

---

UNIVERSIDADE FEDERAL DE OURO PRETO  
DEPARTAMENTO DE CIÊNCIA DE COMPUTAÇÃO  
PROCESSAMENTO DIGITAL DE IMAGENS

**Projeto de Pesquisa**

HUMAN EMOTION DETECTION ON SURVEILLANCE  
CAMERAS

---

**Alunos:**

San Cunha da Silva

Gabriel Fernandes Niquini

## **Resumo**

Com o objetivo de aplicar o conteúdo aprendido na disciplina BCC326 (Processamento de Imagens) foi estudado a implementação de um algoritmo para reconhecimento de expressões humanas afim de debater sobre o tema "Human emotion detection on surveillance cameras" e quais métodos vêm sendo utilizados pela academia.

A solução estudada utiliza a linguagem de programação Python, pois, é uma linguagem cujo a comunidade disponibilizam uma vasta gama de bibliotecas e frameworks que facilitam muito na hora de desenvolver projetos. Uma dessas bibliotecas, a OpenCV é a responsável por pela captura das imagens. Após feita a captura outras bibliotecas de Deep Learning terminam o serviço de classificar qual é a emoção representada naquela imagem.

# 1 Introdução

Reconhecimento de rostos por máquina a partir de imagens fixas e de vídeo está emergindo como uma área de pesquisa ativa abrangendo várias disciplinas como processamento de imagem, reconhecimento de padrões, visão computacional e redes neurais. Além disso, cara a tecnologia de reconhecimento (FRT) tem inúmeras e aplicativos de aplicação da lei. Essas aplicações variam de correspondência estática de fotografias de formato controlado, como passaportes, cartões de crédito, documentos de identidade com foto, carteiras de motorista e fotos de caneca para correspondência em tempo real de imagens de vídeo de vigilância apresentando diferentes restrições em termos de processamento requisitos. [Chellappa et al. \(1995\)](#). Uma descrição de um artigo de 25 anos atrás que mostra a visão atual do cenário, onde várias evoluções foram feitas mas com o avanço das tecnologias é cada dia mais estudado novas formas de garantir acessos privados.

Este trabalho tem como objetivo trazer algumas aplicações descritas na literatura e uma breve análise sobre cada uma. Estudar e implementar uma solução diferentes das analisadas e mostrar os resultados encontrados pelo algoritmo escolhido.

A organização segue a seguinte ordem, primeiro uma breve análise sobre alguns trabalhos encontrados na literatura nos últimos 3 anos. Seguindo com a metodologia aplicada e por fim os resultados com as conclusões tiradas a partir da realização desse projeto.

## 2 Estado da Arte

Esta seção descreve, com o apoio das principais referências bibliográficas, aspectos e questões relevantes ao tema proposto. A base para a realização do projeto inclui temas relacionados à extração de características multimodais e ao reconhecimento de atividades humanas de alto nível.

### 2.1 Trabalhos Relacionados

[Fernández-Caballero et al. \(2016\)](#) apresenta um ambiente inteligente (AI) de regulação de emoções humanas em locais voltados à cuidados com a saúde, tais como hospitais, clínicas e centros especializados de cuidados (maternidades, clínicas psiquiátricas). Os autores propõem uma arquitetura que opera em um pipeline de 3 estágios, chamados de “Detecção Humana”, “Regulação da

emoção” e “Feedback do controle das emoções”. No primeiro, os dados relacionados a emoção do paciente são recebidos. No segundo, são oferecidos a ele/ela diferentes estilos musicais e jogos de luzes/cores, e, no terceiro, é realizado um loop do feedback do controle para avaliar o efeito da regulação das emoções através da detecção delas. Neste artigo é apresentada uma arquitetura para gerenciamento de contexto baseado em domínio. Tais arquiteturas abordam características bem conhecidas da computação pervasiva e fornecem estruturas para facilitar a cooperação entre os desenvolvedores de aplicativos. No entanto, são necessárias fontes de comunicação confiáveis para garantir que não haverá perdas ou delays na transmissão dos dados. Também é necessário constante monitoramento do status dos dispositivos inteligentes para lidar com casos de falha e realocação de recursos para manter a performance do sistema.

Demétrius Alex Scardua (2018) busca inferir as emoções básicas como: raiva, medo, repulsa, surpresa, alegria e tristeza, utilizando-se de alguns passos intermediários para esta tarefa, sendo eles: detecção da face através do método face Detection (FaD); A detecção de elementos constituintes de um rosto humano utilizando o método Feature Detection (FeD); algoritmos de aprendizagem de IA e esquemas de codificações de expressões faciais para posteriormente classificar a emoção presente nas imagens obtidas pelas câmeras. Um Sistema de Reconhecimento de Emoções (SER) pode ser utilizado no âmbito da segurança, na prevenção de tumultos ou associado a um polígrafo, por exemplo. Porém, alguns contratempos podem ser encontrados pelo SER, tais como uma face que está parcialmente escondida ou encoberta, ou uma pessoa que tenha feito aplicação de botox.

Buscando reconhecer alegria, raiva ou medo por meio de ações como andar, pular e sentar, J & Geetha (2017) aplica limites para reduzir a modificação de pixel devido ao ruído da câmera e mudanças nas condições de iluminação. A partir disso, retira do quadro atual o objeto e o compara com uma base de dados pública. São propostos três classificadores, vetores de suporte (SVM), Naïve Bayes e dynamic time warping (DTW). Respectivamente a porcentagem de acerto de cada classificador é 91.98%, 89.91%, e 93.39%, ou seja, os melhores resultados foram encontrados com DTW. É possível verificar que salto com raiva, sentar com medo e salto com medo não conseguem ser identificados com muita precisão e a explicação para isso exige estudos mais aprofundados.

Também utilizando de 3 classificadores (vetor de suporte multi classe (multi-SVM), árvore de decisão e floresta aleatória), Nguyen et al. (2017) propõe uma solução para identificar expressões

positivas, negativas e neutras em tempo real. Os experimentos foram realizados utilizando uma captura de 640 x 480 pixels, o que é uma qualidade baixa para os padrões atuais mas explicável pela necessidade do tempo real. Quatro passos são executados antes da utilização dos classificadores para aumentar a performance, são eles: detecção do rosto humano, extração dos pontos de referência, normalização e cálculo das características correspondentes. O resultado de 70% de média obtido com o multi-SVM, segundo os autores é promissor, pois, devida as condições em que os experimentos foram realizados não eram as ideais.

Por fim [Krestinskaya & James \(2017\)](#) faz um incremento no classificador Min-Max para tentar criar um algoritmo eficiente e direto para reduzir o problema de incompatibilidade de pixels entre classes durante a classificação. O método proposto inclui a aplicação de normalização de pixel para remover offsets de intensidade seguidos de uma métrica Min-Max em um classificador de vizinho mais próximo que é capaz de suprimir outliers de características. Antes do uso do classificador é utilizado uma máscara Gaussiana e detecção de recursos calculando desvio padrão da imagem normalizada. Os resultados mostram que experimentos realizados com o algoritmo proposto tem uma acurácia de 98,57% contra 92,85% do algoritmo original, isso sem utilizar de técnicas complexas e grande poder computacional.

### 3 Metodologia

O projeto escolhido é proposto por 3 etapas principais, primeiro coletando imagens de um dataset para alimentar o treinamento. A segunda etapa também é dividida em 3 etapas, onde na primeira são importadas todas as bibliotecas e funções que serão utilizadas para o treinamento, seguindo assim com o treinamento inicial e a partir desse resultado iniciar a terceira etapa com a utilização do Keras e TensorFlow para refinar o modelo. Por fim é utilizado o modelo gerado como uma das 3 entradas para execução do programa. As outras entradas necessárias são um classificador e um arquivo que executa em tempo real para detecção das emoções.

Foi utilizado um dataset público com imagens representadas com 5 características, sendo elas, feliz, triste, surpreso, raiva e neutro. A partir dessas imagens o modelo começa a ser treinado. Com a utilização dos algoritmos de Deep Learning o modelo passa por um refinamento para melhorar sua precisão.

Para a execução em tempo real, é executado um algoritmo que dentro de um while(true) fica capturando as imagens, toda imagem capturada passa primeiro por um classificador para verificar se aquela imagem é de fato um rosto, caso sim, prossegue o algoritmo, caso contrário, descarta a captura. Após passar pelo classificador a imagem passa por tratamentos para que o modelo de predição seja executado nela. Após tudo pronto, o modelo classifica entre as 5 opções qual tem a maior possibilidade de ser da imagem no momento em que foi capturada.

## 4 Resultados

O programa efetua o que é proposto sem muitos problemas, sendo possível até ter mais de uma pessoa ao mesmo tempo na webcam e prever qual o sentimento de ambas simultaneamente. As vezes não é possível identificar o rosto dependendo da posição em que o mesmo está e tem também certa dificuldade com objetos em movimento, o que o torna inviável para o uso em câmeras em tetos por exemplo.

Comparado a outros programas de mesma função, ele apresenta uma simplicidade grande o que quer dizer que para um programa com fins de educativos é ótimo, já que além de introduzir bem o tema, dá espaço e liberdade criativa para ser mais desenvolvido, porém muito raso, o que deixa a desejar para aquele que quer se aprofundar no assunto.

Ele tem seu funcionamento fazendo previsões baseadas no modelo gerado pelo treino, utilizando a foto(frame) que é tirada da webcam em tempo real e prevendo qual das 5 emoções treinadas tem a maior probabilidade de ser, sendo assim ele é sujeito a erros, já que uma expressão mal capturada ou não conhecida obter um resultado diferente do esperado.

## Referências

- Chellappa, R., Wilson, C. L., & Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5), 705–741.
- Demétrius Alex Scardua, K. M. (2018). *Estudo da identificação de emoções através da inteligência artificial*. Retrieved 2021-03-16, from <https://multivix.edu.br/wp-content/uploads/2018/08/estudo-da-identificacao-de-emocoes-atraves-da-inteligencia-artificial.pdf>
- Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J. M., Castillo, J. C., Lozano-Monazor, E., López, M. T., ... Fernández-Sotos, A. (2016). Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics*, 64, 55-73. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1532046416301289> doi: <https://doi.org/10.1016/j.jbi.2016.09.015>
- J, A., & Geetha, M. (2017, 10). Automatic human emotion recognition in surveillance video. In (Vol. 660, p. 321-342). doi: 10.1007/978-3-319-44790-2\_15
- Krestinskaya, O., & James, A. P. (2017). Facial emotion recognition using min-max similarity classifier. In *2017 international conference on advances in computing, communications and informatics (icacci)* (p. 752-758). doi: 10.1109/ICACCI.2017.8125932
- Nguyen, B. T., Trinh, M. H., Phan, T. V., & Nguyen, H. D. (2017). An efficient real-time emotion detection using camera and facial landmarks. In *2017 seventh international conference on information science and technology (icist)* (pp. 251–255).