Detecção de Imagem para Reconhecimento de Sinais de Libras

Gabriel Ângelo Perez Gasparini Sabaudo[1], **Guilherme Pina Cardim**[1]

[1]Departamento de Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

gabriel.angelo@uel.br, gpcardim@uel.br

Abstract. *Nowadays, technology favors us with great possibilities for creations and has the power to help in our daily lives in different ways, being especially important for the disabled who need their use. The hearing impaired community faces several barriers in society due to the difficulty of communication with other people, fact that makes integration into society difficult; therefore, the use of technologies that stimulate reintegration and promote accessibility are today a reality for many of them.*
The official sign language adopted in the country is Libras (Brazilian Sign Language), but only a small portion of non-disabled Brazilians can use it. Due to its importance, the objective of the following work, through the detection of gestural signals using images and artificial intelligence, is to be able to assist in the education of people who wish to learn sign language and promote greater use of it.

Resumo. *Nos dias atuais a tecnologia nos favorece com grandes possibilidades de criações e possui o poder de auxiliar em nosso cotidiano de diversas maneiras, sendo elas especialmente importantes para deficientes que precisam de seu uso. A comunidade de deficiêntes auditivos encara diversas barreiras na sociedade devido a dificuldade de comunicação com outras pessoas, um fato que dificulta a integração na sociedade; portanto, o uso de tecnologias que estimulam a reintegração e promovem a acessibilidade são hoje uma realidade para muitas delas.*
A linguagem gestual oficial adotada em território nacional é a Libras (Língua Brasileira de Sinais), porém apenas uma pequena parcela de brasileiros não deficientes conseguem usá-la. Devido a sua importância, o objetivo do seguinte trabalho, através de detecção de sinais gestuais utilizando imagens e inteligência artificial, é poder auxiliar na educação de pessoas que desejam aprender a linguagem de sinais e promover maior uso da mesma.

1. Introdução

No tocante à tecnologias voltadas para acessibilidade, existe uma vasta gama de aplicações. O mercado procura atender necessidades para vários tipos de problemas, mas nem sempre todos possuem a atenção necessária. O que vemos hoje em dia é um mundo interligado por soluções que visam praticidade, essencialmente pelo fato da maioria das pessoas possuírem um celular, por exemplo, o que se torna um facilitador para ter acesso a tais aplicações em meio ao cotidiano.

Estes fatores se firmam em relação à deficientes auditivos e o ensino da Libras. A dificuldade relacionada a comunicação com outras pessoas hoje pode ser superada, por

exemplo, pelo uso de aplicativos que ajudam no ensino de sinais gestuais. Quando falamos da linguagem de sinais, nos referimos a uma comunidade muito grande de usuários, e sua importância para essas pessoas é altíssima, principalmente por fazer parte da educação e por formar sua participação na sociedade.

Infelizmente, por muitas vezes os surdos não possuem condições mínimas de atendimento. Em repartições públicas, hospitais, lojas e locais adaptados que lidam com questões de acessibilidade, raramente há alguém preparado para atendê-los [9]. Justamente por esse motivo, a proposta deste trabalho é resolver este tipo de problema com uma aplicação que induzirá o ensino da Libras e o reconhecimento de sinais gestuais com maior facilidade.

O uso de métodos ligados à detecção de alvos em imagens digitais e inteligência artificial se alicerçam muito bem para problemas dessa natureza. O reconhecimento de gestos é uma tecnologia já muito consolidada e utilizada nos últimos anos, fator este que ajuda no desenvolvimento diário de *softwares* ligados a essa área. A partir disto, este trabalho descreve um método baseado na utilização de uma rede neural aplicada em imagens pré selecionadas, onde será feita a tradução para o português de cada sinal de Libras detectado. A vantagem do uso de uma rede neural é seu poder de aprendizado, o que pode garantir resultados finais com maior precisão e eficiência.

Este documento esta dividido entre as seguintes seções:

- A seção 2 apresenta os conceitos que formam a base para o desenvolvimento do trabalho, além de projetos e artigos relacionados ao assunto.
- A seção 3 descreve os objetivos do trabalho.
- A seção 4 discute o passo a passo do métodos a fim de alcançar os objetivos propostos.
- A seção 5 apresenta o cronograma proposto para o desenvolvimento do trabalho.
- A seção 6 mostra a contribuição do trabalho caso os objetivos sejam alcançados.

2. Fundamentação Teórico Metodológica e Estado da Arte

Foi abordado na seção 1 sobre o frequente uso de métodos que englobam a relação homem-máquina para resolver problemas ou dificuldades que se baseiam pelo visual ou por ações. O reconhecimento de gestos é uma das abordagens empregadas com foco em acomodar aplicações avançadas em HCI (Interação Homem-Máquina). Essa tecnologia tornou-se um dos mais importantes campos de pesquisa nessa área por fornecer a base para reconhecer o corpo, as mãos, o movimento das mãos ou postura, e expressões faciais [3].

A visão computacional compreende um grande número de artigos relacionados ao reconhecimento de sinais de Libras, e por este motivo este trabalho revisou uma parcela de alguns destes trabalhos, os quais serão mencionados nesta seção.

Nas subseções abaixo será feita uma breve introdução aos conceitos que englobam os métodos de reconhecimento de imagens e redes neurais.

2.1. Processamento de Imagens

Uma imagem pode ser definida por uma função bidimensional $f(x,y)$, onde x e y são coordenadas espaciais, e a amplitude de f em qualquer par (x,y) é chamada de nível de

cinza naquele ponto da imagem. Quando x, y, e os valores de intensidade de f são finitos e discretos, dizemos que a imagem é uma imagem digital. O campo de processamento de imagens digitais se refere ao processo de imagens digitais por meio de um computador digital [6].

Ou seja, o processamento digital de imagens nada mais é do que a manipulação de uma imagem (ou dado) por computador, de certa maneira que a entrada e a saída do processo são imagens, assim como a Figura 1. O grande objetivo do mesmo consiste em melhorar o aspecto visual de certas feições estruturais, de certa maneira que o analista consiga melhor interpretar, classificar e tomar decisões com base nos dados existentes na imagem [4].

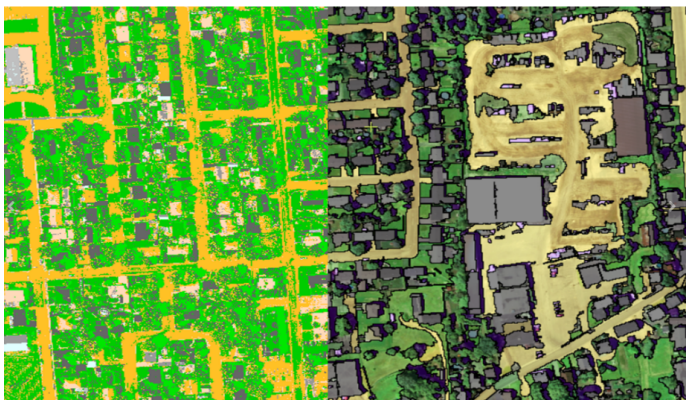

Figura 1. Exemplo de um processamento de imagem, retirada de [5]

Tendo este objetivo como ponto final, o processamento digital de imagens gera inclusive produtos que possam ser posteriormente submetidos a outros processamentos. Um ponto importante a ser destacado dentro do processamento de imagens é a manipulação com o espaço de cores, conceito primordial em trabalhos que abordam o reconhecimento de objetos em imagens [4].

2.2. Modelo de Cores

O uso de cores no processamento de imagens é motivado por dois fatores principais. Primeiro, a cor é um descritor poderoso que muitas vezes simplifica a identificação e extração de objetos de uma cena. Em segundo lugar, os humanos podem discernir milhares de tons e intensidades de cores, em comparação com apenas alguns tons de cinza. Este segundo fator é particularmente importante na análise manual de imagens [6].

Os diversos modelo de cor (também chamados de espaços de cor ou sistema de cor) possuem um propósito para seu uso. Cada espaço possui diferentes características vantajosas para determinadas aplicações, as quais podem se tratar no foco do brilho, saturação, matiz, crominância, sombra, etc.

Para melhor entendimento e contextualização deste campo, alguns dos modelos de cores mais notáveis e utilizados na área de reconhecimento em imagens serão brevemente descritos abaixo.

2.2.1. Modelo RGB

É o sistema de cores mais utilizado em monitores, televisões e câmeras de vídeo. O modelo RGB consiste num espaço de cores aditivas formado pelos canais: Vermelho (*red*), Verde (*green*) e Azul (*blue*).

Este modelo pode ser representado pelo sistema Cartesiano tridimensional de coordenadas, onde os valores primários RGB se estabelecem nos 3 cantos opostos do cubo, como é possível perceber na Figura 2. A combinação das três componentes (nos seus valores máximos) origina a luz branca e a combinação de todos os componentes (nos seus valores mínimos) origina uma cor preta. As outras cores representadas nos cantos restantes (Ciano, Magenta e Amarelo) se formam pelas combinações entre as cores principais.

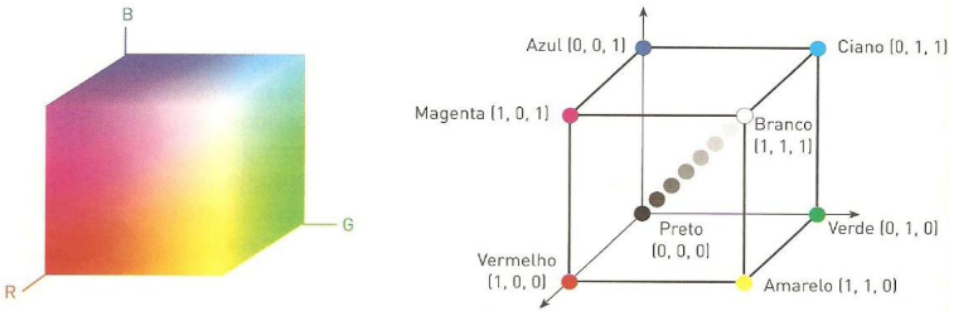

Figura 2. Espaço de cores RGB, retirada de [12]

O modelo RGB é amplamente utilizado em metodologias voltadas para o reconhecimento de pele em imagens, porém não é o sistema mais adequado devido ao fato de que a formação de cores neste espaço não possui correlação com a percepção humana. Ou seja, cores próximas neste espaço (pouca variação das cores primárias) não estão necessariamente próximos em termos de percepção [10].

2.2.2. Modelo YCbCr

O espaço YCbCr é um sinal codificado do RGB, comumente utilizado por estúdios de TV europeus e para trabalhos de compressão de imagens [10]. Neste espaço, a cor é representada pela luma (Y) e por 2 componentes que correspondem à subtração desta luma das componentes azul e vermelha: Cb e Cr. A partir do espaço RGB, pode-se equacionar este espaço de cores segundo as equações abaixo.

$$Y = 0.299R + 0.587G + 0.114B$$
$$Cr = (R - Y) \times 0.713 + 128$$
$$Cb = (B - Y) \times 0.564 + 128$$

A Figura 3 mostra uma representação do espaço YCbCr. Nesta, pode-se observar a variação de luma no eixo Y. Para o valor mínimo de Y, a cor é preta. Já para o valor máximo, esta é branca, independente dos valores de Cb e Cr. Este espaço separa

as componentes de luminância e crominância, representando uma vantagem no uso em algoritmos de reconhecimento de pele.

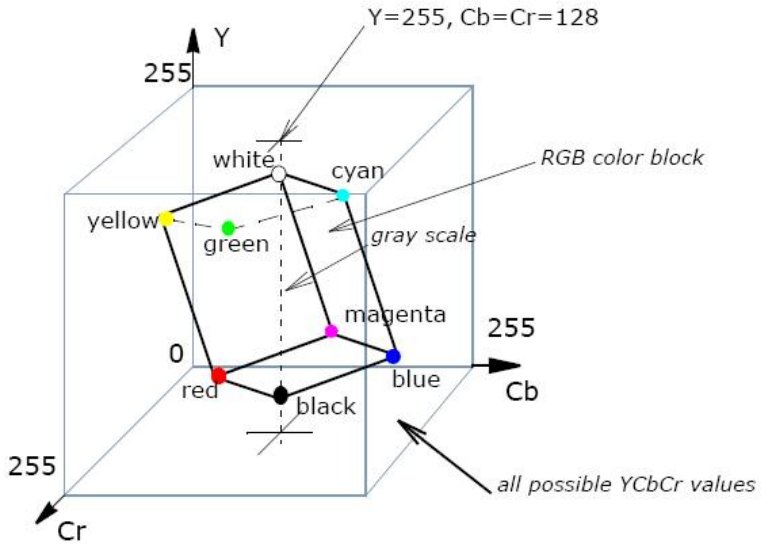

Figura 3. Espaço YCbrCr em relação ao espaço RGB, retirada de [1]

2.2.3. Modelo HSV

O modelo de cor HSV (matiz, saturação, valor) foi desenvolvido para ser mais "intuitivo" na manipulação de cores e foi projetado para aproximar a maneira como os humanos percebem e interpretam as cores.

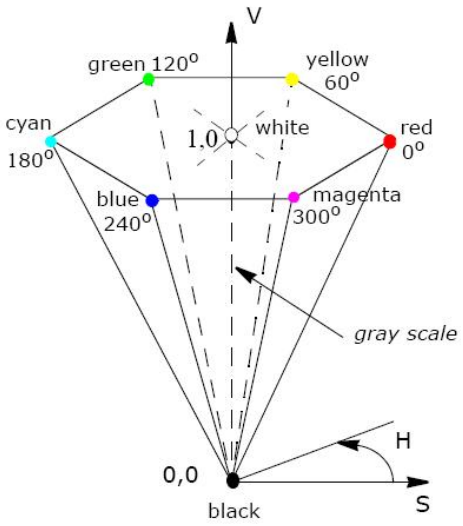

Figura 4. Espaço HSV representado como um hexacone, retirada de [1]

Verificando a Figura 4, é possível perceber que a matiz (H) define a própria cor. Os valores para o eixo de matiz variam de 0 a 360 começando e terminando com vermelho e

passando por verde, azul e todas as cores intermediárias. A saturação (S) indica o grau em que o matiz difere de um cinza neutro. Os valores vão de 0, o que significa sem saturação de cor, a 1, que é a saturação máxima de um determinado matiz em uma determinada iluminação. Já o valor (V), indica o nível de iluminação, variando de 0 (preto, sem luz) a 1 (branco, iluminação total).

O espaço HSV apresenta características desejáveis no tocante à segmentação de pele, tais como o fato das cores serem formadas de maneira intuitiva, além da separação das componentes de crominância e luminância, bastante utilizada nestes algoritmos.

2.3. Segmentação da mão

Como abordado anteriormente, os modelos de cores são amplamente utilizados em trabalhos que precisam, de alguma forma, identificar a pele em um cenário e separá-la. Trabalhos como o de [8] se aproveitam das características dos diferentes modelos de cores e as combinam para obter melhores resultados e cobrir possíveis erros de detecção, garantindo maior acertabilidade no processo de reconhecimento.

A identificação de uma mão, por exemplo, se torna uma tarefa não tão trivial, devido ao fato de que a pele humana possui inúmeras tonalidades. Por isso, os trabalhos voltados para este campo geralmente possuem algoritmos de reconhecimento de pele baseados na definição de intervalos. Estes intervalos, dados por valores referentes aos canais dos espaços de cores, são estipulados com base em diferentes imagens de variados tons de pele e análises a respeito destas.

A partir destas análises, é possível avaliar quais componentes ou espaços de cores permitem a melhor separação dos pixels que são de pele ou não são, além de permitir a criação de regras que sejam genéricas, capazes de reconhecer corretamente os diferentes tons de pele.

2.3.1. Método de Otsu

O método de Otsu (ou Algoritmo de Otsu) é um algoritmo de limiarização por equilíbrio do histograma onde seu objetivo, a partir de uma imagem em tons de cinza, é determinar o valor de um limiar que separe os elementos do fundo e da frente da imagem em 2 cores, o branco e o preto.

O conceito proposto deste método é de iterar por todos os valores possíveis para o threshold (limiar) em uma imagem (ou seja, o intervalo dinâmico da imagem), buscando aquele que minimiza a soma da variância intraclasses da imagem. Esse valor irá corresponder ao melhor threshold para o caso, separando frente e fundo e atribuindo uma cor para cada classe [11].

Primeiramente é cálculado o peso do histograma da imagem, verificando qual lado possui maior peso. O lado mais pesado sofre uma alteração para se tornar mais leve, e este processo se repete até que os dois extremos do histograma se encontrem.

2.4. Aprendizado de Máquina

O conceito de *Machine Learning* caminha praticamente lado a lado com o campo de processamento e detecção de imagens. Quando se trata de dados de imagem, os algoritmos

de ML podem interpretá-los da mesma forma que nossos cérebros. Eles são usados em quase todos os lugares, desde o reconhecimento facial durante a captura de imagens em nossos smartphones, automação de trabalho manual tedioso, carros autônomos e muitos outros exemplos [7].

A partir de imagens dadas como entrada, os algoritmos baseados em ML são capazes de identificar padrões, extrair informações de acordo com objetos e contornos, e classificá-las em conjuntos propriamente especificados pelo desenvolvedor. Estas características tornam o aprendizado de máquina uma poderosa ferramenta no tocante à trabalhos relacionados ao campo de detecção de imagens.

2.4.1. Redes Neurais

As redes neurais são um conjunto pertencente à area de aprendizado de máquina e estão no núcleo dos algoritmos de deep learning. Elas simulam o comportamento do cérebro humano, permitindo que programas de computador reconheçam padrões e resolvam problemas comuns nos campos de inteligência artifical e *machine learning*.

Elas são compostas por camadas de um nó, contendo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada nó, ou neurônio artificial, conecta-se a outro e tem um peso e um limite associados. Se a saída de qualquer nó individual estiver acima do valor do limite especificado, esse nó será ativado, enviando dados para a próxima camada da rede. Caso contrário, nenhum dado será transmitido para a próxima camada da rede [2].

As redes neurais contam com dados de treinamento para aprender e melhorar sua precisão ao longo do tempo. No entanto, uma vez que esses algoritmos de aprendizagem são afinados para precisão, eles são ferramentas potentes na ciência da computação e na inteligência artificial, permitindo-nos classificar e agrupar dados a uma alta velocidade [2].

3. Objetivos

O principal objetivo do projeto é desenvolver uma aplicação que possa identificar, classificar e traduzir sinais de libras através de imagens disponibilizadas.

Os objetivos específicos do projeto são:

- Estudar sobre os métodos de detecção de objetos de interesse em imagens digitais
- Estudar as técnicas de redes neurais cabíveis para o projeto
- Realizar o processamento nas imagens para adquirir características
- Determinar classificações dos sinais para a rede neural de acordo com as características informadas
- Obter a identificação dos sinais de libras a partir da saida dos classificadores neurais

4. Procedimentos metodológicos/Métodos e técnicas

Em primeira instância será levantada uma referência mais ampla nos assuntos de detecção em imagens. Com essa base de referências serão avaliadas as propostas por esses trabalhos e pequisas para realizar primeiramente a identificação das mãos. Baseado no método

escolhido para a identificação, serão estudadas as ferramentas necessárias para realizar o processo.

Após o processo de separação das mãos nas imagens, será feita a extração de características das imagens processadas. As informações coletadas serão passadas como entrada para a rede neural, que passará por um período de treinamento e no fim de seu processo irá determinar as classes que cada sinal pertence. Cada classe deve representar o significado de seu respectivo sinal de Libras.

5. Cronograma de Execução

Atividades:

1. Levantamento de referências no campo de detecção de imagem e objetos;
2. Levantamento de trabalhos relacionados à detecção de Libras;
3. Determinação das ferramentas a serem usadas
4. Estudo da utilização das ferramentas;
5. Implementação do algoritmo para o processamento das imagens;
6. Implementação do algoritmo de rede neural;
7. Treinamento do algoritmo;
8. Avaliações e testes de assertividade do algoritmo;
9. Escrita do TCC;

Tabela 1. Cronograma de Execução

	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai
Atividade 1	X									
Atividade 2	X									
Atividade 3	X	X								
Atividade 4	X	X								
Atividade 5			X	X						
Atividade 6				X	X					
Atividade 7						X				
Atividade 8							X			
Atividade 9							X	X	X	X

6. Contribuições e/ou Resultados esperados

Espera-se que este trabalho possa contribuir de forma prática com as pessoas que trabalham com a comunicação e que desejam entender ou aprender mais sobre a linguagem de Libras, facilitando a comunicação entre pessoas que possuem ou não deficiência auditiva. Além do mais, é esperado que este trabalho ajude no desenvolvimento de outros próximos trabalhos que envolvam o mesmo tema, servindo assim como um estudo de processamento de imagem, detecção de objetos de interesse em imagens digitais e o uso de arquiteturas de redes neurais para gerar e agrupar dados.

7. Espaço para assinaturas

Londrina, *10 de Setembro de 2022*.

Gabriel

Aluno

Guilherme P. Cardim

Orientador

Referências

[1] Japan Earthquake Information Center. Intel® integrated performance primitives 8.2 update 1 reference manual. http://wwweic.eri.u-tokyo.ac.jp/computer/manual/eic2015/doc/intel/en_US/ipp/ipp_manual/GUID-AE698C04-81DB-402B-88E7-2BEED820D4DF.htm, 2015.

[2] IBM Cloud Education. Redes neurais. https://www.ibm.com/br-pt/cloud/learn/neural-networks, Agosto 2020.

[3] S. Du F. T. Timbane and R. Aylward. Hand gesture recognition basend on the fusion of visual and touch sensing data. *Advances in Visual Computing*, 2020.

[4] Adenilson Giovanini. Processamento digital de imagens: o que é? https://adenilsongiovanini.com.br/blog/processamento-digital-de-imagens/, 2016.

[5] GlobalGeo. Processamento digital de imagens. https://www.globalgeo.com.br/servi\unhbox\voidb@x\setbox\z@\hbox{c}{\lineskiplimit-\maxdimen\unhbox\voidb@x\vtop{\baselineskip\z@skip\lineskip.25ex\everycr{}\tabskip\z@skip\halign{#\crcr\unhbox\z@\crcr\hskip\hideskip\char24\hskip\hideskip\crcr}}}os/processamento-digital-de-imagens/.

[6] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Pearson, 3rd edition, 2007.

[7] Vihar Kurama. Ml-based image processing. https://nanonets.com/blog/machine-learning-image-processing/, September 2021.

[8] Luciana R. Veloso e Waslon T. A. Lopes Oeslle A. S. Lucena, Ítalo de P. Oliveira. Detecção de pele baseada em modelos de cor. *Simposio Brasileiro de Telecomunicações*, 2016.

[9] Graciele Kerlen Pereira. Curso de libras. https://ufsj.edu.br/portal2-repositorio/File/incluir/libras/curso_de_libras_-_graciele.pdf, September 2010.

[10] Hebert Luchetti Ribeiro. Reconhecimento de gestos usando segmentação de imagens dinâmicas de mãos baseada no modelo de mistura de gaussianas e cor de pele. *Tese de mestrado, Escola de Engenharia de São Carlos - Universidade de São Paulo.*, 2006.

[11] Leonardo Torok. Método de otsu. *Instituto de Computação – Universidade Federal Fluminense (UFF)*, 2015.

[12] João Pedro Ferreira Valente. Modelo rgb. `https://sites.google.com/site/aimcjv/modelo-rgb`.