

Estrutura de Reconhecimento Facial com Raspberry PI para Segurança Inteligente Utilizando Técnicas de Aprendizagem de Máquina

Açucena de Gois Parente*, Iago Magalhães de Mesquita**, Sarah Frota Alves**,
Francisco Aldinei Pereira Aragão**

*Grupo IDSoft, Sobral, CE 62030-020 BR (e-mail: acucenagois@grupoidsoft.com.br).

**Eixo de Controle de Processo Industriais - Instituto Federal de Educação, Ciência e Tecnologia do Estado do Ceará (IFCE), Sobral 62042-030 BR (e-mail: [iago.magalhaes,mesquita61,sarah.frota.alves08}@aluno.ifce.edu.br](mailto:{iago.magalhaes,mesquita61,sarah.frota.alves08}@aluno.ifce.edu.br), aldinei@ifce.edu.br)

Abstract: This paper proposes an electronic security system of significant robustness, applying techniques of facial biometry. These systems are recurrent in the literature, and therefore have provided the advancement of artificial intelligence and computer vision techniques. Three techniques for facial attribute extraction and classification of images, both static and dynamic, with focus of the facial detection and recognition process were evaluated. The presented results demonstrate the relevance of the studied techniques, with accuracy above 98%. For field testing, a low-cost prototype was developed to suggest a functional and commercial solution.

Resumo: Este trabalho propõe um sistema de segurança eletrônica de significativa robustez, aplicando técnicas de biometria facial. Sistemas assim são recorrentes na literatura, que vem evoluindo junto com o avanço da inteligência artificial e visão computacional. Foram avaliadas três técnicas para extração de atributos faciais e classificação de imagens, tanto estáticas quanto dinâmicas, com foco no processo de detecção e reconhecimento facial. Os resultados apresentados demonstram a relevância das técnicas estudadas, com acurácia acima de 98%. Para o teste em campo, foi desenvolvido um protótipo de baixo custo de modo a sugerir uma solução funcional e comercial.

Keywords: Electronic security; Facial biometry; Artificial intelligence.

Palavras-chaves: Segurança eletrônica; Biometria facial; Inteligência artificial.

1. INTRODUÇÃO

O avanço da eletrônica e da computação tem proporcionado a criação de novas tecnologias que abrangem diversos setores comerciais, tais como, sistemas de segurança, análise de imagens médicas, sistemas de rastreamento, etc. Em particular, quanto aos sistemas de segurança (controle de acesso e/ou câmeras de segurança eletrônica), empresas têm buscado inovar quanto a alterações de conformidade, funcionalidade e segurança, com capacidade de reconhecer padrões complexos da imagem dos usuários, de modo a dificultar falsificação e entrada não autorizada na área protegida pelo sistema. Além disso, a utilização de sistemas computacionais aplicados ao contexto de segurança eletrônica tem provado significativo valor comercial, tendo em vista o uso em diversos segmentos empresariais de modo a proteger seus patrimônios e prover segurança a terceiros.

Em sistemas de controle de acesso, é comum a utilização da tecnologia por rádio frequência (*Radio Frequency Identification - RFID*). Porém, apesar da tecnologia apresentar robustez para a aplicação indicada, o *RFID* não fornece autenticação ao titular do cartão, uma vez que, qualquer indivíduo não autorizado, mas que detenha o cartão cadastrado pode ter acesso consentido pelo sistema, ou seja,

acesso “falso-positivo”. Logo, este trabalho propõe um sistema de biometria facial para detecção e reconhecimento de usuários, que também é integrado a um sistema de controle de acesso com *RFID*. A identificação por rádio frequência é utilizada apenas como componente complementar e não como único sistema de identificação e acesso.

É importante mencionar que a demanda por tecnologias biométricas é crescente. Tais tecnologias compreendem dispositivos de captura de informação biométrica, que através de uma base de dados e técnicas de software, realizam a manipulação da informação, identificando usuários conforme suas características físicas ou comportamentais. As principais tecnologias consistem em: reconhecimento de impressões digitais, reconhecimento de voz, reconhecimento de íris, reconhecimento facial, entre outras (SIMONYAN; ZISSERMAN, 2014). O desenvolvimento das técnicas biométricas tem sido possível devido a análise e processamento de imagens digitais, utilizando técnicas de Visão Computacional (VC) e Inteligência Artificial (IA).

A IA consiste no emprego de técnicas de aprendizagem para resolução de problemas complexos e geralmente não lineares. Tais técnicas empregam modelos matemáticos preditores para

classificação. O campo da IA tem crescido bastante nos últimos anos, principalmente em decorrência dos recentes algoritmos na área de *Machine Learning* e *Deep Learning* (Géron, 2019).

O principal problema da detecção de faces consiste em determinar se a imagem arbitrária representa uma face humana ou não, de modo a retornar as coordenadas da face reconhecida. Como solução, a recente literatura da área propõe técnicas como a *Convolutional Neural Network - CNN* (algoritmo de *Deep Learning*) e *Support Vector Machine - SVM*, com resultados superiores às técnicas tradicionais de visão computacional.

O objetivo deste trabalho é empregar as seguintes técnicas: a) *Haar Cascade*; b) *Histogram of Oriented Gradients - HOG com Support Vector Machine - SVM*; c) *Convolutional Neural Network - CNN*. com finalidade de buscar reconhecer faces em um hardware de baixo custo. Para classificação e reconhecimento das faces foi utilizada apenas uma *CNN*. A maior contribuição que este trabalho se propõe é no embarque da aplicação de forma a testar em formato de Mínimo Produto Viável - MVP.

Considerando o exposto, este artigo foi organizado da seguinte forma: (2) Revisão bibliográfica dos principais algoritmos utilizados para a detecção e o reconhecimento facial; (3) Descrição completa do sistema proposto; (4) Apresentação dos resultados; (5) Conclusão.

2. ESTADO DA ARTE

A face humana é utilizada para diferentes fins de pesquisa: identificação facial de pessoas em sistemas de segurança eletrônica, reconhecimento de expressões faciais aplicado à segurança, medicina, psicologia, etc. Para construir sistemas de reconhecimento, a detecção da face consiste na primeira etapa. Entretanto, a detecção do rosto em uma imagem estática ou dinâmica é um processo complexo, tendo em vista que envolve diferentes variações, como, variação de cor da pele, textura, etc.

Nesta seção, são apresentados diferentes técnicas ou métodos, conforme a literatura da área, para solução do processo de detecção facial.

2.1 Haar Cascade

O método *Haar Cascade* consiste em uma técnica de detecção de objetos proposta por Paul Viola e Michael Jones (VIOLA; JONES, 2001). O detector de objetos Viola-Jones conquistou grande destaque devido a sua eficiência em detectar faces com elevada taxa de acertos, baixa taxa de falso positivo e principalmente devido ao seu baixo custo computacional (ARAUJO, 2010).

Subdividido em três partes, o algoritmo de Viola-Jones primeiramente utiliza filtros de *Haar*, exemplificados na Figura 1, para representar a imagem em um espaço característico. Isto é realizado utilizando-se a “Imagem Integral” a qual é responsável por aumentar o ganho

computacional, pois as operações são realizadas sobre retângulos ao invés de operar sobre *pixels*.

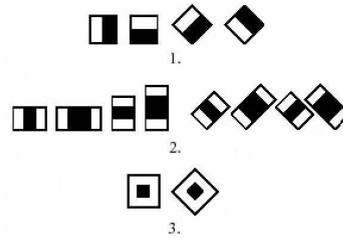


Fig. 1 Filtros de *Haar*. (1) – Recursos de Ponta; (2) – Recursos de Linha; (3) – Recursos de Centro. Fonte: Adaptado de (VIOLA; JONES, 2001).

Em um segundo passo utiliza-se um classificador baseado no método *AdaBoost*, responsável por selecionar as características mais relevantes extraídas no passo anterior (VIOLA; JONES, 2001). Em cada iteração do *AdaBoost*, um conjunto de classificadores fracos h_j é ajustado para minimizar o erro de classificação. Cada um destes classificadores corresponde a uma característica $f_j(x_n)$, onde $j = 1, \dots, J$ e J é o total de características. Dado um limiar θ_j e uma paridade p_j , a regra de classificação pode ser dada por:

$$h_j(x_n) = \{1, \text{ se } p_j f_j(x_n) > p_j \theta_j, \text{ caso contrário.} \quad (1)$$

onde a paridade p_j indica a direção da desigualdade (ARAUJO, 2010).

Para obter um melhor desempenho, utilizam-se classificadores em cascata. A Figura 2 ilustra o funcionamento dos classificadores em cascata.

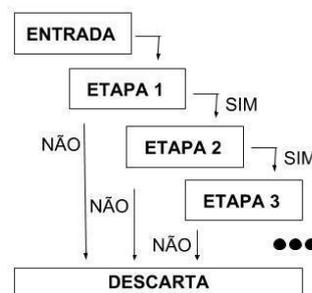


Fig. 2 Fluxograma do algoritmo em cascata. Fonte: Autor.

Para realizar a detecção utilizam-se as características, as quais são escalonadas em diversos tamanhos já que não se sabe o tamanho nem a posição da face na imagem alvo, e então são aplicadas em todas sub-janelas da imagem para verificar se a característica está presente ou não na imagem. Há uma série de classificadores já definidos e salvos em arquivos *XML* que podem ser utilizados diretamente por ferramentas como *OpenCV - Open Source Computer Vision Library*.

2.2 HOG + SVM

O descritor baseado em *Histogram of Oriented Gradients (HOG)* foi proposto por Dalal e Triggs, e é baseado na

avaliação de histogramas locais normalizados dos gradientes orientados da imagem (Guo et al., 2021). O vetor de características é formado pela concatenação dos histogramas normalizados dos gradientes orientados presentes em uma janela de detecção.

Para o treinamento do classificador, as imagens são cortadas em um tamanho fixo. Quando o *HOG* de um rosto é extraído, as características da imagem são adquiridas de forma que, mesmo sob diferentes condições, como luminosidade, o descritor de imagem mantenha-se o mesmo, ou o mais próximo possível. A extração do *HOG* de uma imagem começa com a repartição da imagem em pequenas regiões espaciais denominadas como células, que podem ser ajustadas de acordo com a necessidade (PANG et al., 2011). Também nota-se que, tendo os histogramas, eles são submetidos ao *SVM* para classificá-los.

O *Support Vector Machine (SVM)* é um classificador linear, independentemente da distribuição dos dados serem lineares ou não (Wu et al., 2004). É um classificador de aprendizado supervisionado utilizado em diversas aplicações de reconhecimento de padrões, como: imagens, fala e previsões de séries temporais.

Considerando dados de treinamento de duas classes em um modelo linear, o objetivo do *SVM* é definir um hiperplano que classifique corretamente as classes (Wu et al., 2004). Para esta tarefa podem existir diversos hiperplanos que satisfaçam a classificação. Entretanto, a escolha do hiperplano deve ser feita com o intuito de aumentar a generalização do classificador, dando a capacidade de classificar dados que não estavam presentes no treinamento. Assim o *SVM* modela um hiperplano que cria a maior margem possível entre duas classes. Este hiperplano tem a mesma distância em relação aos pontos mais próximos entre duas classes (PANG et al., 2011).

O processo de extração do histograma e da classificação com o *SVM* pode ser visto na Figura 3.

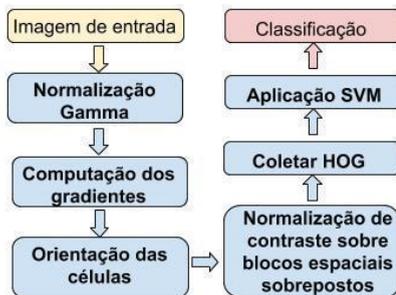


Fig. 3 Fluxograma de funcionamento do algoritmo *HOG+SVM*. Fonte: Autor.

2.3 CNN

As *Convolutional Neural Network (CNNs)*, como muitas técnicas de aprendizado de máquina, são baseadas em arquiteturas biológicas encontradas na natureza. As *CNNs* provêm de um trabalho realizado nos anos 60 por Hubel e

Wiesel que, através de experimentos com gatos e primatas, descobriram que a capacidade visual é organizada em um conjunto de células sensíveis a regiões chamadas campos receptivos. Estas células podem ser classificadas em três tipos: simples, complexas e super complexas. As células simples são ativadas por padrões simples e as demais por padrões de alta complexidade. A partir deste estudo é possível inferir que a combinação de padrões gera os objetos e cenas, produzindo assim a representação visual (Goodfellow et al., 2016). Dessa forma, as *CNNs* surgiram e tentam representar esta ideia através de dados em duas dimensões, assim como as imagens. Sendo assim, as *CNNs* se tornam técnicas promissoras para o reconhecimento de imagens (Goodfellow et al., 2016), sendo introduzidas na literatura oficialmente por (LECUN et al., 1998).

Na figura 4 é possível ver o funcionamento genérico de uma *CNN*.

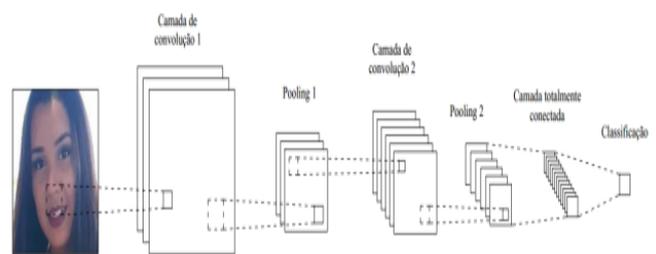


Fig. 4 Rede Neural Convolutiva Genérica. Fonte: Adaptado de (Goodfellow et al., 2016).

2.3.1 CAMADA DE CONVULUÇÃO

A camada de convolução é a camada de construção da rede. Nesta camada os parâmetros são filtros (*kernels*) com baixa receptividade, mas que se estendem por toda a imagem. Cada um desses filtros, aplicado a toda a imagem, consiste em um mapa de características (Goodfellow et al., 2016). Eles dão origem a uma estrutura conectada localmente que percorre toda a extensão do volume de entrada.

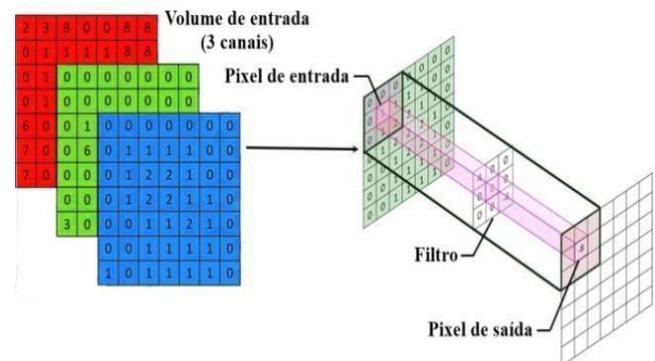


Fig. 5 – Processo de convolução. Fonte: Adaptado de (Goodfellow et al., 2016).

2.3.2 POOLING

O *pooling* resume as saídas de grupos de neurônios vizinhos no mesmo filtro, ou seja, esta camada serve para reduzir progressivamente o tamanho espacial da representação intermediária, como mostrado na figura 6. Assim, o uso do *pooling* reduz a quantidade de parâmetros e computação, portanto, também combate o *overfitting* (Goodfellow et al., 2016).

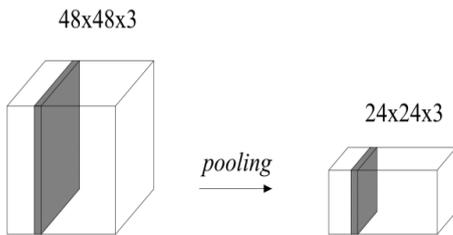


Fig. 6 – Processo de *Pooling*. Fonte: Adaptado de (Goodfellow et al., 2016).

2.3.3 ReLU – UNIDADES LINEARES RETIFICADAS

As ReLUs são camadas que aplicam funções de ativação de não-saturação. Essas funções aumentam a não linearidade da função de decisão da rede, sem afetar os campos receptivos e pode ser calculada por:

$$f(x) = \max(0, x) \quad (2)$$

2.3.4 CAMADA TOTALMENTE CONECTADA

É onde, de fato, inicia-se o aprendizado da rede. Nesta camada os mapas de características são convertidos de duas dimensões para uma e, assim como nas redes Multi Camadas – MLP (*Multi-Layer Perceptron*), todos os neurônios estão conectados a todos os neurônios ativos da camada anterior. Assim, é possível afirmar que esta parte da rede é similar às Redes Neurais regulares (LECUN et al., 1998). Conseqüentemente, suas ativações podem ser computadas como a multiplicação de matrizes somadas a um *bias*.

Essas camadas são formadas por neurônios, ou seja, uma analogia a unidades de processamento, e o termo “totalmente conectado” significa que todos os neurônios da camada anterior estão conectados a todos os neurônios da camada seguinte (ARAUJO, 2010). Em termos matemáticos, um neurônio pode ser descrito pelas equações (3) e (4):

$$u_k = \sum_{j=1}^m w_{kj} x_j, \quad (3)$$

$$y_k = \varphi(u_k + b_k), \quad (4)$$

em que x_1, x_2, \dots, x_m são os m sinais de entrada, w_{k1}, w_{k2}, w_{km} são os pesos sinápticos do neurônio k , e b_k correspondente ao viés ou *bias*, responsável por realizar o deslocamento da função de ativação definida por $\varphi(\cdot)$.

2.4 CURVA ROC

A representação gráfica de curvas ROC (*Receiver Operating Characteristic*) é uma técnica que serve para visualizar, organizar e selecionar classificadores com base no seu desempenho. Um melhor desempenho está intimamente relacionado com uma maior capacidade discriminante (ARAUJO, 2015).

A análise da curva ROC teve sua origem na teoria de detecção de sinal. Sendo assim, é baseada em duas quantidades que contêm toda a informação, uma designada por taxa de verdadeiros positivos, e outra designada por taxa de falsos positivos. Além disso, pode-se definir a curva ROC de duas formas diferentes, uma mais restritiva, em termos de razão de verossimilhança, e outra mais geral, em termos de variação de decisão (EGAN, 1975).

A curva ROC é representada num espaço bidimensional, onde no eixo do x estão representados os valores da taxa de falsos positivos e no eixo do y estão representados os valores da taxa de verdadeiros positivos. O gráfico ROC retrata a relação entre os benefícios e os custos. Vários pontos são importantes no espaço ROC, entre eles, o ponto inferior esquerdo (0,0) ligado ao ponto superior direito (1,1) que representa uma estratégia completamente aleatória, isto é, não classifica como falso positivo nem como verdadeiro positivo, e uma curva que tende para o ponto (0,1) representa uma classificação perfeita (ARAUJO, 2015).

3. PROTÓTIPO DO SISTEMA PROPOSTO

Como mencionado anteriormente, este trabalho propõe um sistema de controle de acesso combinando tecnologia RFID e reconhecimento facial, de modo a garantir dois níveis de segurança.

A figura 7 apresenta o protótipo desenvolvido em laboratório. O item 1 corresponde a plataforma *Raspberry Pi*, utilizada para embarcar os algoritmos de identificação e reconhecimento facial, enquanto o item 2 consiste no sistema de radiofrequência.



Fig. 7 Protótipo do sistema onde (1) é onde funciona todo o protótipo do reconhecimento e (2) é o sistema com tecnologia RFID. Fonte: Autor.

O diagrama simplificado e em blocos da plataforma *Raspberry Pi* é apresentado na figura 8. Vale ressaltar que para a instalação física do sistema não é necessário *mouse* e teclado. O sistema já funcionará automaticamente. Outro ponto importante é que foi utilizada a Câmera *Raspberry Pi* v2 com 8MP e sensor *Sony IMX219*, por se pensar em um sistema de baixo custo.

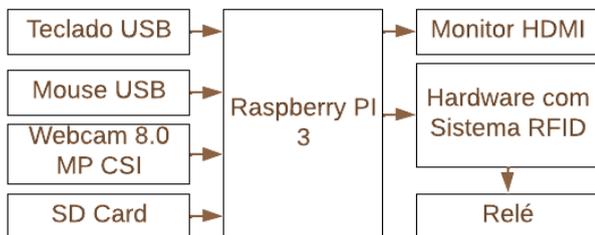


Fig. 8 Fluxograma do protótipo do sistema de controle de acesso. Fonte: Autor.

3.1 OPERAÇÃO DO SISTEMA

A metodologia abordada neste trabalho segue basicamente as seguintes etapas: detecção facial, alinhamento dos pontos faciais e reconhecimento facial.

3.1.1 DETECÇÃO FACIAL

A fim de selecionar o método mais eficiente de detecção, para este trabalho foram comparados 3 métodos: *Haar Cascade* da *OpenCV*, *HOG+SVM* da *Dlib* e uma *CNN* também implementada com a *Dlib*.

Como comentado anteriormente, o detector de faces da *OpenCV* já foi pré-treinado com 4916 imagens de faces e 9544 imagens de não faces (VIOLA; JONES, 2001), seguindo o algoritmo do classificador *Adaboost*.

O *Dlib* é um *kit* de ferramentas aberto, que contém algoritmos e ferramentas de aprendizado de máquina com a finalidade de criar *softwares* complexos e resolver problemas do mundo real. Ele vem sendo usado tanto na indústria quanto na academia em uma ampla gama de domínios incluindo robótica, telefones celulares e grandes ambientes de computação de alto desempenho.

Para a detecção com a técnica *HOG+SVM* foi usado a *Dlib* configurada para encontrar rostos humanos mais ou menos voltados para frente. É criado usando uma ferramenta que executa o classificador sobre uma pirâmide da imagem (FELZENSZWALB et al., 2009).

Para o último teste de detecção foi utilizado também a ferramenta *Dlib*. Foi usado um modelo pré-treinado com imagens retiradas de vários *datasets* abertos, como *ImageNet*, *AFLW*, *Pascal VOC*, o conjunto de dados *VGG*, entre outros.

A detecção feita pelos três métodos pode ser vista nos dois exemplos das figuras 9, sendo uma sem escala e a outra com escala, respectivamente.

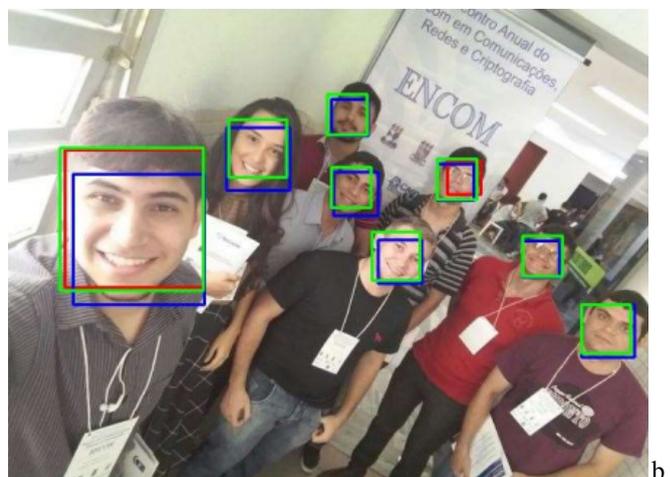
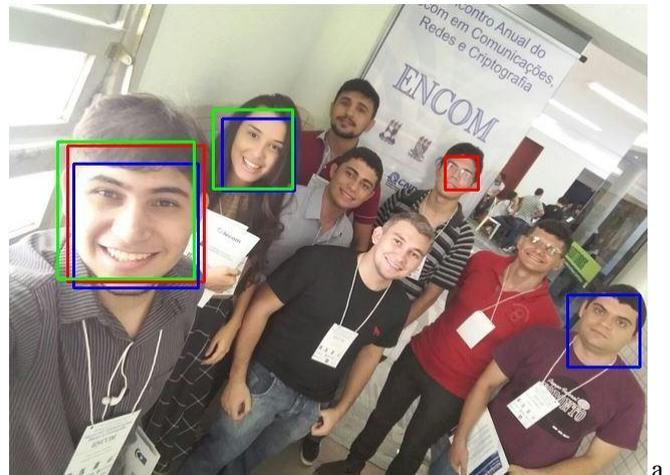


Fig. 9 a) Detecção facial sem escala (Vermelho - *Haar*; Azul - *HOG*; Verde - *CNN*). Fig. 9 b) Detecção facial com escala (Vermelho - *Haar*; Azul - *HOG*; Verde - *CNN*). Fonte: Autor.

3.1.2 ALINHAMENTO DOS PONTOS FACIAIS

Esse processo foi pensado com o intuito de facilitar o reconhecimento facial. Nesta etapa, logo após a detecção da face, ocorre a marcação dos pontos faciais principais, que para este trabalho, são os cantos externo e interno dos dois olhos e logo abaixo do nariz, totalizando 5 pontos para o preditor.

Com o preditor, foi marcado na face detectada os cinco pontos (figura 10.a). Em seguida é feito o alinhamento, que nada mais é do que colocar os quatro pontos relacionados aos olhos, no mais aproximado de uma reta horizontal e validar com uma reta perpendicular à anterior em direção ao ponto do nariz (figura 10.b).

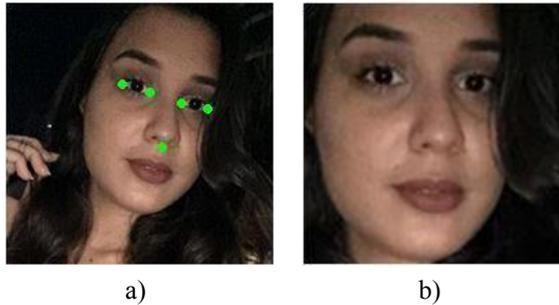


Fig. 10 a) face de entrada. b) Face alinhada. Fonte: Autor.

3.1.3 RECONHECIMENTO FACIAL

Baseado no estudo bibliográfico pode-se notar que a *CNN* requer um bom processador para treiná-la, portanto todo o treinamento, tanto para detecção quanto para o reconhecimento, foi feito no *Google Colab*. O mesmo é um serviço de nuvem gratuito, adequado à programação na linguagem *Python*, com *GPU Tesla K80* disponível. O serviço foi criado para pesquisadores de IA e desenvolvedores de aplicativos de aprendizagem profunda.

O reconhecimento foi feito utilizando o modelo de rede *CNN* da ferramenta *Dlib*. Esse modelo é uma rede *ResNet* com 29 camadas Convolucionais. É essencialmente uma versão da rede *ResNet-34* proposto por (HE et al., 2016) com algumas camadas removidas e o número de filtros por camada reduzidos pela metade. Este modelo tem uma precisão de 99,38%. Isso é comparável a outros modelos de última geração e significa que, dadas duas imagens faciais, ele prevê corretamente se as imagens são da mesma pessoa 99,38% do tempo. Essa rede mantém o padrão de classificação com um treino de, no mínimo, três imagens por pessoa.

3.1.4 APLICAÇÃO

Em sequência foi decidido, sabendo dos resultados, usar a detecção facial feita por *HOG+SVM*, o alinhamento dos pontos do rosto e o reconhecimento facial pela *CNN*.

Implementou-se, então, um *script* para funcionar com imagens de entradas vindas de uma *webcam* associada ao *Raspberry Pi*.

4. RESULTADOS

Com todos os algoritmos de detecção implementados, os mesmos foram testados com 15 imagens para que possa ser possível gerar a matriz de confusão (Tabela 1). Das imagens, 9 eram positivas e 6 eram negativas.

Tabela 1 – Matriz de confusão. Fonte: Autor.

		Predito/Real	Face	Não Faces
Sem escala	Haar	Face	5	2
		Não Face	4	4
		Total	9	6

	HOG	Face	3	0
		Não Face	6	6
		Total	9	6
	CNN	Face	3	0
		Não Face	6	6
		Total	9	6
Com escala	Haar	Face	5	2
		Não Face	4	4
		Total	9	6
	HOG	Face	9	0
		Não Face	0	6
		Total	9	6
	CNN	Face	7	0
		Não Face	2	6
		Total	9	6

Comparando as técnicas de detecção facial implementadas neste trabalho foi possível gerar a curva *ROC* (figura 11) para uma melhor análise de eficiência.

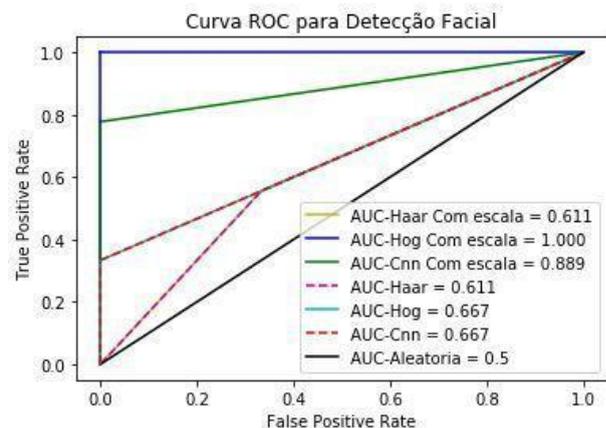


Fig. 11 Curva *ROC* da detecção facial. Fonte: Autor.

Analisando a curva é possível notar que o algoritmo *HOG+SVM* obteve um melhor resultado quando usado com escala, já que, de acordo com a curva, obtém-se uma acurácia de valor 1. Também para fins comparativos, foram analisados a precisão, sensibilidade e o coeficiente de phi ou *MCC* – avalia a qualidade da classificação, baseado nas equações:

$$P = \frac{T_p}{T_p + F_p} \quad (5)$$

$$S = \frac{T_p}{T_p + F_n} \quad (6)$$

$$MCC = \frac{T_p * T_n - F_p * F_n}{\sqrt{(T_p + F_p) * (T_p + F_n) * (T_n + F_p) * (T_n + F_n)}} \quad (7)$$

Onde:

P – precisão

S – sensibilidade

MCC – coeficiente de phi

Tp – verdadeiros positivos

Fp – falsos positivos

Fn – falsos negativos

Tn – verdadeiros negativos

Assim, foi possível gerar a tabela 2 com os valores calculados. Mesmo levando esses novos valores em consideração, o *HOG+SVM* ainda seria uma melhor escolha para este trabalho.

Tabela 2 – Resultados para Análise. Fonte: Autor.

	Sem escala			Com escala		
	Haar	HOG	CNN	Haar	HOG	CNN
Acurácia	0.61	0.66	0.66	0.61	1	0.88
Precisão	0.71	1	1	0.71	1	1
Sensitividade	0.55	0.33	0.33	0.55	1	0.77
MCC	0.21	0.40	0.40	0.21	1	0.76

Para o reconhecimento facial, a rede foi treinada com três imagens de cada pessoa, e testada em 6 fotos de grupo havendo resposta favorável da rede em todas as imagens, como mostram os exemplos das figuras 12 e 13.



Fig. 12 Reconhecimento Facial (exemplo 1). Fonte: Autor.



Fig. 13 Reconhecimento Facial (exemplo 2). Fonte: Autor.

A figura 14 mostra a detecção e o reconhecimento ocorrendo já no protótipo.

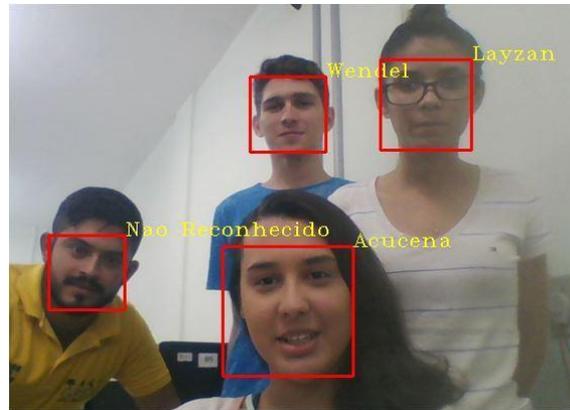


Fig. 14 Reconhecimento de faces cadastradas e não cadastradas. Fonte: Autor.

5. CONCLUSÃO

Neste trabalho é proposto um sistema de reconhecimento de características faciais. Esse sistema consistiu de um conjunto de etapas listadas a seguir:

- Detecção da face com o algoritmo *HOG+SVM*
- Alinhamento facial para facilitar a classificação;
- Reconhecimento da face feito por uma rede *CNN*.

A primeira etapa do sistema proposto é comparada com dois outros métodos: *Haar Cascade* e *CNN*. Para todos os casos, o treinamento e os testes foram feitos sob as mesmas condições.

O protótipo está em pleno funcionamento e respondendo de forma favorável ao esperado (com latência de aproximadamente 10 segundos), ou seja, detectando as faces cadastradas e, em conjunto com o sistema *RFID*.

É importante mencionar que, o sistema só permite acesso a determinado ambiente, quando ocorre a correspondência entre biometria facial e o cartão *RFID*.

AGRADECIMENTOS

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Campus Sobral, Pró-reitoria de Pesquisa e Inovação e ao Laboratório de Eletrônica e Sistemas Embarcados pelo apoio, financiamento e auxílio no desenvolvimento.

REFERÊNCIAS

ARAUJO, G. Algoritmo para reconhecimento de características faciais baseado em filtros de correlação. Dissertação do curso de Pós-Graduação em Engenharia Elétrica COPPE. Rio de Janeiro. Pág 65. 2010.

- ARAUJO, J. Avaliação do desempenho de indicadores com base na metodologia ROC [Receiver Operating Characteristic]. Dissertação de Mestrado em Bioinformática. Braga - Portugal. Pág 72. 2015.
- EGAN, J. Signal Detection Theory and ROC-analysis Academic Press series in cognition and perception Cognition and Perception Series. Academic Press. 1975. ISBN: 0122328507, 9780122328503.
- FELZENSZWALB, P. et al. Object Detection with Discriminatively Trained Part Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 32, n. 9, p. 1–20, 2009.
- GÉRON, A. Mãos à Obra: Aprendizado de Máquina com Scikit-learn & TensorFlow. Alta Books, Rio de Janeiro, Brasil (2019).
- GOODFELLOW, I. et al. A.: Deep learning. 1 st edn. The MIT press, Cambridge, Massachusetts London, England (2016).
- GUO, S; LIU F; et al. HSPOG: An Optimized Target Recognition Method Based on Histogram of Spatial Pyramid Oriented Gradients. Tsinghua Science and Technology. 2021, 26 (4): 475-483. <https://doi.org/10.26599/TST.2020.9010011>.
- HE, K. et al. Identity mappings in deep residual networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 9908 LNCS, p. 630–645, 2016.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, v. 86, n. 11, p. 2278–2323, 1998.
- LU, W; YANG, M. Face Detection Based on Viola-Jones Algorithm Applying Composite Features. 2019 International Conference on Robots & Intelligent System (ICRIS). DOI 10.1109/ICRIS.2019.00029.
- PANG, Y. et al. Efficient HOG human detection. Signal Processing, v. 91, n. 4, p. 773–781, 2011.
- SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- WU, T; LIN, C; WENG, R. Probability estimates for multi-class classification by pairwise coupling. Journal of Machine Learning Research (JMLR), Yoram Singer, aug. 2004.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, v. 1, p. I-511-I-518, 2001.