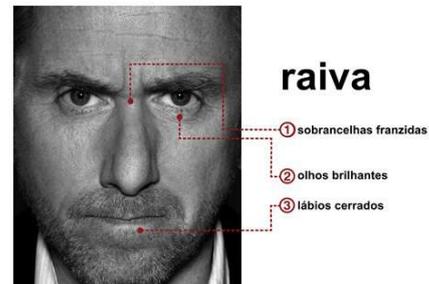
(d) Expressão de Espanto



(e) Expressão de Nojo



(f) Expressão de Raiva



(g) Expressão de Tristeza



Fonte – Lessa (2011)

A expressão matemática para a detecção de canto de Harris é descrita como na equação 4.1, onde E é a diferença entre a janela original e a janela deslocada, u é o deslocamento da janela na direção x , v é o deslocamento da janela na direção y , $w(x, y)$ é a janela de ponderação na posição (x, y) , funcionando como uma máscara e garantindo que apenas a janela desejada seja usada, I é a intensidade da imagem em uma posição (x, y) , $I(x + u, y + v)$ é a intensidade da imagem deslocada e $I(x, y)$ é a intensidade da imagem original. Logo, a medida de mudança de intensidade para $[u, v]$ pode ser aproximada pela equação 4.2, onde M é uma matriz 2x2 calculada a partir de derivadas de imagem, conforme apresentado na equação 4.3 (SHI et al., 1994).

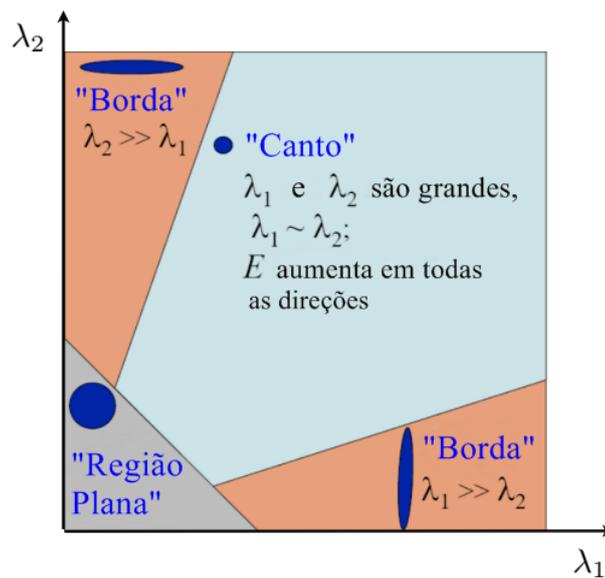
$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2 \quad (4.1)$$

$$E(u, v) \approx [u, v] M \begin{bmatrix} u \\ v \end{bmatrix} \quad (4.2)$$

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad (4.3)$$

Assumindo que os autovalores da matriz sejam λ_1 e λ_2 , seus significados matemáticos são expressos na Figura 7.

Figura 7 – Significados matemáticos dos autovalores λ_1 e λ_2 .



Fonte – Todorovic (2013, p. 21)

Caso $\lambda_1 \approx 0$ e $\lambda_2 \approx 0$, então o pixel (u, v) não possui características de interesse, caso $\lambda_1 \approx 0$ e λ_2 tiver algum valor positivo grande ou λ_1 tiver algum valor positivo grande

e $\lambda_2 \approx 0$, então uma borda é encontrada, e se λ_1 e λ_2 tiverem grandes valores positivos, então um canto é encontrado (RABIEE, 2015).

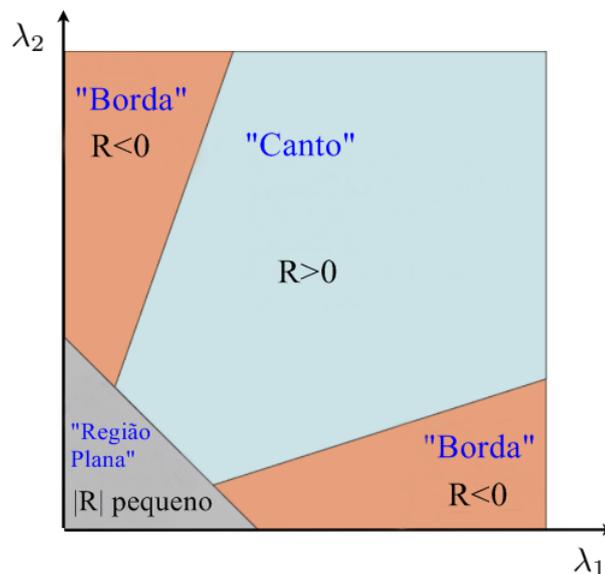
Podemos observar na função de resposta de canto de Harris apresentada na equação 4.4, que os valores próprios da matriz podem ajudar a determinar a adequação de uma janela, calculando uma pontuação R para cada janela, no qual todas as janelas com uma pontuação R maior do que um determinado valor são cantos, portanto, são bons pontos de rastreamento.

$$R = \det(M) - \alpha \cdot \text{traço}(M)^2 = \lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2 \quad (4.4)$$

onde a constante $\alpha=(0.04 \text{ a } 0.06)$ (GERIG, 2012).

Sendo assim, R depende apenas dos autovalores de M . Logo, R é maior que zero para um canto, R é negativo com grande magnitude para uma aresta e o $|R|$ é pequeno para uma região plana, conforme podemos visualizar na Figura 8.

Figura 8 – Representação da função resposta R .



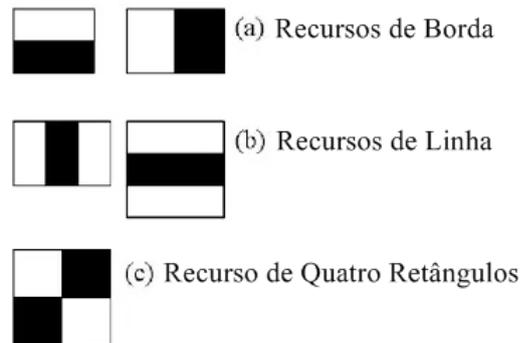
Fonte – Szeliski (2006, p. 5)

4.4 Classificador em Cascata de Haar

O Cascata de Haar é um algoritmo baseado em recursos para detecção de objetos que foi proposto em 2001 por Paul Viola e Michael Jones (VIOLA; JONES, 2001) em seu artigo “*Rapid Object Detection using a Boosted Cascade of Simple Features*”. Os autores apresentaram um novo e mais rápido método de processamento de imagens e detecção de

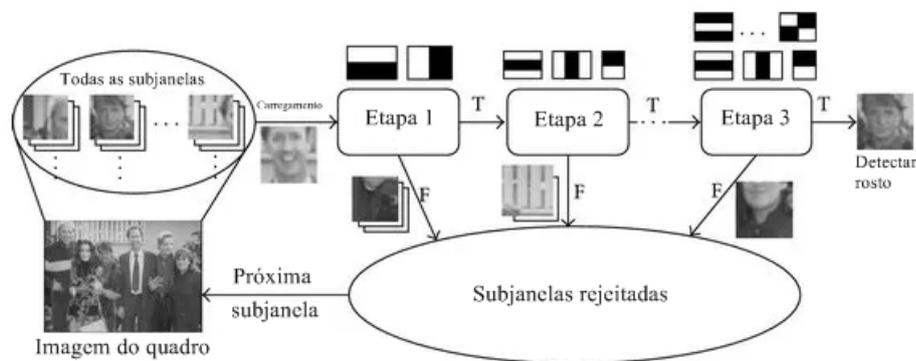
rostos usando recursos retangulares, conforme mostrado nas Figuras 9 e 10 (MAINDOLA, 2021).

Figura 9 – Tipos de recursos do Classificador em Cascata de Haar.



Fonte – Maindola (2021)

Figura 10 – Estrutura em cascata para classificadores Haar.



Fonte – Maindola (2021)

Neste trabalho, utilizamos o classificador em cascata de Haar para identificar faces frontais, carregando alguns dados pré-treinados do *OpenCV* por meio do carregamento do arquivo XML necessário, conforme mostrado na Figura 11.

Figura 11 – Exemplo de comando do *OpenCV*.

```
[ ] face_cascade = cv.CascadeClassifier("gdrive/MyDrive/TCC/Alexia-TCC-2021-1/implementação/haarcascade_frontalface_alt.xml")
```

4.5 Filtro de Gabor

O filtro de Gabor, originalmente introduzido por Dennis Gabor (GABOR, 1946), é um filtro linear cuja resposta ao impulso é definida por uma função harmônica multiplicada por uma função gaussiana (KAUR; JINDAL, 2014).

O filtro de Gabor têm sido utilizado em muitas aplicações como segmentação de textura, para separar várias texturas em uma imagem; reconhecimento óptico de caracteres, para reconhecer automaticamente letras manuscritas, placas numéricas e outdoors; reconhecimento de objetos, devido suas representações de frequência e orientação serem semelhantes às do sistema visual humano; gerenciamento de dimensão fractal; detecção de bordas; identificação de retina e codificação de imagens (JOSHI, 2014).

Neste trabalho, foi utilizado o filtro de Gabor 2D, cuja a fórmula é apresentada na equação 4.5.

$$f(x, y, \omega, \theta, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[\frac{-1}{2} \left(\left(\frac{x}{\sigma_x} \right)^2 + \left(\frac{y}{\sigma_y} \right)^2 \right) + j\omega(x \cos \theta + y \sin \theta) \right] \quad (4.5)$$

onde σ é a distribuição espacial, ω é a frequência e θ é a orientação. As variáveis x e y são sempre valores relativos à origem (NETO, 2016).

O filtro de Gabor 2D foi aplicado nos pontos detectados pelo Detector de Harris para ressaltar as informações nas regiões dos pontos, utilizando os parâmetros $\omega=0.2$ e $\theta=0$, escolhidos por apresentarem um melhor resultado entre os valores testados. Essa informação foi combinada com a imagem original (somada), conforme mostrado na Figura 12.

Figura 12 – Função Harris e Gabor.

```

def gab(im):
    freq=0.2
    theta=0

    im1=img_as_float(im)
    imb=np.ubyte(im)
    ime=im1[:, :, 0]
    detected_faces = face_cascade.detectMultiScale(imb)

    if detected_faces.size!=0:
        column=detected_faces[0][0]
        row=detected_faces[0][1]
        width=detected_faces[0][2]
        height=detected_faces[0][3]
        imr=im[row:row+height, column:column+width]
        cs=column
        ls=row
        imc=corner_harris(imr,method='k', k=0.04)
        listah=corner_peaks(imc,indices=True,num_peaks=100,threshold_rel=None).astype(int)
        dl=5
        dc=5
        for l,c in listah:
            if (l-dl>=0) and (l+dl<= (im1.shape[0]-1)):
                if (c-dc>=0) and (c+dc<= (im1.shape[1]-1)):
                    real_g,imag_g=gabor(ime[l-dl+ls:l+dl+1+ls,c-dc+cs:c+dc+1+cs],freq,theta=theta)
                    ime[l-dl+ls:l+dl+1+ls,c-dc+cs:c+dc+1+cs]=ime[l-dl+ls:l+dl+1+ls,c-dc+cs:c+dc+1+cs]+real_g

        ime1=ime+np.abs(ime.min())
        imef=ime1/ime1.max()

    else:
        imef=im1
        imm=np.zeros((im1.shape[0],im1.shape[1],1))
        imm[:, :, 0]=imef
    return imm

```

4.6 Preparação dos Dados

As imagens frontais presentes nas bases de dados utilizadas foram divididas em diretórios *train/* (treinamento), *test/* (teste) e *val/* (validação), contendo um subdiretório para cada uma das 7 classes de emoções, onde as imagens estão organizadas. Posteriormente, foi utilizado o *ImageDataGenerator* para aplicar os filtros e a função *flow_from_directory()* no gerador de dados para criar um iterador para cada um dos diretórios *train/*, *test/* e *val/*. No argumento *class_mode*, foi especificado o padrão ‘*categorical*’ para rótulos 2D, carregando as imagens com o tamanho de 250x250 pixels no argumento *target_size* e corrigindo o tamanho do lote em 16, conforme mostra a Figura 13.

Figura 13 – Gerador de dados e preparação dos iteradores na base dados Jaffe.

```
[ ] datagen=ImageDataGenerator(preprocessing_function=gab)

[ ] train_data = datagen.flow_from_directory('gdrive/MyDrive/projetos/TCC/Alexia - TCC - 2021/base de dados/dadosjaffe/train', target_size=(250, 250),
color_mode='grayscale', classes=None, class_mode='categorical', batch_size=16, shuffle=True, seed=7)

Found 147 images belonging to 7 classes.

[ ] test_data = datagen.flow_from_directory('gdrive/MyDrive/projetos/TCC/Alexia - TCC - 2021/base de dados/dadosjaffe/test', target_size=(250, 250),
color_mode='grayscale', classes=None, class_mode='categorical', batch_size=16, shuffle=False, seed=7)

Found 66 images belonging to 7 classes.

[ ] val_data = datagen.flow_from_directory('gdrive/MyDrive/projetos/TCC/Alexia - TCC - 2021/base de dados/dadosjaffe/val', target_size=(250, 250),
color_mode='grayscale', classes=None, class_mode='categorical', batch_size=16, shuffle=True, seed=7)

Found 66 images belonging to 7 classes.
```

O mesmo processo foi realizado para as bases de dados KDEF e CK+ 48.

A quantidade de imagens utilizadas por diretório em cada uma das bases de dados é apresentada na Tabela 2.

Tabela 2 – Quantidade de imagens por diretório.

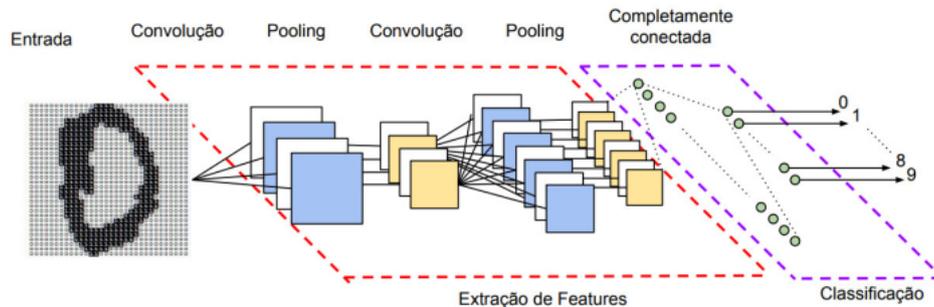
Nome da base de dados	Treinamento	Teste	Validação
JAFFE	147	66	66
KDEF	686	147	147
CK+ 48	783	104	94

4.7 Rede Neural Convolutacional

Uma Rede Neural Convolutacional (RNC) é uma rede neural que emprega a operação de convolução como uma de suas camadas, composta por diversos neurônios (KETKAR; MOOLAYIL, 2021). Considerada uma das principais categorias para fazer reconhecimento e classificações de imagens, as RNCs são amplamente utilizadas em detecção de objetos e reconhecimento facial (PRABHU, 2018).

De acordo com (VARGAS; PAES; VASCONCELOS, 2016), podemos imaginar cada neurônio sendo conectado à um conjunto de pixels da camada anterior e que a cada uma dessas conexões se atribui um peso. Logo, uma saída referente à combinação das entradas de um neurônio, utilizando os respectivos pesos de cada uma de suas conexões é produzida e passada para a próxima camada. Os pesos atribuídos às conexões de um neurônio podem ser interpretados como uma matriz que representa o filtro de uma convolução de imagens no domínio espacial, também conhecido com kernel ou máscara. Um exemplo representativo de uma RNC e suas diferentes camadas é apresentado na Figura 14.

Figura 14 – RNC e suas camadas.



Fonte – Vargas, Paes e Vasconcelos (2016)

Entre as principais vantagens da utilização de redes neurais convolucionais estão o baixo custo exigido em seu processamento e sua alta acurácia na classificação de um objeto-alvo (OLIVEIRA, 2021).

Para a construção deste modelo de rede neural convolucional, foi utilizada a biblioteca *Keras*¹³ executada em *TensorFlow*¹⁴, por meio do serviço de nuvem gratuito *Google Colab*¹⁵. Em seguida, foi criado um modelo sequencial, utilizando camadas *Conv2D*, *MaxPooling2D*, *Flatten*, *Dense* e as funções de ativação *Rectified Linear Unit (ReLU)* e *Sigmoid*. As finalidades destas funções de ativação e camadas são descritas nos tópicos a seguir:

- *ReLU*: Aplica a função de ativação da unidade linear retificada (KERAS, 2015d).
- *Sigmoid*: Aplica a função de ativação sigmóide. Para valores pequenos (<-5), retorna um valor próximo a zero, e para valores grandes (>5) o resultado da função se aproxima de 1 (KERAS, 2015d).
- *Conv2D*: Cria um kernel de convolução que é convoluído com a entrada da camada para produzir um tensor de saídas (KERAS, 2015a).
- *MaxPooling2D*: Reduz a resolução da entrada ao longo de suas dimensões espaciais (altura e largura) tomando o valor máximo em uma janela de entrada (de tamanho definido por `pool_size`) para cada canal da entrada (KERAS, 2015e).
- *Flatten*: Achata a entrada sem afetar o tamanho do lote (KERAS, 2015c).

¹³ <<https://keras.io/>>

¹⁴ <<https://www.tensorflow.org/>>

¹⁵ <<https://colab.research.google.com/>>

- *Dense*: Camada regular de rede neural densamente conectada. Cada neurônio na camada densa recebe entrada de todos os neurônios de sua camada anterior (KERAS, 2015b).

Durante a fase de compilação e treinamento, foi utilizada a função *Categorical Crossentropy*, que calcula a perda de entropia cruzada entre rótulos e previsões. Para otimização do modelo foi utilizado o método estocástico de gradiente denominado *Adam Optimizer Algorithm*, com a métrica acurácia e 50 épocas, conforme é mostrado na Figura 15. O atributo *shuffle* define se os dados de treinamento devem ser embaralhados antes de cada época.

Figura 15 – Construção do Modelo de Rede Neural Convolutacional (RNC).

```
[ ] model = Sequential()
model.add(Conv2D(16, (3, 3), input_shape=(250, 250,1), data_format='channels_last'))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Conv2D(32, (3, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Conv2D(64, (3, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Flatten())
model.add(Dense(512))
model.add(Activation('relu'))
model.add(Dense(256))
model.add(Activation('relu'))
model.add(Dense(128))
model.add(Activation('relu'))
model.add(Dense(64))
model.add(Activation('relu'))
model.add(Dense(7))
model.add(Activation('sigmoid'))

model.compile(optimizer='Adam', loss='categorical_crossentropy', metrics=['accuracy'])

[ ] hist=model.fit(train_data, validation_data=val_data, batch_size=16, epochs=50, shuffle=True)
```

A saída apresentada na Figura 16 mostra todas as camadas usadas nesta rede.

Figura 16 – Estrutura geral do modelo.

```

Model: "sequential"

```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 248, 248, 16)	160
activation (Activation)	(None, 248, 248, 16)	0
max_pooling2d (MaxPooling2D)	(None, 124, 124, 16)	0
conv2d_1 (Conv2D)	(None, 122, 122, 32)	4640
activation_1 (Activation)	(None, 122, 122, 32)	0
max_pooling2d_1 (MaxPooling2D)	(None, 61, 61, 32)	0
conv2d_2 (Conv2D)	(None, 59, 59, 64)	18496
activation_2 (Activation)	(None, 59, 59, 64)	0
max_pooling2d_2 (MaxPooling2D)	(None, 29, 29, 64)	0
flatten (Flatten)	(None, 53824)	0
dense (Dense)	(None, 512)	27558400
activation_3 (Activation)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
activation_4 (Activation)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
activation_5 (Activation)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
activation_6 (Activation)	(None, 64)	0
dense_4 (Dense)	(None, 7)	455
activation_7 (Activation)	(None, 7)	0

```

Total params: 27,754,631
Trainable params: 27,754,631
Non-trainable params: 0

```

4.8 Métricas de Avaliação de Desempenho

Os resultados dos experimentos apresentados neste trabalho foram avaliados por meio de diferentes métricas de avaliação de desempenho, conhecidas como Acurácia, Precisão, F1-Score e Revocação, cujas fórmulas matemáticas são apresentadas respectivamente na Tabela 3.

Tabela 3 – Métricas de Avaliação de Desempenho.

Métrica	Fórmula
Acurácia	$\frac{VP + VN}{VP + FN + VN + FP} \quad (4.6)$
Precisão	$\frac{VP}{VP + FP} \quad (4.7)$
F1-Score	$2 * \frac{(precisão * recall)}{(precisão + recall)} \quad (4.8)$
Revocação	$\frac{VP}{VP + FN} \quad (4.9)$

VP: verdadeiros positivos; FN: falsos negativos; FP: falsos positivos; VN: verdadeiros negativos (BIOINFO, 2021).

- Acurácia: Considerada uma das métricas mais simples e importantes, a acurácia avalia o percentual de acertos do modelo, podendo ser obtida pela razão entre a quantidade de acertos e o total de entradas (BIOINFO, 2021). Na classificação multirótulo, essa função calcula a precisão do subconjunto (SCIKIT-LEARN, 2021a).
- Precisão: É a capacidade do classificador não rotular como positiva uma amostra que é negativa. Avalia a quantidade de verdadeiros positivos sobre a soma de todos os valores positivos (SCIKIT-LEARN, 2021c).
- F1-Score: Pode ser interpretado como uma média harmônica de precisão e recall, onde um F1-Score atinge seu melhor valor em 1 e pior em 0 (SCIKIT-LEARN, 2021b).
- Revocação: Avalia a capacidade do método de detectar com sucesso resultados classificados como positivos (BIOINFO, 2021). O melhor valor é 1 e o pior valor é 0 (SCIKIT-LEARN, 2021d).

Além das métricas de avaliação apresentadas, o relatório de classificação também inclui algumas informações adicionais (LAURENTI, 2020):

- Suporte: Número de observações para cada classe.
- Média macro: Média aritmética de uma métrica entre as sete classes.

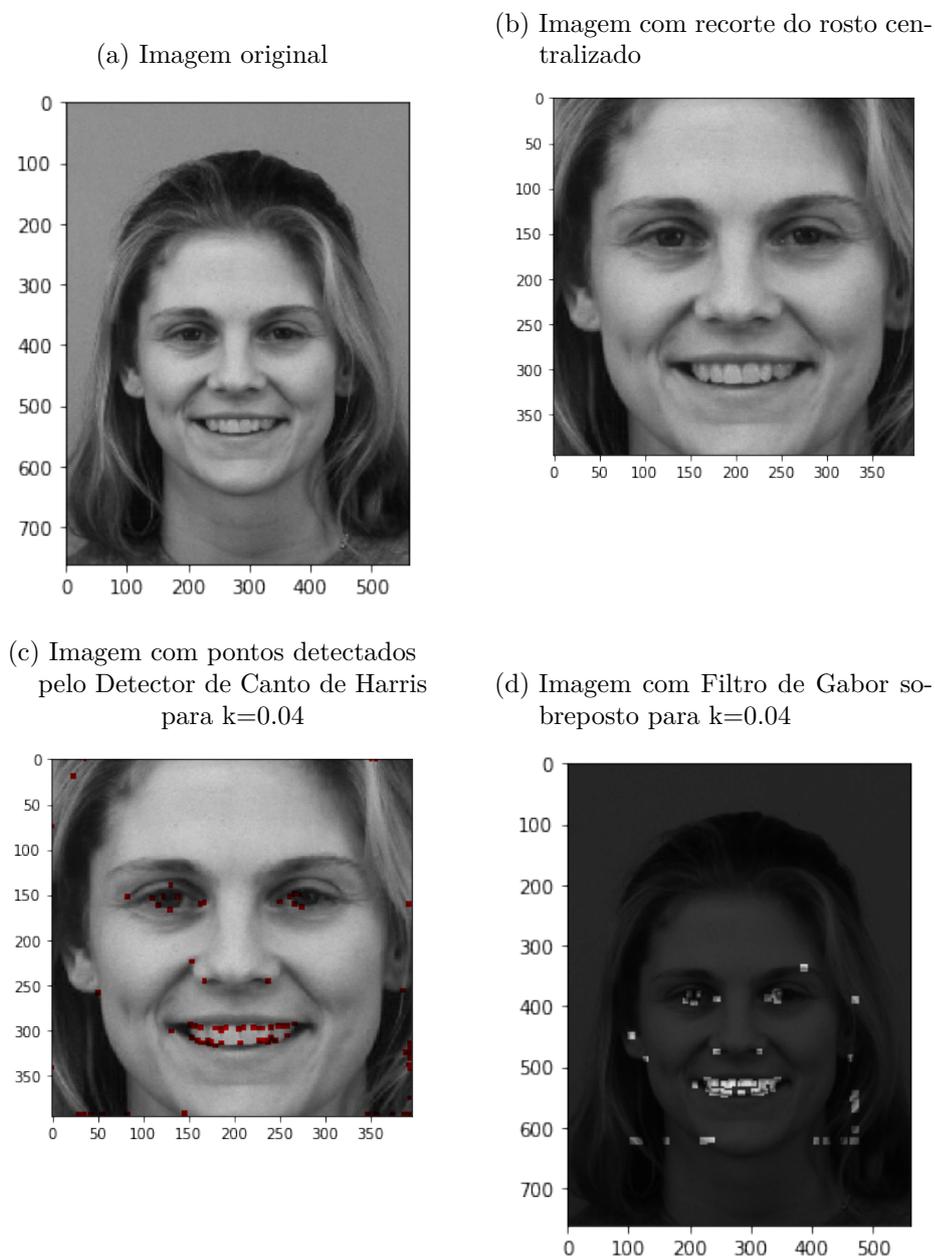
- Média ponderada: a média ponderada é calculada dividindo a soma (métrica de interesse * peso) pela soma (pesos).

5 RESULTADOS

A fim de analisar os parâmetros para o detector de canto de Harris e identificar os valores que apresentam as maiores taxas de acerto na identificação de emoções em imagens de faces, foram realizados testes utilizando oito valores para o parâmetro k nas três bases de dados escolhidas, sendo eles: 0, 0.02, 0.04, 0.05, 0.06, 0.1, 0.15 e 0.2.

Um exemplo de imagem utilizada em sua versão original, recortada, com pontos detectados pelo Detector de Harris para $k=0.04$ e com o filtro de Gabor sobreposto é apresentado na Figura 17.

Figura 17 – Exemplo de imagem com pontos detectados pelo Detector de Canto de Harris e com filtro de Gabor sobreposto.



Como podemos observar na Tabela 4, $k=0.04$ apresentou o melhor resultado, obtendo 89% de validação de acurácia na base de dados JAFFE.

Tabela 4 – Acurácias e Perdas JAFFE (treinamento e teste).

Parâmetro k	Acurácia	Validação da Acurácia	Perda	Validação da Perda
$k=0$	1.0	0.86	8.76	0.77
$k=0.02$	1.0	0.81	6.37	1.02
$k=0.04$	1.0	0.89	1.41	0.59
$k=0.05$	1.0	0.84	5.54	0.74
$k=0.06$	1.0	0.83	1.03	0.83
$k=0.1$	1.0	0.80	6.06	1.12
$k=0.15$	1.0	0.81	1.24	0.85
$k=0.2$	1.0	0.84	1.93	0.87

Os gráficos apresentados nas Figuras 18 e 19 demonstram o processo de treinamento na base de dados JAFFE.

Figura 18 – Gráfico da acurácia do modelo na base de dados JAFFE para $k=0.04$.

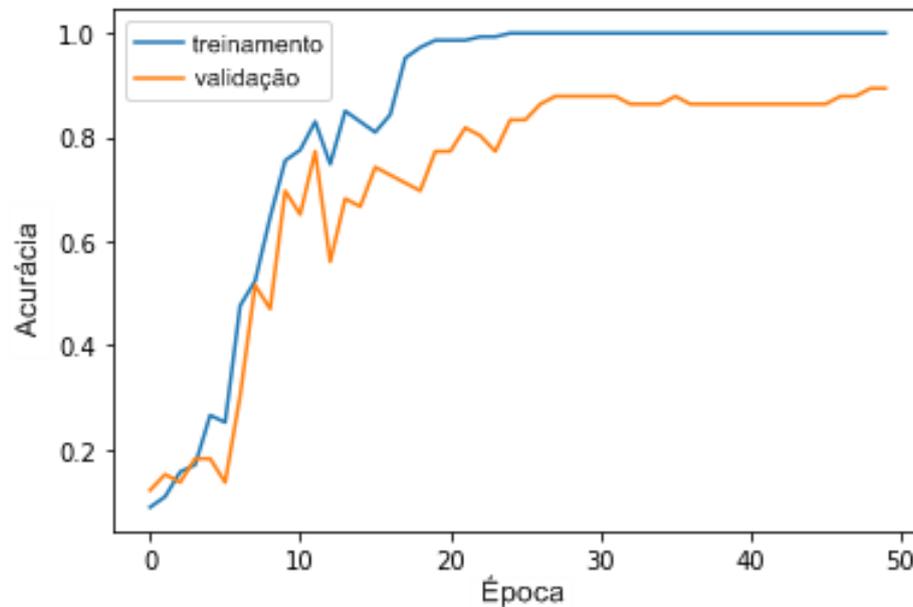
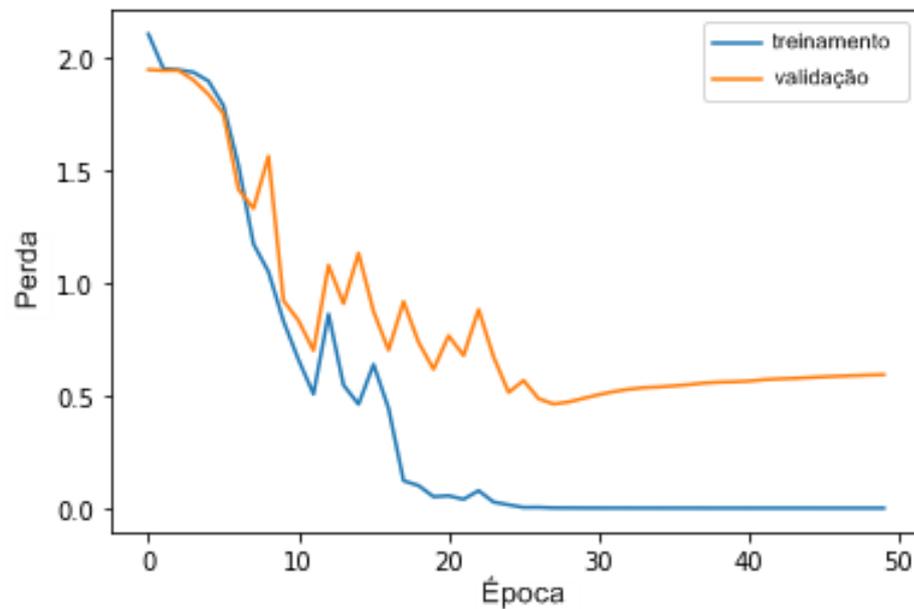


Figura 19 – Gráfico da perda do modelo na base de dados JAFFE para $k=0.04$.



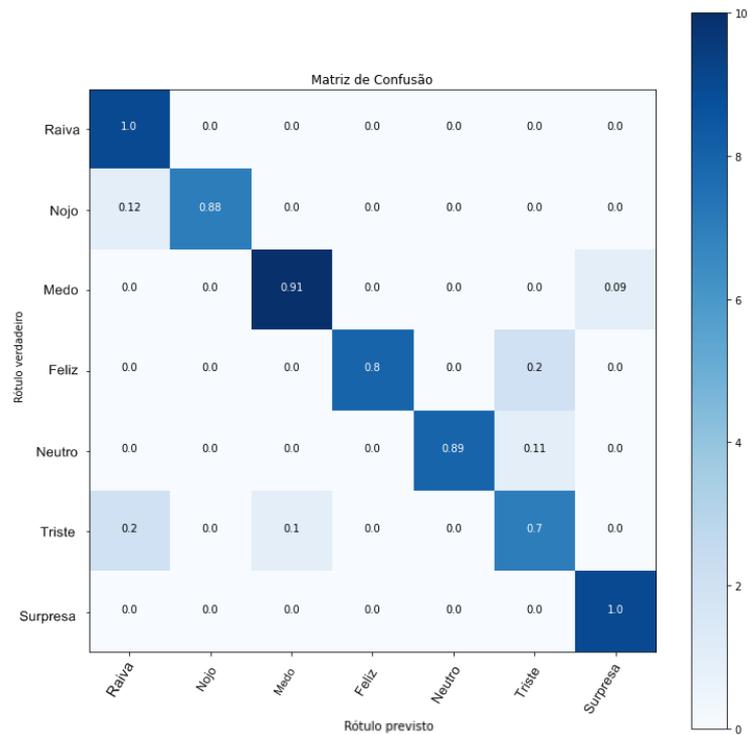
As medidas de desempenho precisão, revocação e f1-score são apresentadas na figura 20.

Figura 20 – Avaliação de desempenho do reconhecimento de emoções na base de dados JAFFE para $k=0.04$.

Relatório de Classificação

	precisão	revocação	f1-score	suporte
Raiva	0.75	1.00	0.86	9
Nojo	1.00	0.88	0.93	8
Medo	0.91	0.91	0.91	11
Feliz	1.00	0.80	0.89	10
Neutro	1.00	0.89	0.94	9
Triste	0.70	0.70	0.70	10
Surpresa	0.90	1.00	0.95	9
acurácia			0.88	66
média macro	0.89	0.88	0.88	66
média ponderada	0.89	0.88	0.88	66

A matriz de confusão resultante é apresentada na Figura 21, onde podemos verificar que a revocação mais alta foi alcançado nas classes de emoção "raiva"(100%) e "surpresa"(100%) e a mais baixa foi alcançado na classe "triste"(70%).

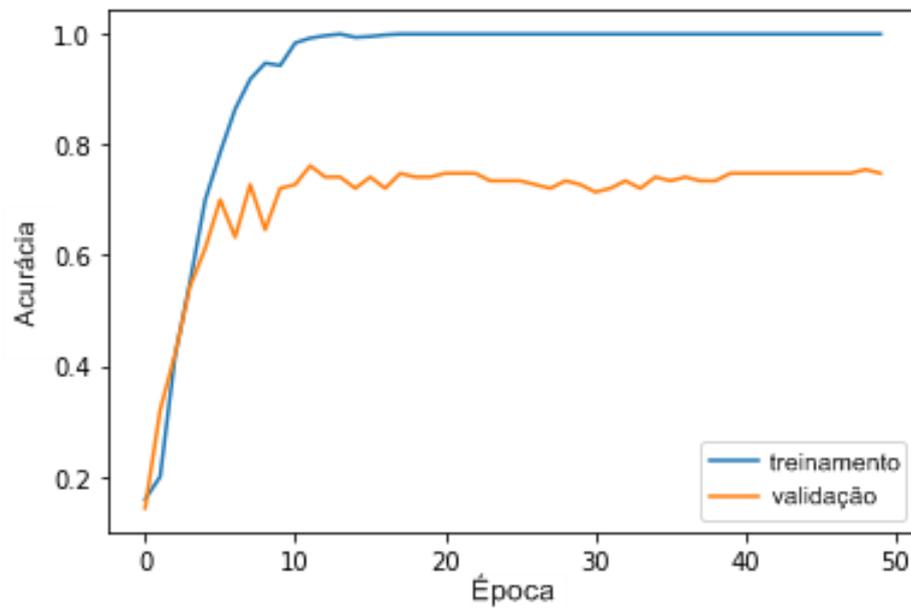
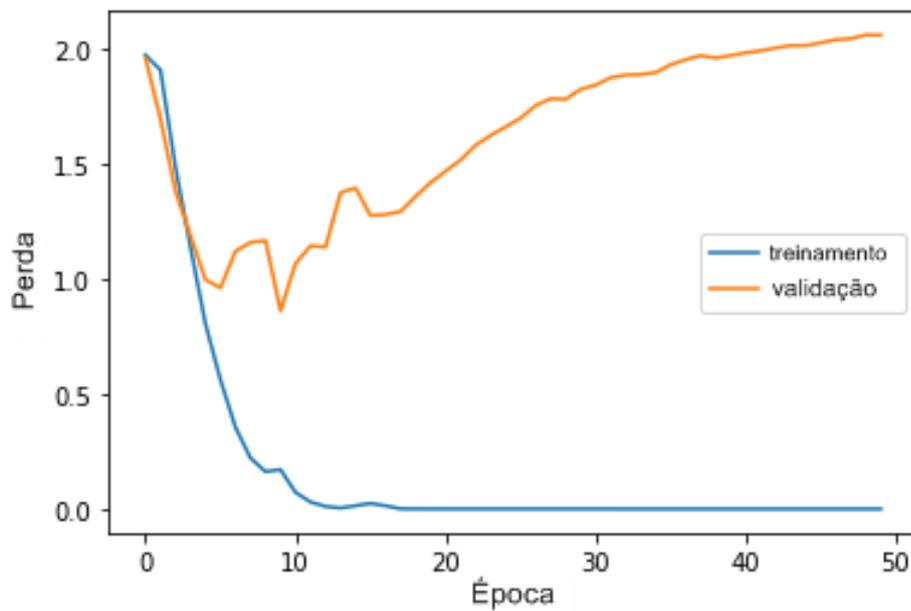
Figura 21 – Matriz de Confusão JAFFE para $k=0.04$.

Na base de dados KDEF, observa-se na Tabela 5 que $k=0.05$ apresentou o melhor resultado, obtendo 82% de validação de acurácia.

Tabela 5 – Acurácias e Perdas KDEF (treinamento e teste).

Parâmetro k	Acurácia	Validação da Acurácia	da Perda	Validação da Perda
k=0	1.0	0.78	1.75	2.13
k=0.02	1.0	0.80	1.02	1.81
k=0.04	1.0	0.80	1.54	1.92
k=0.05	1.0	0.82	1.01	1.97
k=0.06	1.0	0.78	1.36	1.91
k=0.1	1.0	0.81	2.09	2.04
k=0.15	1.0	0.77	9.89	1.74
k=0.2	1.0	0.80	8.01	1.67

Os gráficos apresentados nas Figuras 22 e 23 demonstram o processo de treinamento na base de dados KDEF.

Figura 22 – Gráfico da acurácia do modelo na base de dados KDEF para $k=0.05$.Figura 23 – Gráfico da perda do modelo na base de dados KDEF para $k=0.05$.

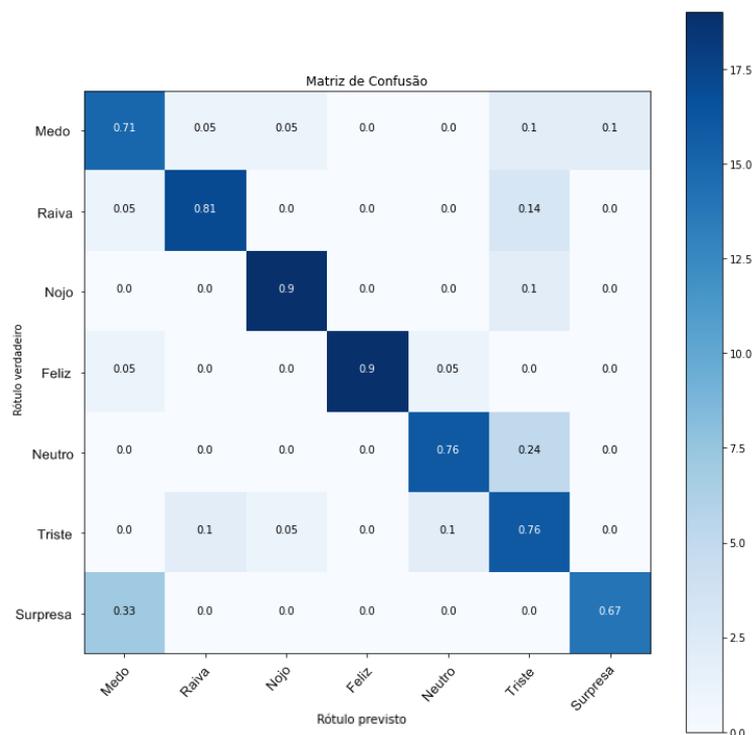
As medidas de desempenho precisão, revocação e f1-score são apresentadas na figura Figura 24.

Figura 24 – Avaliação de desempenho do reconhecimento de emoções na base de dados KDEF para $k=0.05$.

Relatório de Classificação				
	precisão	revocação	f1-score	suporte
Medo	0.62	0.71	0.67	21
Raiva	0.85	0.81	0.83	21
Nojo	0.90	0.90	0.90	21
Feliz	1.00	0.90	0.95	21
Neutro	0.84	0.76	0.80	21
Triste	0.57	0.76	0.65	21
Surpresa	0.88	0.67	0.76	21
acurácia			0.79	147
média macro	0.81	0.79	0.79	147
média ponderada	0.81	0.79	0.79	147

A matriz de confusão resultante é apresentada na Figura 25, onde podemos verificar que a revocação mais alta foi alcançada nas classes de emoção "nojo"(90%) e "feliz"(90%) e a mais baixa foi alcançada na classe "surpresa"(67%).

Figura 25 – Matriz de Confusão KDEF para $k=0.05$.



Podemos observar na Tabela 6 que os testes na base de dados CK+ 48 obtiveram

o melhor resultado, onde $k=0.04$ apresentou 100% de sucesso em todas as métricas de avaliação de desempenho utilizadas.

Tabela 6 – Acurácias e Perdas CK+ 48 (treinamento e teste).

Parâmetro k	Acurácia	Validação da Acurácia	da Perda	Validação da Perda
k=0	1.0	0.98	1.52	0.10
k=0.02	1.0	0.97	1.49	0.12
k=0.04	1.0	1.0	4.73	0.0066
k=0.05	1.0	0.99	3.88	0.05
k=0.06	1.0	0.98	3.85	0.06
k=0.1	1.0	0.98	1.89	0.04
k=0.15	1.0	0.99	3.36	0.10
k=0.2	1.0	0.94	1.03	0.20

Os gráficos apresentados nas Figuras 26 e 27 demonstram o processo de treinamento na base de dados CK+ 48.

Figura 26 – Gráfico da acurácia do modelo na base de dados CK+ 48 para $k=0.04$.

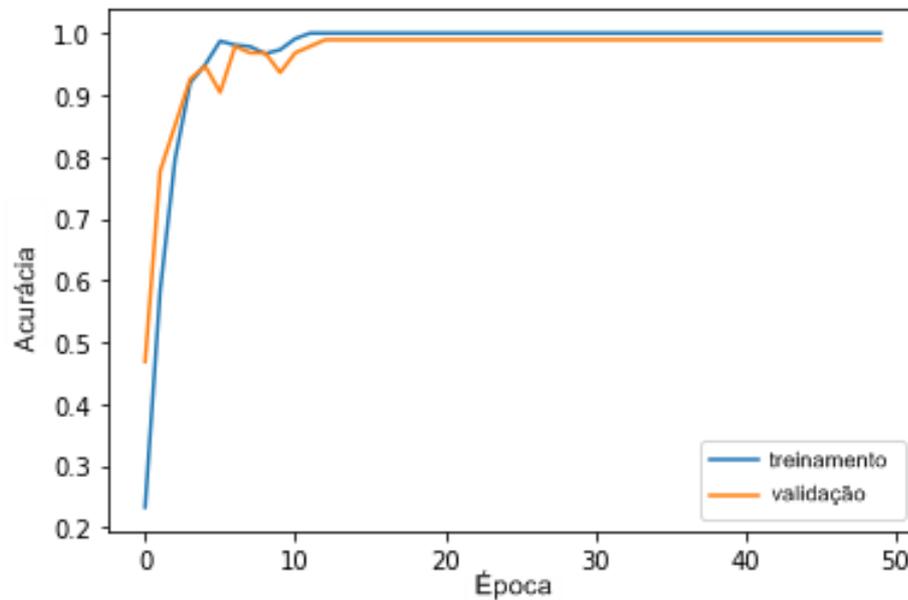
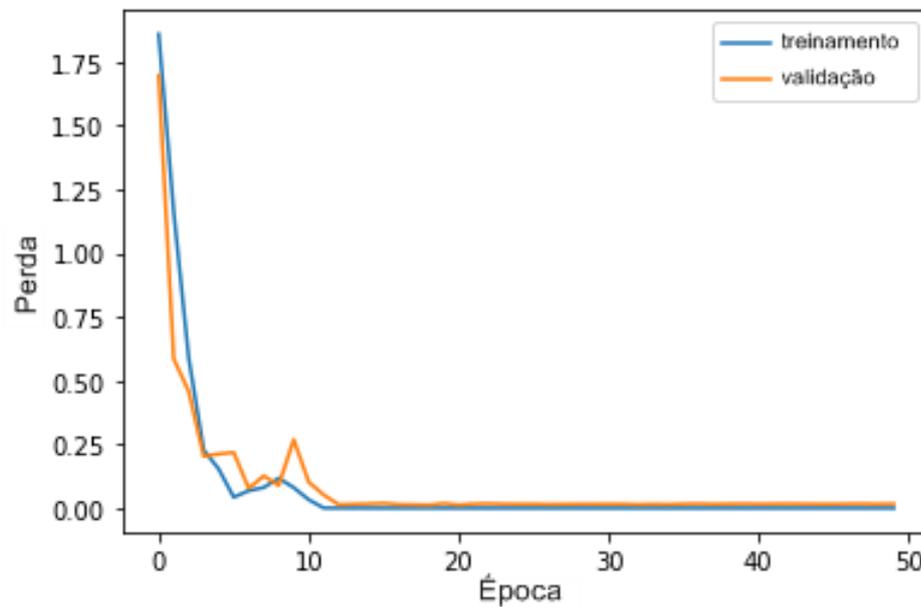


Figura 27 – Gráfico da perda do modelo na base de dados CK+ 48 para $k=0.04$.



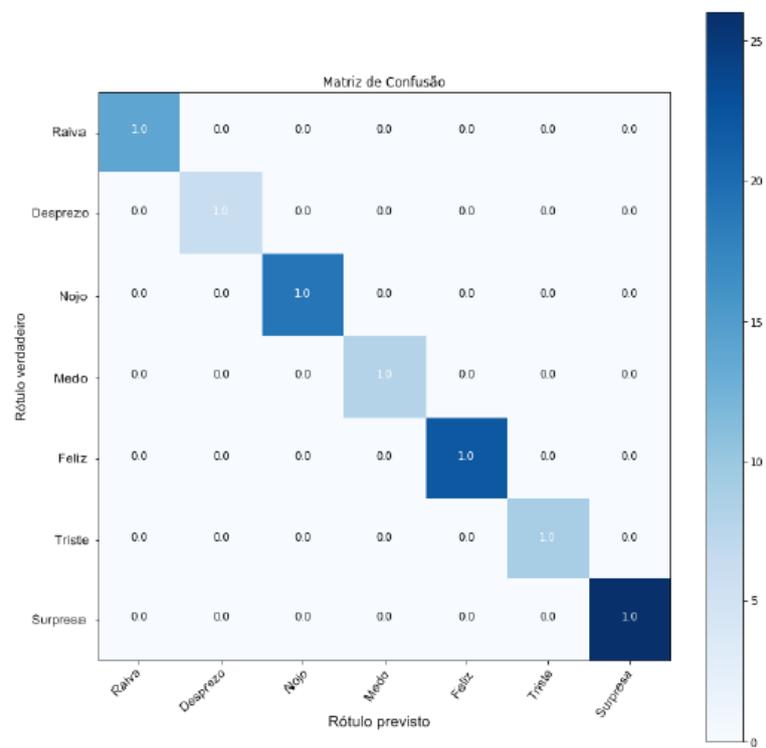
As medidas de desempenho precisão, revocação e f1-score são apresentadas na figura 28.

Figura 28 – Avaliação de desempenho do reconhecimento de emoções na base de dados CK+ para $k=0.04$.

Relatório de Classificação

	precisão	revocação	f1-score	suporte
Raiva	1.00	1.00	1.00	14
Desprezo	1.00	1.00	1.00	6
Nojo	1.00	1.00	1.00	19
Medo	1.00	1.00	1.00	8
Feliz	1.00	1.00	1.00	22
Triste	1.00	1.00	1.00	9
Surpresa	1.00	1.00	1.00	26
acurácia			1.00	104
média macro	1.00	1.00	1.00	104
média ponderada	1.00	1.00	1.00	104

A matriz de confusão resultante é apresentada na Figura 29, onde podemos observar que foi obtido 100% de revocação em todas as emoções na base de dados CK+.

Figura 29 – Matriz de Confusão CK+ 48 para $k=0.04$.

6 CONSIDERAÇÕES FINAIS

6.1 Conclusão

O desenvolvimento no presente estudo teve como objetivo a utilização do detector de canto de Harris na identificação de emoções em imagens de faces. Foi realizada uma análise a fim de verificar a maneira como a alteração dos valores do parâmetro k é capaz de influenciar nos resultados, a fim de identificarmos os melhores parâmetros para o detector de canto de Harris. Por meio dos testes realizados nas três bases de dados apresentadas, foi possível verificar que os melhores valores para k foram 0.04 e 0.05. Na base de dados JAFFE, obteve-se 89% de acurácia com $k=0.04$ e 100% de acurácia na CK+ 48 também com $k=0.04$. Na KDEF, o melhor resultado foi obtido com $k=0.05$, resultando em 82% de acurácia.

Este estudo comprovou que o detector de canto de Harris é eficiente no processo de detecção de cantos como características faciais presentes nas imagens, auxiliando no processo de identificação de emoções quando combinado com outras tecnologias como RNCs, que tornaram-se cada vez mais utilizadas em aplicações que necessitam de predições e adaptações para determinados cenários.

6.2 Trabalhos Futuros

Como possíveis trabalhos futuros proponho:

- Realizar o mesmo experimento com melhores recursos, para otimização do tempo.
- Realizar experimentos com diferentes modelos de rede neural e variações de épocas.
- Incluir outras bases de dados a serem testadas, como por exemplo AffectNet e RAF-DB.
- Realizar testes em imagens com variações de dimensão, rotação, brilho e ruído.

REFERÊNCIAS

- ALI, M. F.; KHATUN, M.; TURZO, N. A. Facial emotion detection using neural network. **International Journal of Scientific Engineering Research**, v. 11, n. 8, p. 1–8, 2020. Citado na página 25.
- ÁLVAREZ, V. M. et al. A method for facial emotion recognition based on interest points. In: IEEE. **2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)**. [S.l.], 2018. p. 1–4. Citado na página 25.
- ANTONINI, G.; POPOVICI, V.; THIRAN, J.-P. Independent component analysis and support vector machine for face feature extraction. In: SPRINGER. **International Conference on Audio-and Video-Based Biometric Person Authentication**. [S.l.], 2003. p. 111–118. Citado na página 27.
- ARDIZZONE, E.; CASCIA, M. L.; MORANA, M. Probabilistic corner detection for facial feature extraction. In: SPRINGER. **International Conference on Image Analysis and Processing**. [S.l.], 2009. p. 461–470. Citado na página 37.
- BEHAINE, C. A.; SCHARCANSKI, J. Enhancing the performance of active shape models in face recognition applications. **IEEE Transactions on Instrumentation and Measurement**, IEEE, v. 61, n. 8, p. 2330–2333, 2012. Citado na página 27.
- BHOWMIK, M. K.; SHIL, S.; SAHA, P. Feature points extraction of thermal face using harris interest point detection. **Procedia Technology**, Elsevier, v. 10, p. 724–730, 2013. Citado na página 28.
- BIOINFO. **Métricas de Avaliação em Machine Learning**. 2021. Disponível em: <<https://bioinfo.com.br/metricas-de-avaliacao-em-machine-learning/>>. Citado na página 49.
- CHAUGULE, V. et al. Product review based on optimized facial expression detection. In: IEEE. **2016 Ninth International Conference on Contemporary Computing (IC3)**. [S.l.], 2016. p. 1–6. Citado na página 29.
- DERPANIS, K. G. The harris corner detector. **York University, Citeseer**, v. 2, 2004. Citado na página 37.
- EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. **Journal of personality and social psychology**, American Psychological Association, v. 17, n. 2, p. 124–129, 1971. Citado na página 25.
- FRIESEN, E.; EKMAN, P. Facial action coding system: a technique for the measurement of facial movement. **Palo Alto, Consulting Psychologists Press**, v. 3, n. 2, p. 5, 1978. Citado na página 25.
- GABOR, D. **Theory of communication: Journal of the Institute of Electrical Engineers**, **93**, 429–457. [S.l.]: DGZAAW, 1946. Citado na página 43.
- GERIG, G. **Keypoint Detection: Harris Operator**. 2012. Disponível em: <<http://www.sci.utah.edu/~gerig/CS6320-S2013/Materials/Harris-corner-detection.pdf>>. Citado na página 41.

- GHIMIRE, D. et al. Facial expression recognition based on local region specific features and support vector machines. **Multimedia Tools and Applications**, Springer, v. 76, n. 6, p. 7803–7821, 2017. Citado na página 37.
- GOELEVELN, E. et al. The karolinska directed emotional faces: a validation study. **Cognition and emotion**, Taylor & Francis, v. 22, n. 6, p. 1094–1118, 2008. Citado na página 36.
- GONDHI, N. K. et al. An efficient algorithm for facial landmark detection using haar-like features coupled with corner detection following anthropometric constraints. In: IEEE. **2017 2nd International Conference on Telecommunication and Networks (TEL-NET)**. [S.l.], 2017. p. 1–6. Citado na página 29.
- HAMOUIZ, M. **Feature-based affine-invariant detection and localization of faces**. Tese (Doutorado) — University of Surrey, 2004. Citado na página 27.
- HAN, Y.; CHEN, P.; MENG, T. Harris corner detection algorithm at sub-pixel level and its application. In: ATLANTIS PRESS. **2015 International Conference on Computational Science and Engineering**. [S.l.], 2015. Citado na página 37.
- HARRIS, C. G.; STEPHENS, M. A combined corner and edge detector. In: CITESEER. **Alvey vision conference**. [S.l.], 1988. v. 15, n. 50, p. 10–5244. Citado na página 37.
- HUSSAIN, S. **Emotion detection from frontal facial image**. Tese (Doutorado) — BRAC University, 2013. Citado na página 25.
- JI, Y.; IDRISSEI, K. Facial expression recognition by automatic facial parts position detection with boosted-lbp. In: IEEE. **2009 Fifth International Conference on Signal Image Technology and Internet Based Systems**. [S.l.], 2009. p. 28–35. Citado na página 35.
- JOSHI, P. **Understanding Gabor Filters**. 2014. Disponível em: <https://prateekvjoshi.com/2014/04/26/understanding-gabor-filters/>. Acessado em: 24 fev. 2022. Citado na página 43.
- KAMAROL, S. K. A. et al. Spatiotemporal feature extraction for facial expression recognition. **IET Image Processing**, IET, v. 10, n. 7, p. 534–541, 2016. Citado na página 29.
- KAUR, A.; JINDAL, G. Texture based image segmentation using gabor filters. **International Journal of Engineering Science & Advanced Technology [IJESAT], IJASCSE**, Citeseer, v. 3, n. 3, 2014. Citado na página 43.
- KERAS. **Conv2D layer**. 2015. Disponível em: <https://keras.io/api/layers/convolution_layers/convolution2d/>. Citado na página 46.
- KERAS. **Dense layer**. 2015. Disponível em: <https://keras.io/api/layers/core_layers/dense/>. Citado na página 47.
- KERAS. **Flatten layer**. 2015. Disponível em: <https://keras.io/api/layers/reshaping_layers/flatten/>. Citado na página 46.
- KERAS. **Layer activation functions**. 2015. Disponível em: <<https://keras.io/api/layers/activations/>>. Citado na página 46.

- KERAS. **MaxPooling2D layer**. 2015. Disponível em: <https://keras.io/api/layers/pooling_layers/max_pooling2d/>. Citado na página 46.
- KERAS. **About Keras**. 2020. Disponível em: <<https://keras.io/about/>>. Citado na página 33.
- KESARWANI, P.; CHOUDHARY, A. K.; MISRA, A. K. Automatic facial feature point extraction by using harris algorithm. **Akgec International Journal of Technology**, v. 9, n. 1, p. 1–6, 2013. Citado na página 28.
- KETKAR, N.; MOOLAYIL, J. Convolutional neural networks. In: **Deep Learning with Python**. [S.l.]: Springer, 2021. p. 197–242. Citado na página 45.
- LAURENTI, G. **Confusion Matrix and Classification Report**. 2020. Disponível em: <<https://medium.com/swlh/confusion-matrix-and-classification-report-88105288d48f>>. Citado na página 49.
- LESSA, J. **Expressões Faciais: comunicação de emoções, veículo de significados**. 2011. Disponível em: <<http://imagemnacomunicacao.blogspot.com/2011/02/facial-expressions-communicate-emotions.html>>. Citado na página 39.
- LISSETTI, C. L.; RUMELHART, D. E. Facial expression recognition using a neural network. In: **FLAIRS Conference**. [S.l.: s.n.], 1998. p. 328–332. Citado na página 25.
- LUCEY, P. et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: **IEEE. 2010 ieee computer society conference on computer vision and pattern recognition-workshops**. [S.l.], 2010. p. 94–101. Citado na página 36.
- LUNDQVIST D.; FLYKT, A.; ÖHMAN, A. **About KDEP**. 1998. Disponível em: <<https://kdef.se/home/aboutKDEF.html>>. Citado na página 36.
- LYONS, M. J.; KAMACHI, M.; GYOBA, J. Coding facial expressions with gabor wavelets (ivc special issue). **arXiv preprint arXiv:2009.05938**, 2020. Citado na página 35.
- MA, L.; KHORASANI, K. Facial expression recognition using constructive feedforward neural networks. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 34, n. 3, p. 1588–1595, 2004. Citado na página 25.
- MAINDOLA, G. **Face Detection with HAAR Cascade in OpenCV Python**. 2021. Disponível em: <<https://machinelearningknowledge.ai/face-detection-with-haar-cascade-in-opencv-python/>>. Citado na página 42.
- MATPLOTLIB. **Matplotlib: Visualization with Python**. 2021. Disponível em: <<https://matplotlib.org/>>. Citado na página 34.
- MEHENDALE, N. Facial emotion recognition using convolutional neural networks (ferc). **SN Applied Sciences**, Springer, v. 2, n. 3, p. 1–8, 2020. Citado na página 25.
- NETO, J. do E. S. B. **Filtros de Gabor**. 2016. Disponível em: <<http://www.lcad.icmc.usp.br/~jbatista/procimg/2016/gabor.pdf>>. Citado na página 43.

NOOR, J. et al. Facial expression recognition using hand-crafted features and supervised feature encoding. In: IEEE. **2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)**. [S.l.], 2020. p. 1–5. Citado na página 30.

OLIVEIRA, A. Redes neurais convolucionais: Análise de imagens de vídeo em tempo real para identificação de faces e verificação do uso de máscaras de proteção. Pontifícia Universidade Católica de Goiás, 2021. Citado na página 46.

OPENCV. **About OpenCV**. 2022. Disponível em: <<https://opencv.org/about/>>. Citado na página 34.

PAPERSWITHCODE. **CK+ (Extended Cohn-Kanade dataset)**. 2018. Disponível em: <<https://paperswithcode.com/dataset/ck>>. Citado na página 36.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011. Citado na página 34.

PRABHU, R. **Understanding of Convolutional Neural Network (CNN) — Deep Learning**. 2018. Disponível em: <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>. Acessado em: 24 fev. 2022. Citado na página 45.

PRIYA, R. V. Emotion recognition from face. **International Journal of Pure and Applied Mathematics**, v. 119, n. 16, p. 1–10, 2018. Citado na página 30.

RABIEE, H. **Digital Image Processing Edge and Keypoint Detection**. 2015. Disponível em: <http://ce.sharif.edu/courses/94-95/1/ce933-1/resources/root/Lectures/Lecture11_EdgeAndKeypoinDetection.pdf>. Citado na página 41.

RAJESH, K.; NAVEENKUMAR, M. A robust method for face recognition and face emotion detection system using support vector machines. In: IEEE. **2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)**. [S.l.], 2016. p. 1–5. Citado na página 25.

RAMOS, Y. D. Y. Estimativa da pose da cabeça em imagens monoculares usando um modelo no espaço 3d. 2013. Citado na página 28.

RASHMI; SAINI, G. S.; BAGCHI, A. Human facial emotion recognition using harris point and support vector machine. **International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)**, v. 7, 2018. Citado na página 30.

RESEARCH, G. **Colaboratory**. 2017. Disponível em: <<https://research.google.com/colaboratory/faq.html>>. Citado na página 33.

ROSA, C. C.; BUENO, F. **Identificação de emoções mediante expressões faciais em imagens de webcam utilizando técnicas de aprendizagem de máquina em ambiente web**. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2013. Citado na página 28.

- SÁNCHEZ, J.; MONZÓN, N.; NUEZ, A. S. D. L. An analysis and implementation of the harris corner detector. **Image Processing On Line**, 2018. Citado na página 30.
- SCHMID, C.; MOHR, R.; BAUCKHAGE, C. Evaluation of interest point detectors. **International Journal of computer vision**, Springer, v. 37, n. 2, p. 151–172, 2000. Citado na página 37.
- SCIKIT-LEARN. **sklearn.metrics.accuracy_score**. 2021. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>. Citado na página 49.
- SCIKIT-LEARN. **sklearn.metrics.f1_score**. 2021. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html>. Citado na página 49.
- SCIKIT-LEARN. **sklearn.metrics.precision_score**. 2021. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html>. Citado na página 49.
- SCIKIT-LEARN. **sklearn.metrics.recall_score**. 2021. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html>. Citado na página 49.
- SHAWON, A. **CKPLUS**. 2019. Disponível em: <<https://www.kaggle.com/shawon10/ckplus>>. Citado na página 37.
- SHI, J. et al. Good features to track. In: IEEE. **1994 Proceedings of IEEE conference on computer vision and pattern recognition**. [S.l.], 1994. p. 593–600. Citado 2 vezes nas páginas 37 e 40.
- SZELISKI, R. **Notes on the Harris Detector**. 2006. Disponível em: <<https://courses.cs.washington.edu/courses/cse576/06sp/notes/HarrisDetector.pdf>>. Citado na página 41.
- TENSORFLOW. **TensorFlow**. 2021. Disponível em: <<https://www.tensorflow.org/>>. Citado na página 33.
- TODOROVIC, S. **ECE 468 / CS 519: Digital Image Processing Gradients, Harris Corners**. 2013. Disponível em: <http://web.engr.oregonstate.edu/~sinisa/courses/OSU/ECE468/lectures/ECE468_7.pdf>. Citado na página 40.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. **Proceedings of the xxix conference on graphics, patterns and images**. [S.l.], 2016. v. 1, n. 4. Citado 2 vezes nas páginas 45 e 46.
- VERMA, S. K.; PARASHAR, G. Evaluation of corner detection algorithms for human emotion modeling. **IJCA Proceedings on National Conference on IPR, Future Technology, Optimization and Management**, 2018. Citado na página 29.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: IEEE. **Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001**. [S.l.], 2001. v. 1, p. I–I. Citado na página 41.

- WALT, S. V. D.; COLBERT, S. C.; VAROQUAUX, G. The numpy array: a structure for efficient numerical computation. **Computing in science & engineering**, IEEE, v. 13, n. 2, p. 22–30, 2011. Citado na página 34.
- WALT, S. van der. **scikit-image image processing in python**. 2021. Disponível em: <<https://scikit-image.org/>>. Citado na página 33.
- WIDANAGAMAACHCHI, W. N. Facial emotion recognition with a neural network approach. **University of Colombo**, 2009. Citado na página 27.
- WU, T. et al. Learning to make facial expressions. In: IEEE. **2009 IEEE 8th International Conference on Development and Learning**. [S.l.], 2009. p. 1–6. Citado na página 38.
- ZHAO, Y. **Human emotion recognition from body language of the head using soft computing techniques**. Tese (Doutorado) — University of Ottawa, 2012. Citado na página 27.