

Reconhecimento de LIBRAS em Imagens Utilizando Histograma de Gradientes Orientados e Máquina de Vetor Suporte

Renan Praciano I. Leal Sandes, Rassa Bezerra Rocha,

Resumo—Acessibilidade é um tema que atrai muitos pesquisadores com intuito de proporcionar melhorias na qualidade de vida de pessoas que apresentam algum tipo de deficiência. Este trabalho apresenta a implementação de algoritmo que utiliza Histograma de Gradientes Orientados e Máquina de Vetor de Suporte para o reconhecimento de imagens que contém símbolos do idioma LIBRAS. Os resultados apresentam uma taxa de acerto de, no mínimo, 75%.

Index Terms—Processamento de Imagens, Extração de Características, Histograma de gradientes orientados, Máquina de vetor suporte.

I. INTRODUÇÃO

SEGUNDO o Instituto Brasileiro de Geografia e Estatística (IBGE), o Brasil tem cerca de 23,9% da população com algum tipo de deficiência, dentre elas visual, auditiva, motora e mental ou intelectual [1].

Particularmente no caso da auditiva, as pessoas que possuem tal deficiência têm a opção de se comunicar por meio da Língua Brasileira de Sinais (LIBRAS), instituída como meio legal de comunicação e expressão pela Lei federal Nº10.436 de 24 de abril de 2002 [2]. No entanto, embora LIBRAS seja o segundo idioma do Brasil, boa parte da população é leigo nesse idioma, fazendo com que a comunicação com pessoas com deficiência auditiva não seja fluente.

Como definido por [3], a LIBRAS é uma modalidade gestual-visual que utiliza de um meio visual-espacial na elaboração de sinais, sendo é necessário a visualização dos gestos para o entendimento da mensagem. Os sinais de LIBRAS são compostos por cinco parâmetros:

- Configuração das mãos (CM) - Formas que são postas as mãos para a execução do sinal;
- Ponto de Articulação (PA) - Lugar onde incide a mão configurada para a execução do sinal;
- Movimento (M) - Deslocamento das mãos durante a execução do sinal;
- Orientação ou Direcionalidade (O/D) - Direção da movimentação das mãos durante o sinal;
- Expressão facial e/ou corporal(EF/C) - Complementos ao sinal feitos pelo corpo ou pela face para dar vida e entendimento ao sinal executado.

Este artigo tem o objetivo de desenvolver um sistema que auxilie a interação entre indivíduos que possuem pouca ou

nenhuma audição com pessoas que não têm o conhecimento com o idioma LIBRAS. Dessa forma, pretende-se implementar um sistema que seja capaz de reconhecer o alfabeto em LIBRAS e o converter para o idioma Português. É proposto que, inicialmente, esse sistema seja aplicado em determinados ambientes, como restaurantes, shoppings e aeroportos.

A parte inicial do sistema de conversão consiste no reconhecimento de gestos do idioma LIBRAS em imagens. Desde os anos 90 o estudo de reconhecimento de gestos para as mais variadas aplicações têm sido feitos. Em [4] é relatada parte dessa história. Para o reconhecimento, este trabalho fez o uso de um descritor que utiliza os Histogramas de Gradientes Orientados (HOG) para que tais características extraídas sejam inseridas em um classificador Máquina de Vetor Suporte.

Além desta seção introdutória, este artigo está dividido em mais três seções. A Seção II descreve o algoritmo de reconhecimento de gestos, enquanto as Seção IV e Seção V apresentam, respectivamente os resultados obtidos e as considerações finais do trabalho.

II. O ALGORITMO

Em [5], são discutidas características importantes para o reconhecimento de gestos. Um sistema a base de reconhecimento de gestos deve ser independente do fundo, insensível à luz e impessoal para que seja possível robusta detecção dos gestos. Os HOG são caracterizados por ter boa insensibilidade à luz, e conseguem reconhecer os contornos muito bem, como utilizado para o reconhecimento de pessoas em imagens em [6].

Para a extração do vetor de características HOG de imagens 128x128 *pixels*, inicialmente foi realizada a aplicação de máscaras espaciais $[-101]$ e $[-101]^T$ para extrair os gradientes como em [7]

Em seguida, faz-se necessário o cálculo dos ângulos e redução dos mesmos. Para isso, a imagem foi dividida em blocos de tamanho 8x8 *pixels* e, para cada bloco, é calculado o histograma contendo 9 *bins* com a magnitude ponderada. Os *bins* são 0, 20... 140, 160. Caso um elemento esteja entre bin_i e bin_{i+1} , a magnitude ponderada para o bin_i é dada por: $mag_i = mag * (\theta - bin_i) / bin_{i+1} - bin_i$. A Figura 1 ilustra o histograma para uma imagem do banco de imagens utilizado para testes [6], [7].

Posteriormente, os histogramas foram normalizados em blocos de 16x16 *pixels*, ou seja, os histogramas de 4 células dos blocos obtido foram normalizados utilizando a norma euclídi-

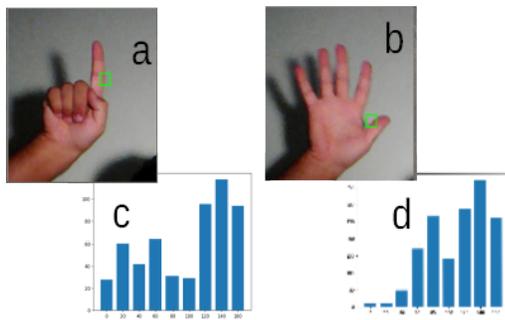


Figura 1. a) Amostra da classe 1 do banco de imagens. b) Amostra da classe 5 do banco de imagens. c) Histograma para a região 8x8 demarcada na imagem (a) d) Histograma para a região 8x8 demarcada na imagem (b)

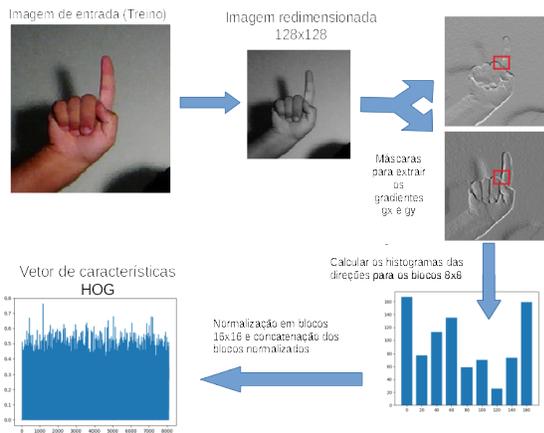


Figura 2. Processo de extração do vetor de características HOG.

ana, com o objetivo de reduzir a influência da iluminação desigual. A normalização euclidiana é dada por: $\|v\| = \sqrt{\sum a_i^2}$ onde os histogramas serão divididos por $\|v\|$ e a_i são os elementos dos histogramas das células.

Os vetores normalizados são então concatenados formando o vetor de características HOG. Para as imagens utilizadas [3] os vetores HOG contém 8100 elementos que são aplicados à algoritmos de aprendizado de máquina para então realizar a [4] classificação de tais imagens. A Figura 2 ilustra uma visão do processo por inteiro.

III. MÁQUINA DE VETOR DE SUPORTE -SVM

Primeiramente descrito em [8], a SVM é um método de aprendizado de máquina supervisionado baseado em um mapeamento não linear afim de encontrar um hiperplano que separa as classes de dados (limiar de decisão). Neste trabalho a SVM [8] será implementada com o auxílio da biblioteca `sklearn.svm`.

IV. RESULTADOS

Para avaliar o desempenho do sistema desenvolvido, um bando de imagens foi criado a partir de 100 imagens, sendo 50 pertencentes à classe 5 e 50 pertencentes à classe 1 (Exemplos expostos na figura 1 (a) e (b)). As imagens que compõem o banco de imagens não apresentam problemas de oclusão e não foram processadas antes de serem aplicadas ao método

proposto. Além disso possuem um bom contraste entre a mão e o fundo da imagem, como visto na figura 1.

O algoritmo implementado com a linguagem de programação Python recebe imagens pertencentes à dois grupos distintos que se diferenciam por conterem imagens de duas posições de mãos, conforme a figura 1. O programa implementado segue as etapas apresentadas na figura 2.

Para a etapa de treino, foram realizados 30 imagens de cada classe, enquanto 20 imagens de cada classe foram utilizadas no teste do algoritmo proposto. Os resultados indicam boa taxa de classificação, com taxa de acertos ($TaxaDeAcertos = (acertos/erros)$) de 85% para a classe 1 e 75% para a classe 5.

V. CONSIDERAÇÕES FINAIS

Este artigo apresenta a implementação de um algoritmo de classificação de imagens que ilustram símbolos do idioma LIBRAS, no intuito de ser aplicado a um sistema de conversão LIBRAS-Português. O método extrai os histogramas de gradientes orientados das imagens e os utilizam como entrada para a classificação utilizando a Máquina de Vetor de Suporte.

Os resultados indicam que o algoritmo apresenta bons resultados para o objetivo proposto, com taxa de acertos ($TaxaDeAcertos = (acertos/erros)$) de 85% para a classe 1 e 75% para a classe 5.

Como trabalhos futuros, pretende-se aumentar a quantidade de treino e testes do algoritmo, incluindo novas classes de imagens. Além disso, é objetivo futuro, após a classificação, a conversão da informação contida na imagem reconhecida para o Português.

REFERÊNCIAS

- [1] IBGE, *Normas de apresentação tabular*, 3rd ed. Rio de Janeiro: Centro de Documentação e Disseminação de Informações. Fundação Instituto Brasileiro de Geografia e Estatística, 218.
- [2] *Lei nº 10.436/2002*, Plenário. Relator: Paulo Renato Souza ed., Brasil. Casa Civil, Brasília, 24 de janeiro de 2007, 24 de abr. 2002. [Online]. Available: https://www.planalto.gov.br/ccivil_03/leis/2002/110436.htm
- [3] M. Honora and M. Frizanco, *Livro ilustrado de Língua Brasileira de Sinais: desvendando a comunicação usada pelas pessoas com surdez*. Ciranda Cultural, 2008.
- [4] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.
- [5] M. Kölsch and M. Turk, "Robust hand detection." in *FGR*, 2004, pp. 614–619.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.
- [7] O. A. S. Lucena and W. T. A. Veloso, L. R. and Lopes, "Implementação de um sistema de reconhecimento de objetos em imagens," *Revista de Tecnologia da Informação e Comunicação*, vol. 6, no. 2, pp. 34–42, 2016.
- [8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.