Cocou Médard Tossou

Marcos Costa Braga

Estudo e implementação de técnicas de processamento de imagens para reconhecimento facial

BRASÍLIA, DF

2020

Cocou Médard Tossou

Marcos Costa Braga

Estudo e implementação de técnicas de processamento de imagens para reconhecimento facial

Monografia, apresentada no Instituto de

Educação Superior de Brasília (IESB), para

obtenção do grau de Bacharelado em Ciência da Computação sob a orientação do Prof. Thiago Raposo Milhomem.

BRASÍLIA, DF

2020

Cocou Médard Tossou

Marcos Costa Braga

Estudo e implementação de técnicas de processamento de imagens para reconhecimento facial

**BANCA EXAMINADORA - APROVADO POR:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Thiago Raposo Milhomem - Orientador

Centro Universitário Instituto de Educação Superior de Brasília, DF

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. José Márcio

Centro Universitário Instituto de Educação Superior de Brasília, DF

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof.

Centro Universitário Instituto de Educação Superior de Brasília, DF

Brasília, DF, \_\_\_\_\_\_\_\_ de\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ de 2020.

Esse trabalho é dedicado à Deus,

que nos deu sabedoria para escrever cada página

e iluminou os caminhos para superar as adversidades.

Sem Ele, seria impossível.

**AGRADECIMENTOS**

Somos gratos à Deus, que encheu nossos corações de luz e contribuiu com a nossa cumplicidade. Agradecemos aos familiares, por acreditarem e apoiarem nosso sonho, muito obrigado pelo carinho. Aos nossos colegas de sala, muito obrigado pela ajuda em vários momentos de dificuldade. E, por fim, obrigado a Julie Larissa e a Natália Wesgueber por todo apoio, paciência e companheirismo nesses meses de muito trabalho.

**RESUMO**

Neste trabalho, pretende-se explorar os métodos computacionais de reconhecimento facial. Uma Rede Neural Convolucional (CNN) será criada e treinada com o objetivo de selecionar imagens que contenham faces, e somente estas imagens serão enviadas para o reconhecimento facial, maximizando a assertividade do reconhecimento. Propõem-se uma abordagem para o problema aplicando a Transformada Discreta do Cosseno (DCT) em um rosto para reconhecer suas características e compará-las com as características dos rostos previamente cadastrados, através da Relação Sinal-Ruído de Pico (PSNR). O objetivo do trabalho é realizar o reconhecimento facial com razoável confiabilidade e assertividade, a partir dos métodos escolhidos.

**Palavras-chaves:** Detecção Facial, Reconhecimento Facial, Redes Neurais Convolucionais, Transformada Discreta do Cosseno, Relação Sinal-Ruído de Pico.

**ABSTRACT**

In this project, we intend to explore computational methods of facial recognition. A Convolutional Neural Network (CNN) will be created and trained to select images that contain faces, then only these images will be sent for facial recognition, maximizing the assertiveness of the recognition. An approach to the problem is proposed by applying the Discrete Cosine Transform (DCT) on a face to recognize its characteristics and compare them with the characteristics of previously registered faces, using the Peak Signal-Noise Ratio (PSNR). The objective of this work is to perform facial recognition with reasonable reliability and assertiveness, based on the chosen methods.

**Keywords:** Facial Detection, Facial Recognition, Convolutional Neural Networks, Discrete Cosine Transform, Peak Signal-to-Noise Ratio.

LISTA DE FIGURAS

[Figura 1 – Diagrama simplificado de um corte transversal do olho humano 13](#_Toc39680131)

[Figura 2 – Distribuição de Sensores 15](#_Toc39680132)

[Figura 3 – Exemplo do processo de aquisição de uma imagem digital 16](#_Toc39680133)

[Figura 4 – Imagens indexadas 17](#_Toc39680134)

[Figura 5 – Imagem *truecolor* 18](#_Toc39680135)

[Figura 6 – Imagem em escala de cinza 18](#_Toc39680136)

[Figura 7 – Representação simplificada de um neurônio 21](#_Toc39680137)

[Figura 8 – Modelo matemático de um neurônio 21](#_Toc39680138)

[Figura 9 – Diagrama em blocos da aprendizagem 22](#_Toc39680139)

[Figura 10 – Representação simplificada de uma rede neural artificial 23](#_Toc39680140)

[Figura 11 – Retropropagação 24](#_Toc39680141)

[Figura 12 – Rede neural Convolucional 24](#_Toc39680142)

[Figura 13 – Fluxograma do programa 27](#_Toc39680143)

[Figura 14 – Exemplo do Imagens Utilizadas 28](#_Toc39680144)

[Figura 15 – Exemplo de pré-processamento de imagem 29](#_Toc39680145)

[Figura 16 – Precisão e Perda do Conhecimento 29](#_Toc39680146)

[Figura 17 – Exemplo de falsos positivos 30](#_Toc39680147)

[Figura 18 – Exemplo de imagens Utilizadas para teste 31](#_Toc39680148)

[Figura 19 – Compressão da imagem 32](#_Toc39680149)

**LISTA DE ABREVIATURAS E SIGLA**

**CCD** – *Discrete Cosine Transform* (Transformada Discreta do Cosseno)

**CNN** – *Convolutional Neural Network* (Rede Neural Convolucional)

**DCT** – *Discrete Cosine Transform* (Transformada discreta do cosseno)

**MSE** – *Mean Square Error* (Erro Quadrático Médio)

**NFC** – *Near Field Communication* (Comunicação por campo de proximidade)

**PSNR** – *Peak Signal-to-Noise Ratio* (Relação sinal-ruído de pico)

**RGB** – *Red, Green and Blue* (Vermelho, Verde e Azul)

**SUMÁRIO**

[1 INTRODUÇÂO 11](#_Toc39680150)

[1.1 Objetivos 12](#_Toc39680151)

[1.1.1 Geral 12](#_Toc39680152)

[1.1.2 Específicos 12](#_Toc39680153)

[1.2 Organização do trabalho 12](#_Toc39680154)

[2 FUNDAMENTAÇÃO TEÓRICA 13](#_Toc39680155)

[2.1 Visão computacional 13](#_Toc39680156)

[2.1.1 A estrutura do olho humano 13](#_Toc39680157)

[2.1.2 Sensores e aquisição de imagens 15](#_Toc39680158)

[2.2 Representação de imagens digitais 16](#_Toc39680159)

[2.2.1 Imagens indexadas 16](#_Toc39680160)

[2.2.2 Imagem *truecolor* 17](#_Toc39680161)

[2.2.3 Imagem em escala de cinza 18](#_Toc39680162)

[2.3 Biometria 19](#_Toc39680163)

[2.4 Detecção Facial 19](#_Toc39680164)

[2.5 Redes Neurais 20](#_Toc39680165)

[2.5.1 Processos de aprendizagem 21](#_Toc39680166)

[2.5.2 Perceptron 22](#_Toc39680167)

[2.5.3 Redes Neurais de Múltiplas Camadas 23](#_Toc39680168)

[2.5.4 Redes Neurais Convolucionais 24](#_Toc39680169)

[2.6 Reconhecimento Facial 25](#_Toc39680170)

[2.7 Transformada Discreta do Cosseno 25](#_Toc39680171)

[2.8 Relação Sinal-Ruído de Pico 26](#_Toc39680172)

[3 DESENVOLVIMENTO E RESULTADOS OBTIDOS 27](#_Toc39680173)

[4 CONCLUSÃO 33](#_Toc39680174)

[4.1 Sugestão para trabalhos futuros 33](#_Toc39680175)

# INTRODUÇÂO

O reconhecimento facial é utilizado como uma forma de autenticação e identificação de pessoas. Assim como a impressão digital, a face é um sistema biométrico, ou seja, cada face é única e pode ser usada para diferenciar os seres vivos. As tecnologias de reconhecimento biométrico estão sendo cada vez mais buscadas para aumentar a segurança e autenticidade.

Um estabelecimento com sistema de autenticação por cartão de Comunicação por Campo de Proximidade (NFC) por exemplo, pode ser burlado quando não monitorado por um ser humano ou por outro sistema de reconhecimento biométrico. Pois qualquer um que estiver com este cartão NFC, mesmo que não seja o dono dele, poderá acessar os locais liberados por aquele cartão NFC.

O reconhecimento facial é um processo mais complexo que o reconhecimento de impressão digital e de retina, pois vários fatores, como objetos na frente do rosto, baixa luminosidade ou inclinação da posição do rosto, podem atrapalhar a assertividade do sistema.

Para auxiliar no reconhecimento facial, a detecção facial faz um pré-processamento da imagem coletado, com o objetivo verificar se há faces na imagem e mandar para o algoritmo de reconhecimento facial a posição da face, facilitando assim a extração de características.

Neste projeto a detecção facial será realizada por uma Rede Neural Convolucional (CNN). A CNN será treinada de forma supervisionada para que ela detecte faces nas imagens apresentadas.

O reconhecimento facial será realizado através da comparação de características, obtidas através da Transformada Discreta do Cosseno (DCT).

## Objetivos

Apresentam-se, a seguir, os objetivos gerais e específicos desse trabalho.

### Geral

O objetivo deste trabalho é o estudo e a implementação de um método de reconhecimento facial, propondo-se implementar uma CNN e treiná-la utilizando-se um banco de imagens para a detecção facial. Em conjunto, uma abordagem por PSNR para reconhecimento facial, utilizando-se da DCT para captura das características da face.

### Específicos

* Implementar uma CNN e treiná-la para a detecção de face;
* Realizar a extração de características faciais para o reconhecimento facial;
* Realizar o reconhecimento facial com razoável confiabilidade e assertividade.

## Organização do trabalho

Esse trabalho está organizado em 04 capítulos, além desse, nos quais serão abordados, resumidamente, os seguintes tópicos:

* **Capítulo 2** – **Fundamentação Teórica**: Conceitos necessários para a compreensão do trabalho;
* **Capítulo 3** – **Desenvolvimento e Resultados Obtidos**: Demonstração do programa desenvolvido e apresentação dos testes realizados;
* **Capítulo 4** – **Conclusão**: Comparação dos resultados esperados e obtidos;

# FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo demonstrar os principais conceitos necessários para o entendimento do trabalho.

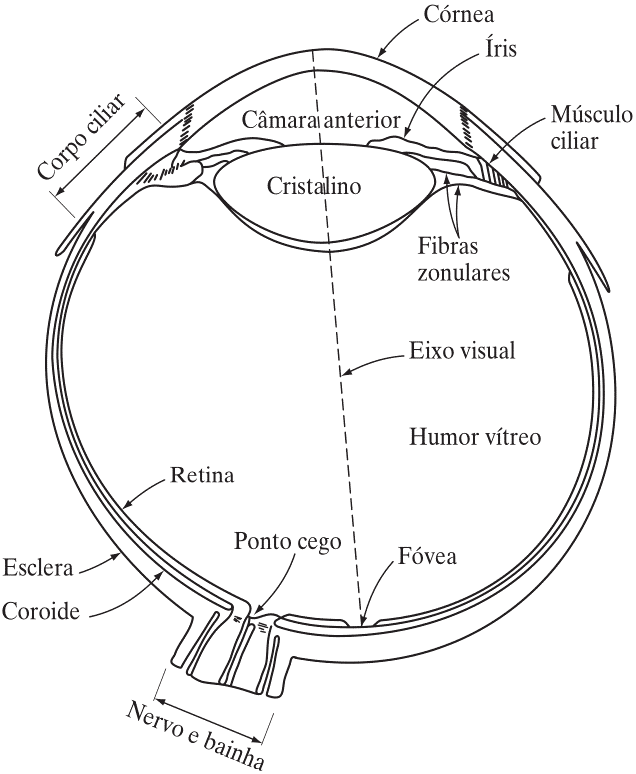
## Visão computacional

O objetivo deste capítulo é demonstrar sobre os métodos usados para a captura de imagens por um olho humano para a interpretação visual e por uma câmera digital para o processamento digital.

### A estrutura do olho humano

O olho tem um formato parecido com uma esfera, como ilustrado na Figura 1, e tem o seu diâmetro médio de 20 mm. Possuem três membranas que os envolvem, são eles: a córnea, a coroide e a retina. (GONZALEZ; WOODS, 2010)

Figura 1 – Diagrama simplificado de um corte transversal do olho humano



Fonte: (GONZALEZ; WOODS, 2010).

A córnea, um tecido resistente e transparente, cobre a superfície anterior do olho e juntamente a ela, tem a esclera, membrana opaca, que envolve o restante do globo ocular. Abaixo da esclera tem a coroide que é altamente vascularizada e tem como função nutrir o olho. A sua cobertura é bastante pigmentada, auxiliando na redução da luminosidade que entra no olho. (GONZALEZ; WOODS, 2010)

Na região anterior da coroide, tem-se o corpo ciliar e a íris, que faz os movimentos de contração e relaxamento para controlar a luminosidade que entra no olho. A íris tem uma região central, chamado de pupila, e tem o seu tamanho aproximado de 2 a 8 mm de diâmetro. A íris é pigmentada, ou seja, a parte colorida do olho. (GONZALEZ; WOODS, 2010)

Na região posterior da Iris tem-se o cristalino, sua formação é composta de camadas de células fibrosas que são sustentadas por fibras que se ligam ao corpo ciliar. O cristalino tem uma coloração levemente amarelada, que fica mais aparente com a idade. A opacidade do cristalino pode ser provocada pela doença catarata causando cegueira total ou parcial. As proteínas contidas no cristalino são responsáveis por absorver parte da luz infravermelha e ultravioleta. (GONZALEZ; WOODS, 2010)

Ao focar em um objeto, a luz que é emitida e/ou refletida deste, cria uma imagem na retina, membrana mais interna do olho. A visão é possibilitada pela distribuição de receptores discretos de luz pela retina, sendo eles cones e bastonetes. (GONZALEZ; WOODS, 2010)

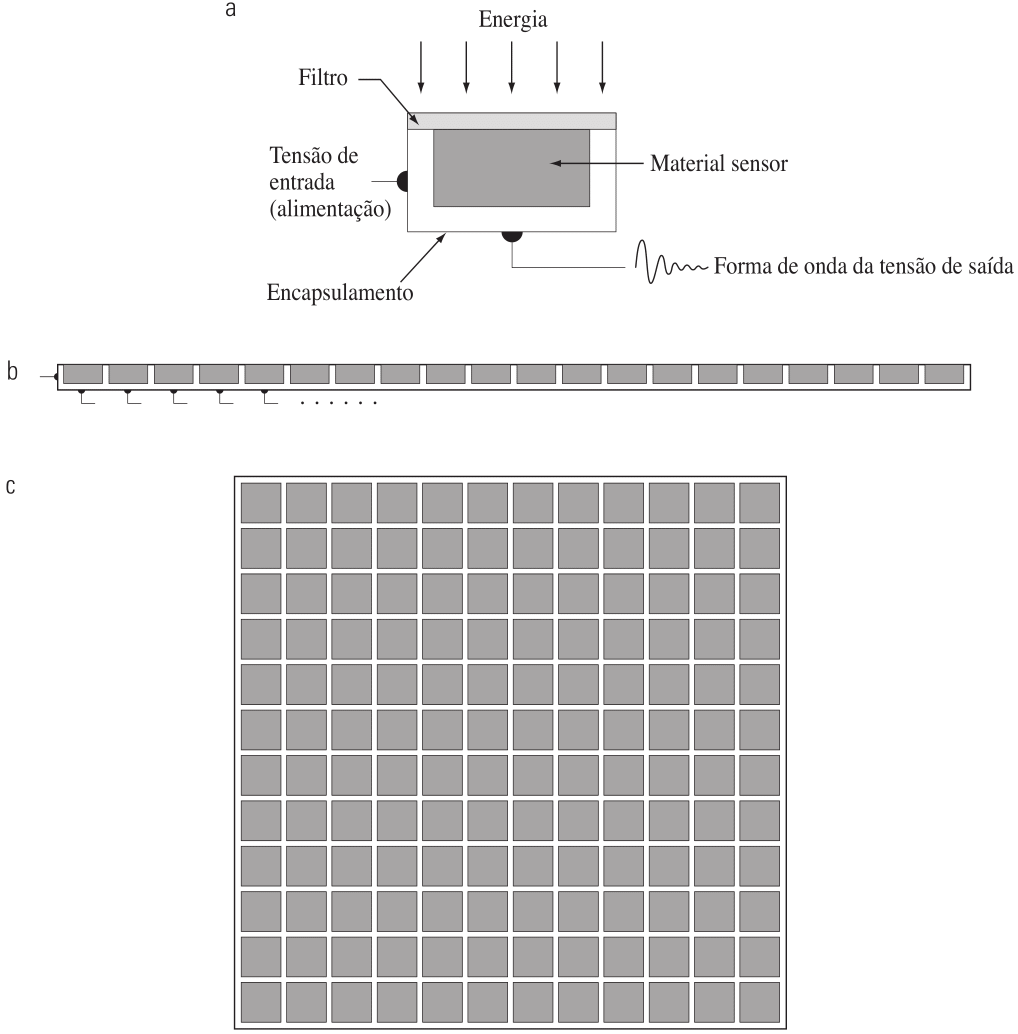
A variação de cones em cada olho é de 6 a 7 milhões, posicionados principalmente na fóvea, região central da retina, sendo altamente sensíveis às cores. Cada um destes cones é conectado em sua própria terminação nervosas, assim podendo discernir pequenos detalhes. Esse tipo de visão pode ser chamado de fotópica ou luz clara. (GONZALEZ; WOODS, 2010)

A quantidade de bastonetes é bem maior que a de cones, cerca de 75 a 150 milhões. Estes são distribuídos na superfície da retina. Por terem uma área extensa e estarem ligados a uma única terminação nervosa, é reduzido a quantidade de detalhes discernidos por eles. A função dos bastonetes é dar uma noção geral do campo de visão, não proporciona visão colorida e são bem sensíveis a baixa luminosidade. Assim se um objeto é colorido quando iluminado pela luz do sol, pode ficar descolorido quando iluminado pela luz da lua cheia. Esse tipo de visão pode ser chamado de visão escotópica ou visão de luz escura. (GONZALEZ; WOODS, 2010)

### Sensores e aquisição de imagens

Os sensores para a captura da imagem podem ser distribuídos de diversas formas e quantidade. Na Figura 2 está ilustrado 3 formas de distribuição. Na Figura 2(a) se tem a utilização de somente 1 sensor que precisa percorrer a imagem tanto na vertical quanto na diagonal para obter a imagem completa. No arranjo dos sensores de linha, mostrado na Figura 2(b), os sensores têm que percorrer a imagem somente na vertical ou somente na horizontal para gerar a imagem completa. O arranjo predominante encontrado nas câmeras digitais é de sensores individuais em forma de matriz bidimensional, como mostra a Figura 2(c). A vantagem de os sensores serem distribuídos em forma de matriz bidimensional, é que uma imagem completa pode ser obtida sem não a necessidade dos sensores se moverem tendo assim uma captura mais rápida. (GONZALEZ; WOODS, 2010)

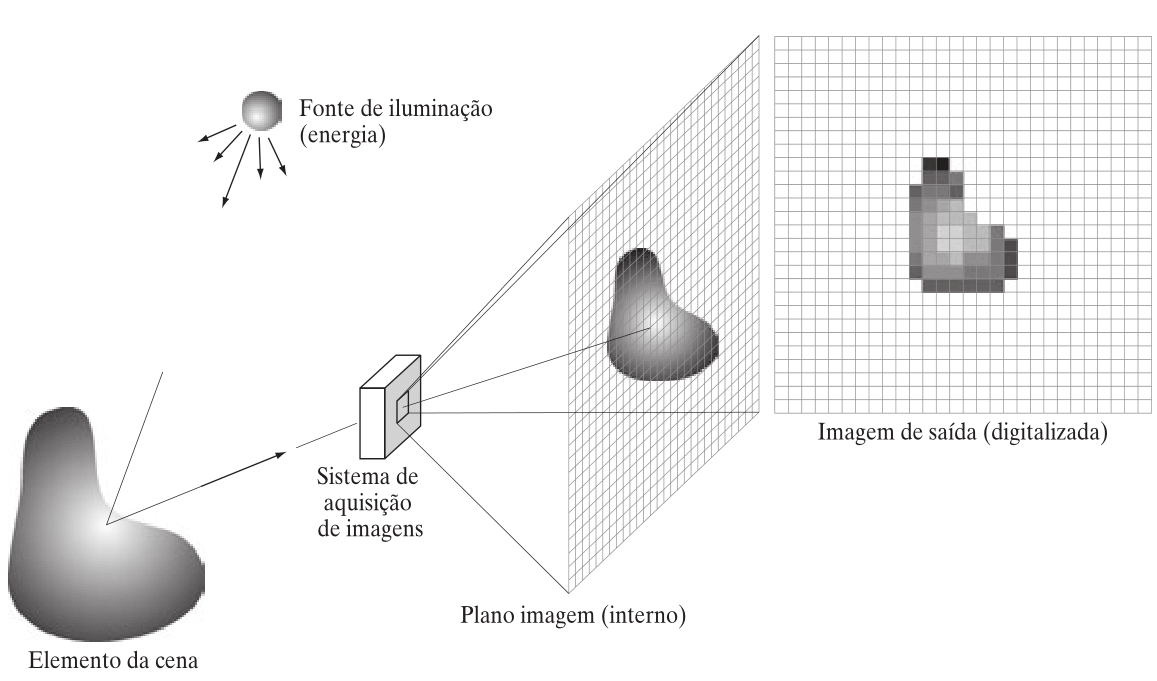
Figura 2 – Distribuição de Sensores



Fonte: (GONZALEZ; WOODS, 2010).

Um sensor típico para câmeras digitais é o Dispositivo de Carga Acoplada (CCD). Os sensores CCD são amplamente utilizados em câmeras digitais e outros instrumentos que utilizam sensores de luz. A principal forma na qual os sensores matriciais são utilizados é mostrada na Figura 3. Os sensores capturam a intensidade da luz emitida e/ou refletida pelo objeto e circuitos digitais e analógicos a transformam em um sinal analógico que é então digitalizado por um outro componente do sistema de aquisição de imagens. A saída é uma imagem digital, como mostra o esquema da Figura 3. (GONZALEZ; WOODS, 2010)

Figura 3 – Exemplo do processo de aquisição de uma imagem digital



Fonte: (GONZALEZ; WOODS, 2010).

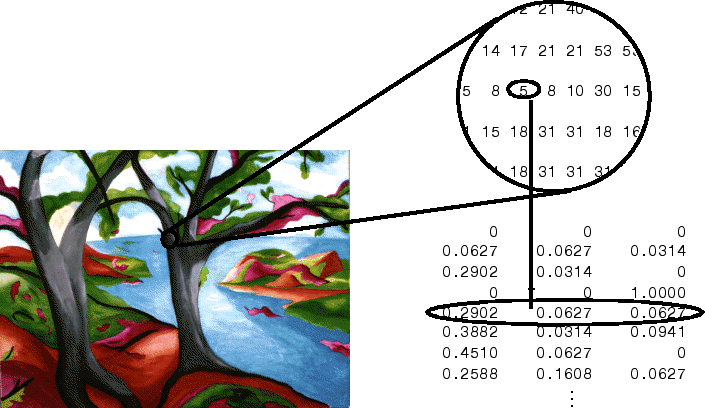
## Representação de imagens digitais

As imagens digitais são representadas por uma matriz bidimensional e cada componente da matriz representa um pixel da imagem. A matriz é uma estrutura fácil de manipular e de visualizar dados. Porém a mesma não é adaptativa a mudanças espaciais, dificultando o manuseio no tamanho ou forma da matriz, normalmente deformando a imagem. (PEDRINI; SCHWARTZ, 2008)

### Imagens indexadas

Para representar uma imagem indexada faz-se necessário duas matrizes. A primeira matriz guarda um número inteiro que se refere a uma linha da segunda matriz. A segunda matriz contém somente 3 colunas e cada linha armazenam o valor da intensidade, em decimais de 0 a 1, das cores vermelho, verde e azul (RGB), que são as cores primárias e fazem parte do sistema de cores aditivas. Misturando as cores RGB é possível representar as outras cores. Por exemplo, quando o valor do RGB está no mínimo (0, 0, 0), resulta na cor preta, quando este está no máximo (1, 1, 1), resulta na cor branca. Na Figura 4 é ilustrado o exemplo de um pixel que apresenta o valor 5 na primeira matriz. Assim, o valor RGB correspondente estará na quinta fileira da segunda matriz. (MATHWORKS, 2019)

Figura 4 – Imagens indexadas

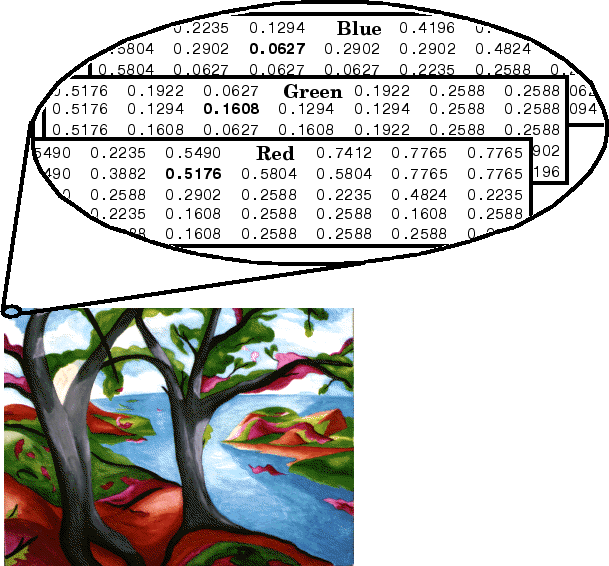


Fonte: (MATHWORKS, 2019)

### Imagem *truecolor*

A imagem *truecolor* necessita de três matrizes do mesmo tamanho, cada matriz para uma cor primária. Então para achar a cor verdadeira de um pixel na posição (x,y) deve-se procurar pela posição (x,y) nas 3 matrizes e misturar os valores RGB correspondentes. A Figura 5 destaca um pixel na posição (2,3). Sua cor verdadeira está na mistura dos valores de RGB coletados na posição (2,3) de cada matriz. (MATHWORKS, 2019)

Figura 5 – Imagem *truecolor*

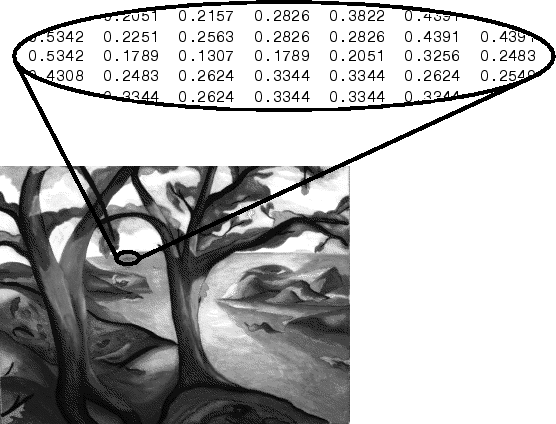


Fonte: (MATHWORKS, 2019)

### Imagem em escala de cinza

Diferente das imagens coloridas, a imagem em escala de cinza necessita de somente uma matriz, pois cada elemento da matriz representa um pixel da imagem. Cada pixel é representado com um valor decimal de 0 a 1, onde 0 significa preto e 1 significa branco. A Figura 6 ilustra os valores da matriz de uma imagem em escala de cinza. A escala de cinza é o estilo de representação de imagem digital que foi utilizado neste projeto. (MATHWORKS, 2019)

Figura 6 – Imagem em escala de cinza



Fonte: (MATHWORKS, 2019)

## Biometria

A biometria é a ciência que estuda as características físicas e comportamentais dos seres vivos e pode ser usada para autenticar a identidade de uma pessoa. O sistema de reconhecimento biométrico traz uma maior exatidão na autenticação de usuários, porém pode se tornar um sistema de autenticação caro, pois necessita de dispositivos extras para a captura das características. (GALVÃO, 2015)

A característica comportamental se baseia em uma ação do ser humano. Como a assinatura, onde a velocidade, o ângulo, e a pressão utilizada para fazer a assinatura são medidas para a autenticação. E a maneira de andar, que também é única. As características comportamentais não são tão usadas como forma de autenticação, pois podem mudar em pouco tempo, então se o sistema de reconhecimento não for constantemente atualizado pode-se ficar prejudicado. (GALVÃO, 2015) (ZHANG, 2009)

A característica física é mais utilizada como forma de autenticação, pois não muda com o tempo. Como a impressão digital, que é única e é formada quando o indivíduo ainda é um feto. A impressão da palma da mão, que tem o mesmo princípio da impressão digital, mas para a palma inteira. A geometria da mão, que são os padrões geométricos de ambas as mãos ou dos dedos. A fala, que pode ser diferenciada pelos padrões acústicos da pessoa. A íris, que é a parte colorida do olho e tem padrões de formação únicos. A retina, que contém um desenho formado por vasos sanguíneos. E o rosto, que é facilmente diferenciado pelos humanos e é a biometria que foi utilizado neste projeto. (GALVÃO, 2015) (ZHANG, 2009)

## Detecção Facial

A detecção facial é uma etapa anterior ao reconhecimento facial visando delimitar a área de captura de características da imagem, otimizando o algoritmo de reconhecimento. (LOPES, 2017)

A assertividade da detecção pode ser prejudicada devido a fatores externos como: posição da câmera que registrou a imagem, expressão facial, presença de elementos ou objetos como barba, bigode, óculos, chapéu e baixa ou alta luminosidade. (LOPES, 2017)

As principais abordagens de detecção de face são: Métodos Baseados no Conhecimento, Métodos Baseados em Características Invariantes, Métodos Baseados em *Templates* e Métodos Baseados na Aparência. (LOPES, 2017)

O Método Baseados na Aparência possui informação prévia sobre o alvo da detecção. As informações necessárias para a ação são capturadas da imagem sem que haja alguma influência externa. Existem três sessões técnicas: *Eigenfaces*, *Hidden Markov Models e* Redes Neurais. (LOPES, 2017)

A *Eigenfaces* utiliza como base a Análise do Componente Principal, na qual as imagens não estão distribuídas aleatoriamente em um espaço amplo e podem ser projetadas com utilização de pixels que fará uma imagem parecida com a face humana sem precisão da imagem. (LOPES, 2017)

A *Hidden Markov Models* está relacionada com o reconhecimento da fala, porém pode ser utilizada para reconhecimento e detecção de faces frontais com análise do cabelo, testa, olhos, nariz e boca, sempre nessa ordem. (LOPES, 2017)

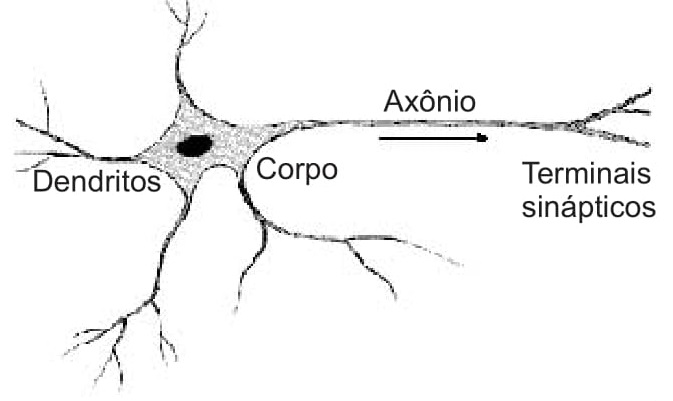
A Rede Neural é uma técnica de reconhecimento de padrões. Após o seu treinamento, com exemplos imagens de faces, a rede neural pode detectar outras faces. E este será o método utilizado neste trabalho. (LOPES, 2017)

## Redes Neurais

A aprendizagem do ser humano passar por um processo chamado de experiência. Porém as máquinas foram projetadas para realizar contas matemáticas e processos previamente estipulados. A principal forma de utilização das redes neurais é para a simulação da aprendizagem do ser humano. Ela simula vários neurônios e a conexão entre eles. Com o grande poder computacional e o processo contínuo de aprendizagem, uma máquina pode realizar tarefas mais complexas, com mais velocidade e assertividade do que um ser humano. (HAYKIN, 2001)

O neurônio biológico contém duas partes importantes para a sua função e comunicação, os axônios e os dendritos, como ilustrado na Figura 7. Os axônios são linhas de transmissão onde se passa a informação, já os dendritos são as zonas receptivas, as que recebem a informação. (HAYKIN,2001)

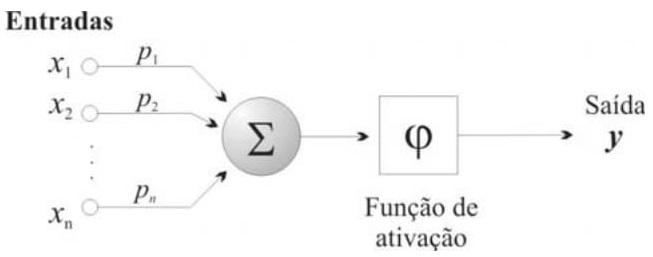
Figura 7 – Representação simplificada de um neurônio



Fonte: (FERNEDA, 2006)

O neurônio artificial, por sua vez, é um conjunto de operações matemáticas que realizam as sinapses, somam os sinais recebidos e realizam a função de ativação que restringe a saída para que a resposta seja ativado ou não-ativado. (HAYKIN, 2001)

Figura 8 – Modelo matemático de um neurônio



Fonte: (FERNEDA, 2006)

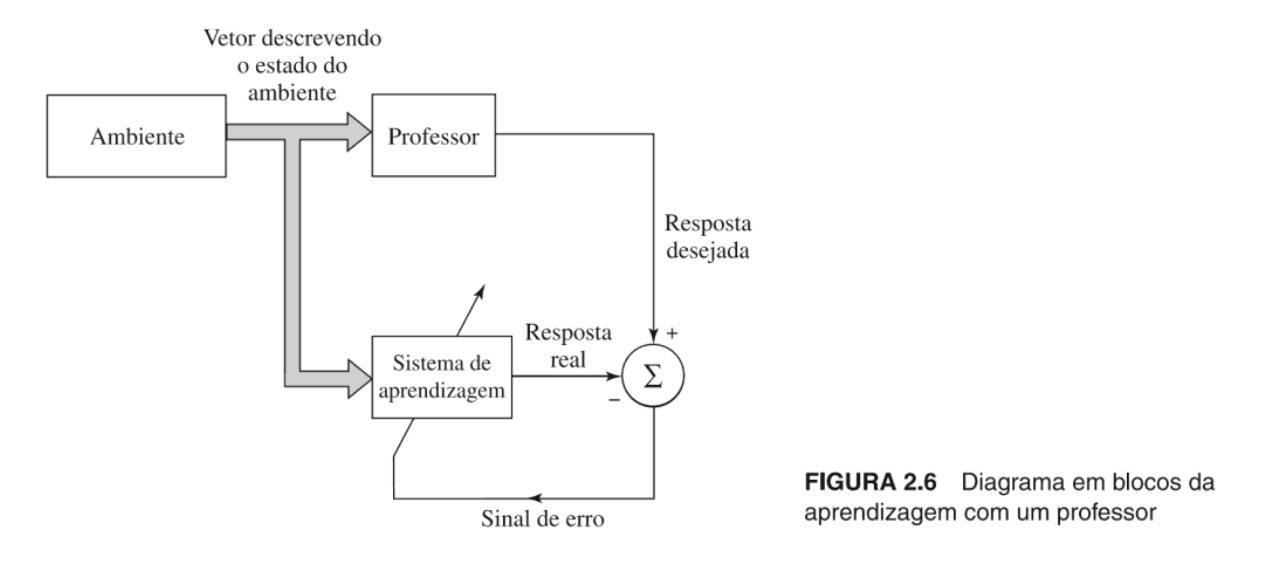
A estrutura do neurônio artificial é basicamente formada por: entradas (x), pesos sinápticos (p), um somador para acumular os sinais de entrada e uma função de ativação que limita o sinal de saída (y) a um valor fixo. (FERNEDA, 2006)

### Processos de aprendizagem

O algoritmo de aprendizagem, modifica os pesos sinápticos da rede neural para alcançar o objetivo desejado. É possível treiná-la por uma aprendizagem supervisionada ou não supervisionada. (HAYKIN, 2001)

A aprendizagem supervisionada é um sistema realimentado fechado, onde é desconhecido o ambiente em que ele se encontra. Pode-se treinar a rede neural com a possibilidade de erro, sendo chamado de superfície de erro como ilustrado na Figura 9. A rede neural é treinada até que tenha uma baixa porcentagem de erro. (HAYKIN, 2001)

Figura 9 – Diagrama em blocos da aprendizagem



Fonte: (HAYKIN, 2001)

Já a aprendizagem não supervisionada, como o nome já diz, não há supervisão, porém, são dadas condições para que seja realizada uma medida independente da tarefa, assim, ao regularizar as estatísticas de erro, a rede neural poderá realizar a tarefa de forma correta. (HAYKIN, 2001)

### Perceptron

A forma mais simples de uma rede neural é formada de apenas uma unidade, chamada de perceptron. Ele é formado de somente um neurônio e tem a função de classificar padrões linearmente separáveis. (PEDRINI; SCHWARTZ, 2008)

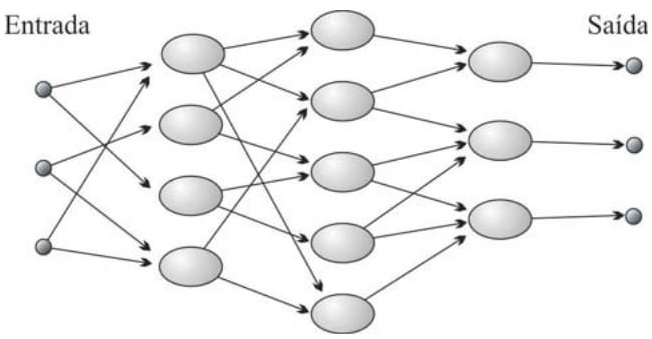
A regra do perceptron apresenta convergência no caso de as amostras estarem linearmente separáveis. O treinamento tem como base a atualização dos pesos para que a rede neural consiga acertar a maioria das entradas de treinamento. (PEDRINI; SCHWARTZ, 2008)

Para separar classes não linearmente separáveis faz-se o uso de redes neurais de múltiplas camadas com saídas não binárias.

### Redes Neurais de Múltiplas Camadas

A rede neural de múltiplas camadas consiste em três partes principais: a camada de entrada, uma ou mais camadas ocultas e uma camada de saída. O sinal mandado pela camada de entrada, passa por cada camada a sua frente alimentado a rede neural, do inglês *Feedforward*, como ilustrado na Figura 10. (HAYKIN, 2001)

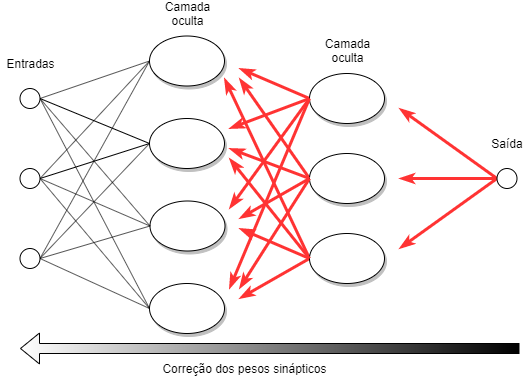
Figura 10 – Representação simplificada de uma rede neural artificial



Fonte: (FERNEDA, 2006)

Para a atualização dos pesos sinápticos, em redes neurais de múltiplas camadas, é utilizado a retropropagação, do inglês *Backpropagation*. Que consiste em voltar da camada de saída em direção a camada de entrada, atualizando os pesos sinápticos, como ilustrado na Figura 11. A correção do erro é feita através da subtração da resposta real pela resposta desejada. (HAYKIN, 2001)

Figura 11 – Retropropagação

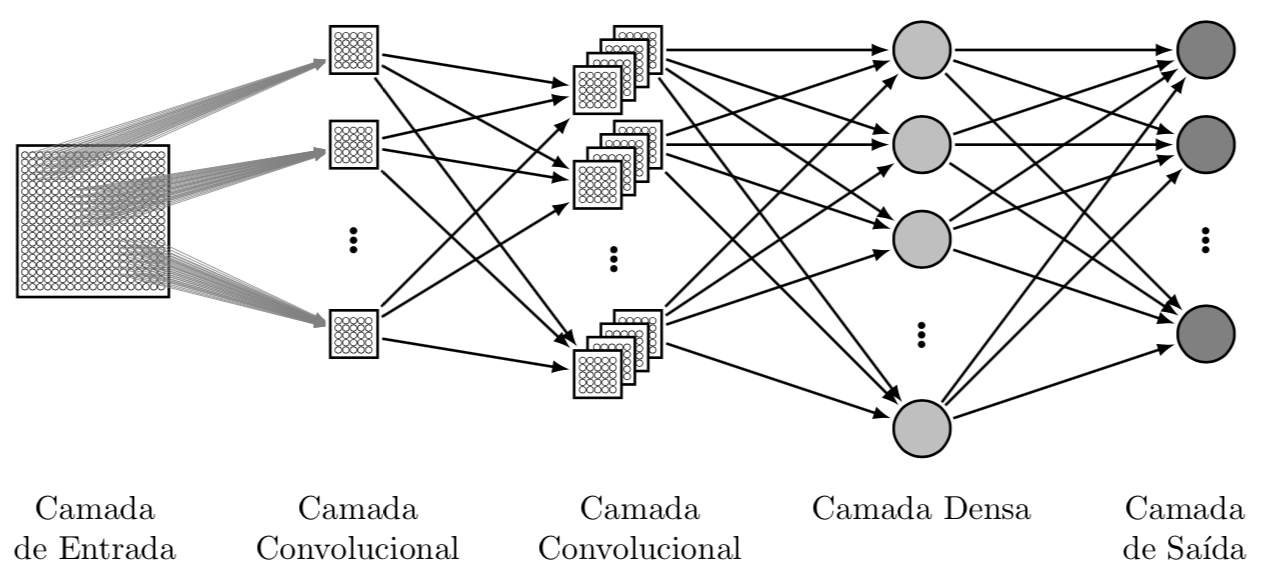


Fonte: (Elaborado pelos Autores, 2019)

### Redes Neurais Convolucionais

A CNN realiza um cálculo convolucional, criando matrizes com as partes mais significativas da imagem, como ilustrado na Figura 12. O neurônio receberá apenas os pixels mais significativos, de forma que os padrões na imagem ficam mais visíveis. (REIS; RAVINDRA, 2017)

Figura 12 – Rede neural Convolucional



Fonte: (SAKURAI, 2019)

## Reconhecimento Facial

Antes retratado em filmes como ficção científica, com o avanço da tecnologia, o reconhecimento facial deixou de ser ficção e hoje tem várias aplicações, sendo as principais delas a autenticação e identificação. (WEST, 2017)

O reconhecimento facial é bastante utilizado como forma de identificação, pois é uma tecnologia não invasiva. É aceita como forma de autenticação pela maioria dos usuários e não necessita de um aparelho com captura de imagem de alta resolução, sendo assim uma tecnologia de baixo custo. Há dois principais tipos de reconhecimento facial. O reconhecimento por padrões da temperatura emitida pelo usuário utilizando uma câmera infravermelha e o reconhecimento por padrões visíveis do rosto que pode ser realizada por uma câmera digital e é a forma utilizada neste projeto. (ZHANG, 2009)

Em 2010 o Facebook começou a implementar a funcionalidade de reconhecimento facial para ajudar a identificar usuários em fotos. E mais de 350 milhões de fotos são carregadas e marcadas por esta tecnologia todos os dias. (WEST, 2017)

A implementação do reconhecimento facial na área policial tem aumentado e gerado resultados. Em 2011 os Estados Unidos da América usaram o reconhecimento facial para confirmar a identidade do corpo de Osama bin Laden, depois que ele foi morto em um ataque. (WEST, 2017)

## Transformada Discreta do Cosseno

A Transformada Discreta do Cosseno (DCT) tem sido amplamente explorada pela comunidade de processamento de imagens, principalmente nas áreas de compressão, filtragem e extração de características. (PEDRINI; SCHWARTZ, 2008)

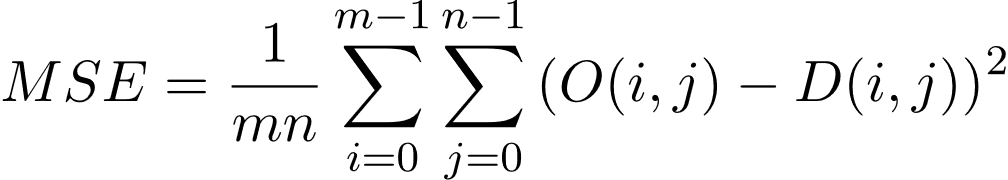
A DCT faz a alocação da energia da imagem em regiões próximas a origem. Após a execução da DCT, muitos coeficientes apresentaram valores próximos a zero, podendo ser eliminados sem que haja perda significativa de informações contidas na imagem. Este processo de eliminar dados menos significativos é utilizado na compressão de arquivo. (PEDRINI; SCHWARTZ, 2008)

O Joint Photographic Expert Group (JPEG) utiliza a DCT como forma de compressão de imagem. O JPEG calcula o DCT para cada blocos de 8x8 na imagem, baseado na fórmula (1), onde, com um vetor de entrada de elementos, em que e representa o núcleo da transformada. (MURALI, 2013)

## Relação Sinal-Ruído de Pico

A Relação Sinal-Ruído de Pico (PSNR) é relação entre o máximo possível de um sinal pela potência de ruído. Quando o valor do PSNR está alto, significa que há uma forte relação entre as potências da imagem original e a imagem de comparação. O PSNR foi utilizado para realizar a comparação entre duas DCTs, a da imagem de entrada e da imagem previamente cadastrada. Este processo foi relizado para todas as imagens cadastradas, após isto, o maior PSNR foi analizado para saber se as caracteristicas das faces eram parecidas. O termo matemático do PSNR é demostrado na fórmula (2). (DANTAS, 2015)

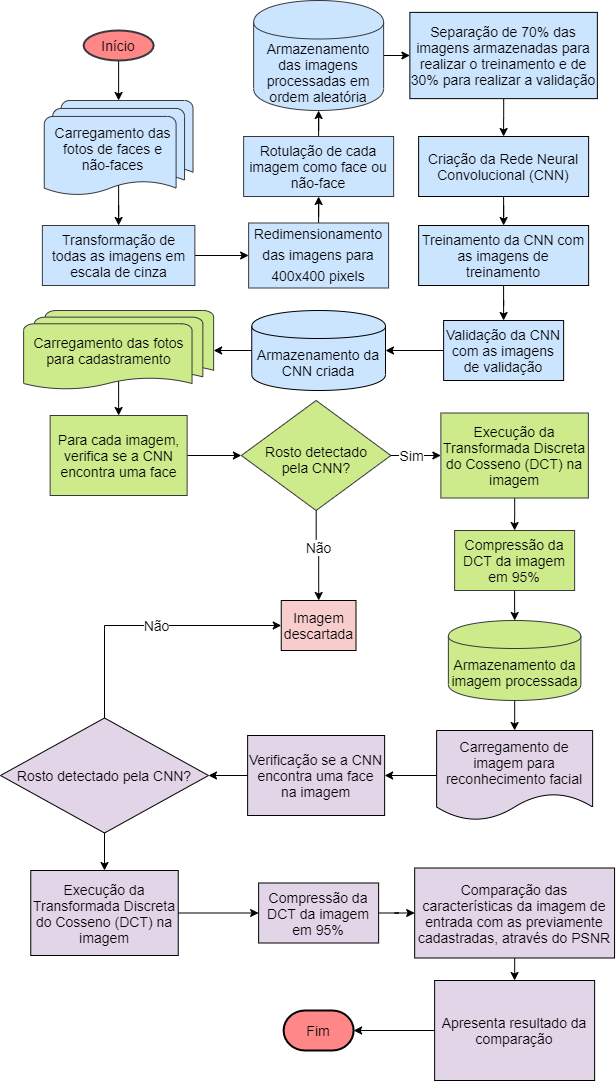
é a diferença da imagem a ser comparada e da imagem de entrada. O MSE é demostrado na fórmula (3).



Sendo que representa a imagem a ser comparada, representa a imagem de entrada, representa o número total de linhas de pixels da imagem, o número de colunas e é o valor da máximo da posição da imagem.

# DESENVOLVIMENTO E RESULTADOS OBTIDOS

Figura 13 – Fluxograma do programa

****

Fonte: (Elaborado pelos Autores, 2020)

O programa criado foi desenvolvido em Python. Dividido em 3 etapas. Como demostrado na Figura 13. Na primeira etapa, a criação da CNN, na segunda o cadastramento das faces e na terceira a comparação de imagens de entrada com as imagens cadastradas.

Para realizar o treinamento da CNN foram coletadas 2000 imagens com faces do banco de imagens Labeled Faces in the Wild (LFW). 2000 imagens sem faces foram coletadas dos sites libreshot.com e shopify.com, que tem imagens livres de direitos autorais. Alguns exemplos das imagens utilizadas estão ilustrados na Figura 14.

Figura 14 – Exemplo do Imagens Utilizadas

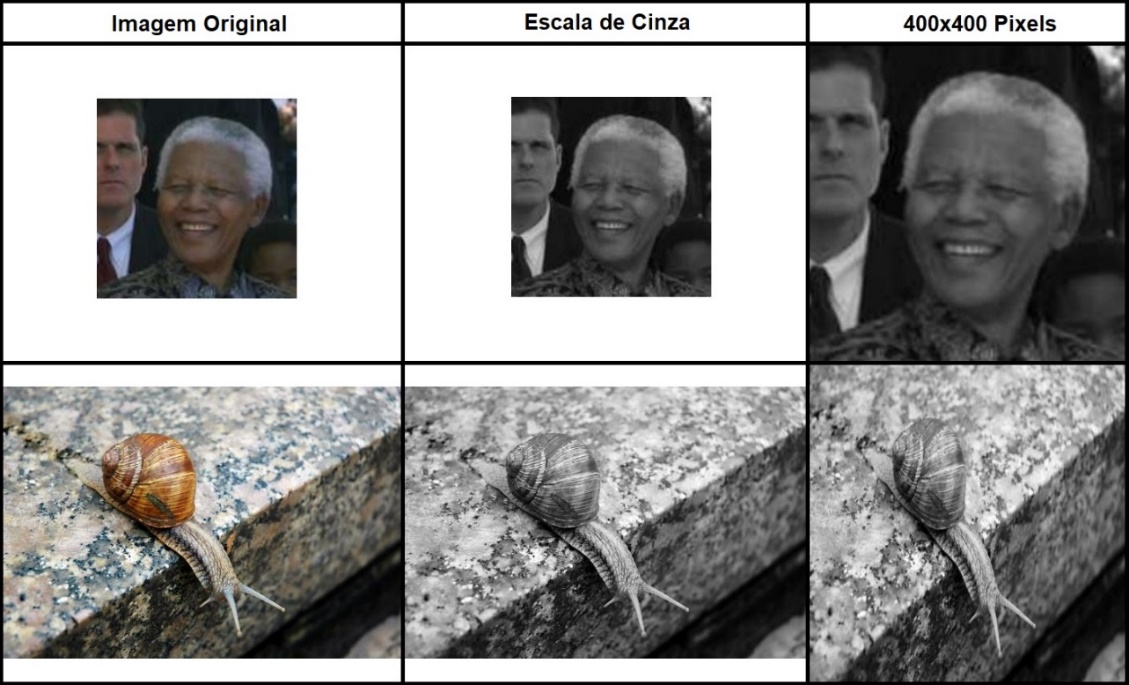


Fonte: (Elaborado pelos Autores, 2019)

Antes de realizar o treinamento da CNN foi realizado um pré-processamento nas imagens, para que todas estivessem em escala de cinza e possuíssem o mesmo tamanho. Na Figura 15, a imagem que contém uma face, originalmente tem o tamanho de 250x250 pixels e a imagem que contém um caramujo de 511x340 pixels. Ao lado estão os resultados da transformação das imagens em escala de cinza e depois o redimensionamento para 400x400 pixels.

Após o pré-processamento das imagens, as imagens são rotuladas em “Face” e “Não Face” e são guardadas aleatoriamente em uma lista para servirem de entrada da CNN.

Figura 15 – Exemplo de pré-processamento de imagem



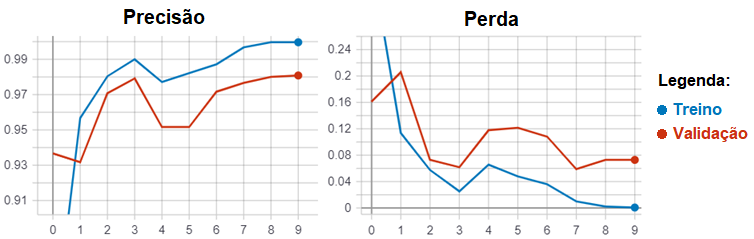
Fonte: (Elaborado pelos Autores, 2019)

Para avaliar a precisão da CNN, 30% da lista com as imagens foi separada para a validação e 70% da lista foi usada para treinamento. A CNN contém 3 camadas convolucionais, 2 camadas ocultas com 64 neurônios em cada uma. A camada de saída contém 1 só neurônio, que utiliza a função de ativação Sigmoide para classificar a amostra não linearmente separável.

O treino de cada passo da CNN demorou em média 540 segundos. Totalizando 5387 segundos para realizar os 10 passos da criação da CNN. A cada passo a precisão da CNN foi aumentando, como demonstrado na Figura 16.

Após o treino e validação, o modelo ficou com precisão de 98% e perda de 7% em imagens que a CNN não tinha visto ainda. O Figura 16 mostra os gráficos de precisão e perda a cada passo da CNN.

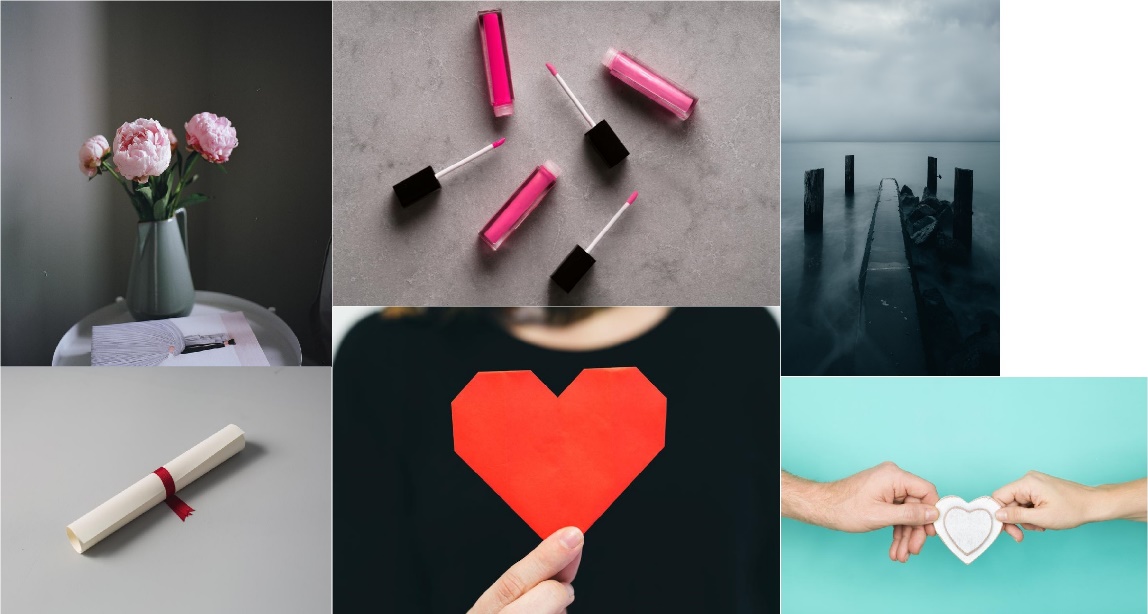
Figura 16 – Precisão e Perda do Conhecimento



Fonte: (Elaborado pelos Autores, 2019)

A CNN teve uma boa precisão de acerto, porém mesmo assim ainda houve erros de imagens que tinham faces e que foram rotuladas como “Não Face”, chamado de falso negativo, e imagens que não continham faces e foram rotuladas como “Face”, chamado de falso positivos, como ilustrado na Figura 17.

Figura 17 – Exemplo de falsos positivos

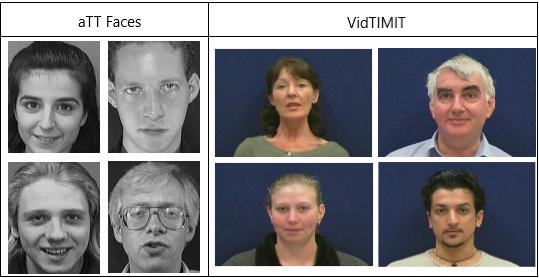


Fonte: (Elaborado pelos Autores, 2019)

Para realizar os testes do reconhecimento facial, 2 bancos de imagens foram escolhidos, sendo eles o aT&T Faces que contém imagens de 40 pessoas e o VidTIMIT que contém imagens de 43 pessoas. Foi utilizado o mesmo pré-processamento da etapa de detecção facial e todas as imagens passaram pela validação da CNN criada para, só então, passar pelo processo de cadastramento. 80 pessoas foram cadastradas com 8 imagens cada, totalizando 640 imagens. Alguns exemplos das imagens utilizadas estão ilustrados na Figura 18.

Foram realizados 2 testes diferentes na etapa do reconhecimento facial. O primeiro consiste em verificar a importância da DCT no reconhecimento facial comparando o resultado a utilização da DCT no programa e o segundo teste, avalia o a precisão do programa, usando como imagens de entrada pessoas que não estavam cadastradas.

Figura 18 – Exemplo de imagens Utilizadas para teste

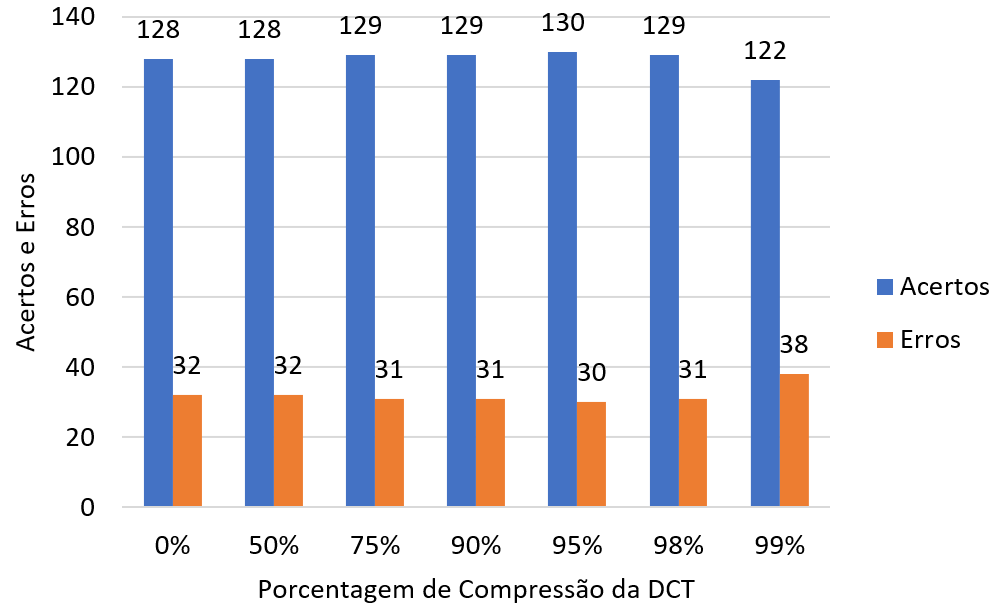
****

Fonte: (Elaborado pelos Autores, 2020)

Para o primeiro teste, 2 imagens de cada pessoa cadastrada foram coletadas, totalizando 160 imagens de entrada. O resultado obtido foi de 130 acertos reconhecimento facial e 30 erros no melhor caso, com 95% de compressão da DCT.

A DCT foi utilizada com o objetivo de obter as características das faces recebidas. Porém, foi observado que o ganho de acertos não foi tão expressivo no uso desta transformada, como demostrado na Figura 19.

Figura 19 – Compressão da imagem



Fonte: (Elaborado pelos Autores, 2020)

Para testar a precisão do programa, 2000 imagens de pessoas que não estavam cadastradas foram usadas como teste. Um valor mínimo foi estabelecido através do maior PSNR das imagens não cadastradas. O resultado foi validado levando-se em consideração . Assim, quanto maior o valor de , mais seletivo o programa se torna, evitando falsos-positivos, porém imagens que poderiam ser consideradas positivas, mas não obtiveram também foram descartadas.

Após a aplicação do valor mínimo o resultado passou de 130 para 107 acertos, e de 30 para 23 erros. Ou seja, 30 imagens foram desconsideradas por não obter . Entretanto, todas as imagens não cadastradas utilizadas no teste, foram desconsideradas, alcançando o objetivo de diminuir os falsos positivos.

# CONCLUSÃO

Este trabalho apresentou o estudo e aplicação do reconhecimento facial. O programa foi dividido em 3 partes principais, detecção facial, através de uma Rede Neural Convolucional, extração de características, através da Transformada Discreta do Cosseno e reconhecimento facial, através da fórmula da Relação Sinal-Ruído de Pico.

A criação e treino da CNN para realizar a detecção de faces trouxe resultados satisfatórios. Porém nas imagens com faces descentralizadas ou distantes da câmera, observou-se dificuldade em detectar a face, resultando em falsos negativos.

A utilização da DCT para extração de características não foi essencial para o reconhecimento facial como esperado pelos autores, pois não teve ganho significativo nos acertos do reconhecimento.

A fórmula da PSNR resultou em uma boa assertividade e precisão do reconhecimento facial.

Dessa forma, pode-se dizer que a implementação das técnicas utilizadas e os estudos feitos foram satisfatórios para que houvesse o reconhecimento facial, mesmo ocorrendo casos de dificuldades na extração de características pela DCT, assim, alcançando os objetivos sugeridos nesse trabalho.

## Sugestão para trabalhos futuros

Outras transformadas podem ser testadas como forma de captura de características da imagem ou para a detecção facial. A etapa de reconhecimento facial também pode ser modificada para ser realizada por uma rede neural. Pode-se optar por outras formas de captura de características da face, como calcular a distância entre os olhos, nariz, boca, testa e queixo ou capturar o formato do rosto com uma câmera infravermelho.

**REFERÊNCIAS BIBLIOGRÁFICAS**

CAMBRIDGE. AT&T Laboratories Cambridge **database of faces.** 1994.

DANTAS, A. M. R. **Codificação de Imagens em Baixo Consumo Para Comunicações em Minas Subterrâneas.** Brasília, DF: UNB,2015

FERNEDA, E. **Redes neurais e sua aplicação em sistemas de recuperação de informação.** Disponível em: <http://www.scielo.br/pdf/ci/v35n1/v35n1a03.pdf> Acessado em 20 de out. 2019.

GALVÃO, M. C. **Fundamentos em segurança da informação.** São Paulo: Pearson Education do Brasil, 2015. (114 p.) ISBN 9788543009452.

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagens.** 3. ed. São Paulo, SP: Pearson Prentice Hall, 2010. (644 p.) ISBN 9788576054016.

HAYKIN, S.**Redes neurais:** princípios e prática. 2. ed. Porto Alegre: Bookman, 2001. (900 p.). ISBN 9788573077186.

LOPES, E. C. **Detecção de Faces e Características Faciais.** Porto Alegre, RS: PUCRS, 2005.

MATHWORKS. **MATLAB:** Graphics(R2019b). Disponível em: <https://www.mathworks.com/help/pdf\_doc/matlab/graphg.pdf> Acessado em 14 de set. 2019.

MURALI, S. et al. **Comparision and analysis of photo image forgery detection techniques**. ArXiv, 2013.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de imagens digitais: princípios, algoritmos e aplicações.** São Paulo, SP: Thomson, 2008. (508 p.). ISBN 978-85-221-0595-3.

REIS, P.; RAVINDRA, S. **Como as redes neurais convolucionais realizam o reconhecimento de imagem**. InfoQ Brasil, São Paulo, SP, Brasil, 2017.

SANDERSON, C.; LOVELL, B.C. **Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference.** Lecture Notes in Computer Science (LNCS), Vol. 5558, pp. 199-208, 2009.

SAKURAI, R. **Implementando a estrutura de uma Rede Neural Convolucional utilizando o MapReduce do Spark.** Disponível em: **<**http://rafaelsakurai.github.io/cnn-mapreduce/> Acessado em 27 de out. 2019.

WEST, J. D. **A brief history of face recognition**. 2017. Disponível em: <https://www.facefirst.com/blog/brief-history-of-face-recognition-software/> Acesso em: 14 nov. 2019.

ZHANG, D. et al. **Advanced pattern recognition technologies with applications to biometrics.** IGI Global, 2009.