

# Metodologia para a segmentação e extração de características de gestos estáticos em imagens.

*Otávio Nathanael Campos de Oliveira 1, Diego Saqui (Orientador), Anderson Pereira das Neves (Coorientador)*

*Instituto Federal de Educação, Ciência e Tecnologia, Corumbá –MS*

---

## RESUMO

O presente artigo aborda uma metodologia que segmenta gestos estáticos de linguagens de sinais, utilizando um descritor composto por características em sua maioria de baixo nível e de aspecto dimensional, extraídas de imagens estáticas com gestos. Para avaliar a relevância dessas características, uma base de dados composta por elas é gerada e utilizada em testes que empregam os algoritmos de classificação *K-nearest neighbours* (KNN), Árvores de Decisão com C4.5 e *Support Vector Machines* (SVM) e avaliam seus resultados por meio da técnica de validação cruzada. Também se utilizam de técnicas de seleção de características com o propósito de se reduzir os custos computacionais e tempo de treinamento, além de tentar proporcionar a melhoria da classificação dos algoritmos considerados. Tal metodologia se apresenta agnóstica quanto à natureza do gesto, visto que, ela não foi desenvolvida para atender especificamente um gesto estático, como por exemplo, os de Libras. A metodologia desenvolvida além de contribuir para elaboração de sistemas de baixo custo que visam realizar a tradução da linguagem de sinais, também pode auxiliar em outros sistemas que fazem uso do reconhecimento de gestos, como jogos e sistemas de navegação. Resultados com uma base contendo todas as características extraídas, dezoito ao todo, proporcionaram aos classificadores KNN, C4.5 e SVM taxas de acurácia de 91%, 88% e 93%, respectivamente. Resultados com as características restantes após a seleção, oito ao todo, foram 92%, 87% e 91%, respectivamente. Tais resultados indicam que a segmentação em conjunto com as características extraídas pode ser adequada para tanto o reconhecimento de gestos estático de uma língua de sinais como Libras, quanto para outro fim que demande o reconhecimento de gestos.

**Palavras-chave:** Gestos, Libras, Visão Computacional, Extração, Segmentação.

---

## 1 INTRODUÇÃO

De acordo com o censo realizado pelo IBGE em 2010, cerca de 5% da população brasileira apresenta alguma deficiência auditiva, e entre eles estão as pessoas que se comunicam em língua de sinais brasileira (Libras). Libras é reconhecida pelas leis brasileiras como um meio de comunicação oficial da comunidade surda, de acordo com Oliveira (2006). Porém, segundo Toscano, Dizeu e Caporali (2005), a linguagem oral para a sociedade é imperativa e qualquer outra forma de comunicação acaba sendo desconsiderada. Embora existam iniciativas sociais que visam mudar essa realidade, têm-se buscado novas abordagens para o problema. Além de outras iniciativas, pesquisadores e desenvolvedores têm tentado contornar essa

situação ao fazer uso de recursos computacionais. O aplicativo Hand Talk e a suíte Vlibras são dois grandes exemplos de novas abordagens, porém ambas se limitam apenas a realizar traduções da Língua Portuguesa para Libras. Contudo, já há estudos que buscam realizar o inverso, sendo o uso de recursos de Visão Computacional (VC) uma das formas abordadas.

Segundo Barelli (2018), VC é o campo onde se tenta modelar a visão humana, através do uso de recursos computacionais como câmeras e sistemas de computadores, para diversos fins.

Um sistema que traduz gestos de Libras tem-se como principal processo analisar e determinar gestos capturados. Após análise, se tem a exposição do significado do gesto. Entre esses dois processos,

está a segmentação do objeto alvo, onde de acordo com Barelli (2018) é um dos procedimentos fundamentais para qualquer sistema de VC, sendo no presente trabalho a etapa onde a área da mão será separada do restante da imagem. Um objeto em uma cena de imagem ou vídeo possui várias características que o separa do resto. Essas características reunidas, segundo Nixon e Aguado (2012), compõem um descritor que será utilizado para dar identidade para o objeto segmentado da imagem.

Vários tipos de características e metodologias são explorados na literatura. Yeo, Lee e Lim (2015) propõem uma abordagem que é basicamente dividida em três módulos sendo estes, câmera, detecção e interface. As principais características extraídas estão relacionadas aos contornos do objeto segmentado, aproximação de polígonos, determinação do centro e raio da palma da mão, definição da região de interesse, extração convexa de defeitos de casco e convexidade, localização e direção determinantes das mãos, determinação das pontas dos dedos. Bastos, Angelo e Loula (2015) utilizaram dois descritores de forma e uma Rede Neural de dois estágios para reconhecer gestos em imagens digitais em seu trabalho, que são histograma orientado a gradientes e momentos invariantes de Zernike. Essa abordagem mostrou altas taxas de reconhecimento, atingindo uma média de 96%. Ibrahim, Selim e Zayed (2018) apresentam em seu trabalho um sistema que traduz sinais de palavras árabes isoladas em texto, obtendo-se uma taxa de reconhecimento de 97%. Características geométricas da mão são extraídas para compor um descritor, onde este é utilizado por um algoritmo classificador baseado em distância euclidiana.

O presente estudo descreve uma metodologia que busca segmentar gestos estáticos de linguagem de sinais em imagens e realizar a extração de características em sua maioria de baixo nível e de aspecto dimensional dos mesmos. Segundo Murthy e Jadon (2009), tais características são consideradas de baixo nível por não mapearem um aspecto muito específico do objeto de interesse, possibilitando descrever de forma geral a identidade de tal objeto. Também estão inclusas características que mapeiam um aspecto bem específico do objeto, porém grande

parte destas se mostram ineficazes na etapa de seleção de características, demonstrando assim a importância das de baixo nível. Por isso, a metodologia no fim se apresenta agnóstica quanto à natureza dos gestos estáticos, ou seja, ela pode ser utilizada tanto para mapear a identidade de gestos da língua de sinais, como Libras, quanto a de gestos utilizados em outros tipos de sistemas, como o de jogos. Uma base de dados contendo essas características é gerada para serem testada por diferentes algoritmos classificadores, a fim de avaliar sua relevância para o campo do reconhecimento de gestos.

## 2 MATERIAIS E MÉTODOS

A primeira parte desta seção trata dos recursos e técnicas utilizados no presente trabalho. A segunda parte descreve o fluxo da metodologia utilizada para realizar a segmentação e extração das características de gestos estáticos.

### 2.1 Recursos e técnicas utilizadas

Os principais recursos utilizados no presente trabalho são a plataforma Anaconda, as bibliotecas OpenCV, Scikit-Learn, Pandas e a linguagem de programação Python. Anaconda (2018) é uma plataforma *open source* que trabalha com ciência de dados que conta com vários recursos como ambientes virtuais que permitem a fácil instalação de recursos como a biblioteca OpenCV, SciKit-Learn, Pandas e Python. OpenCV (2018) é uma biblioteca de processamento de imagens multiplataforma que integra coleções de métodos já prontos que facilitam no desenvolvimento de sistemas de Visão Computacional. Scikit-Learn (2011) é uma biblioteca *open-source* que reúne algoritmos relacionados ao aprendizado de máquina, como os algoritmos classificadores usados no presente trabalho. Pandas (2019) também é uma biblioteca *open-source* que facilita operações e manipulações de base de dados. Python (2018) é uma linguagem de programação de alto nível que tem se tornado vastamente utilizado no campo da ciência de dados. Python foi escolhido principalmente por ser compatível com todos os

recursos citados, e por apresentar uma curva de aprendizado menor do que outras linguagens como C ou JAVA.

Para aplicar e avaliar a metodologia proposta, o presente trabalho utilizou uma base de imagens, de Marie (2018), com cerca de 78000 imagens, subdivididas por cerca de 26 classes, onde a classe é nada mais do que gestos estáticos relacionados à linguagem de sinais. Importante apontar que, a metodologia de segmentação e extração de características não está presa a somente um dado grupo de gestos estáticos pré-definidos. As informações obtidas na fase de extração quando combinadas conseguem descrever de forma geral a natureza de um gesto, diferente de trabalhos como o de Yeo, Lee e Lim (2015), onde os autores determinam quais gestos serão reconhecidos pelo sistema. Um banco de dados composto pelas informações obtidas na etapa de extração é gerado e aplicado em testes, que empregam o uso da técnica seleção de características, validação cruzada, e o uso de três algoritmos classificadores, *K-nearest neighbours* (KNN), Árvore de Decisão C4.5 e *Support Vector Machines* (SVM).

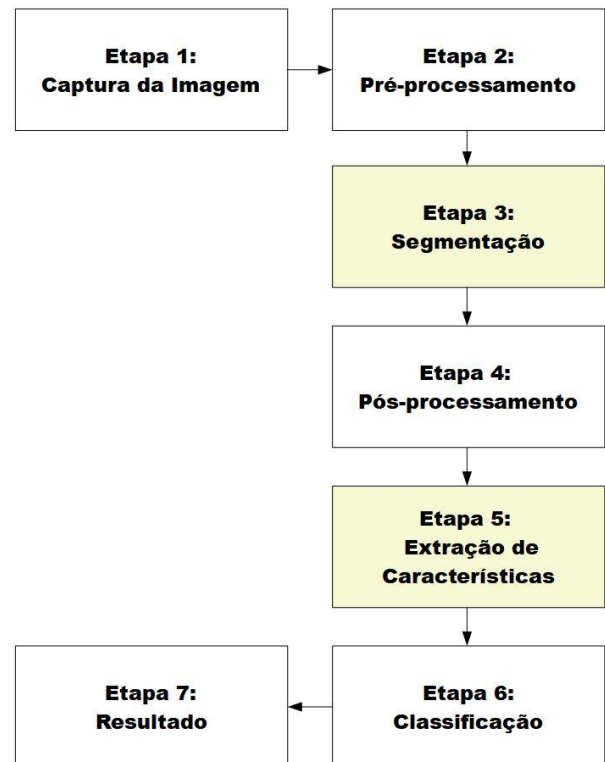
Os algoritmos classificadores utilizados foram implementados na biblioteca SciKit-Learn, e foram escolhidos por serem simples e de fácil uso. Para a avaliação dos classificadores, se é considerado a acurácia geral (*overall accuracy-OA*). Durante os testes, se utiliza a técnica de validação cruzada onde se divide a base em vários grupos para serem intercalados entre o treinamento dos classificadores e o teste dos mesmos quando treinados. No presente trabalho, se utiliza diferentes quantidades de grupos, de 2 a 10 grupos especificamente, para avaliar o desempenho do algoritmo treinado em relação à quantidade de dados disponível em cada grupo em cada teste.

## 2.2 Fluxo da metodologia empregada para a extração e classificação de gestos estáticos

Embora usado de diferentes formas por várias áreas, os sistemas de VC possuem um fluxo básico em comum. Segundo Barelli (2018), O fluxo se inicia com a captura da imagem, após vem pré-processamento, segmentação, pós-processamento,

Data: 06/11/2019

extração de características, classificação e, por último, a exibição do resultado. Esse fluxo pode ser observado na Figura 1, e nota-se que duas etapas estão destacadas das demais, sendo ambas as etapas o principal foco do presente trabalho. Todas as etapas serão explicadas em sequência nos próximos parágrafos.



**Figura 1:** Fluxo básico de um sistema de visão computacional baseado em Barelli (2018).

**Etapa 1: Aquisição da Imagem** - O fluxo se inicia com a aquisição da imagem, o que pode ser realizado tanto câmera digital simples quanto por uma com sensores mais específicos. No caso do presente trabalho, as imagens usadas foram obtidas no repositório do github de Marie (2018), onde se tem cerca de 26 gestos estáticos separados em subdiretórios dentro do conjunto de dados.

**Etapa 2: Pré-processamento** - é aplicado, para melhorar a qualidade da imagem ou para realçar um determinado aspecto da mesma para a próxima fase. Filtro gaussiano e mediano foram aplicados para suavizar os pixels da imagem a fim de deixar sua

coloração mais uniforme. A aplicação de tais filtros nem sempre é muito visível para o olho humano, mas servem como a principal solução contra o dano causado por ruídos, onde segundo Barelli (2018), são pequenas variações aleatórias que o sinal sofre durante o processo da captura da imagem, impossibilitando a leitura de seu valor real. Não podendo controlar o aparecimento desses defeitos na imagem, utiliza-se filtros com baixa intensidade.

**Etapas 3: Segmentação** - A etapa da segmentação visa separar a mão, objeto de interesse, do restante da imagem para que se extraia somente informações relacionadas a ela. Segundo Barelli (2018), essa etapa é fundamental para um sistema de VC visto que sua qualidade pode influenciar diretamente as próximas etapas, seguindo o fluxo descrito na figura 1. Para se segmentar um determinado objeto, é preciso identificar características que o diferencie do restante. No presente trabalho, o aspecto da cor do objeto foi escolhido como principal parâmetro diferenciador, por o mesmo ser facilmente distinguível da cor de vários objetos e elementos que podem estar presentes na imagem.

Estudos como o de Yeo, Lee e Lim (2015) apontam que o padrão de cor YCrCb captura o tom de pele humano de forma mais eficiente, onde Y é o canal que armazena a luminância, Cr e Cb armazenam informações relacionadas ao croma. As imagens capturadas então são convertidas para esse padrão de cor antes de iniciar o processo de segmentação.

Dentro desse processo, Black Hat é aplicado no canal Y da imagem convertida, que armazena os tons amarelos, a fim de aumentar o contraste do objeto de interesse em relação aos outros elementos presentes na imagem. Black Hat é um método bem direto e simples, onde entra com a função de gerar um grande contraste na cor dos objetos que compõe a imagem. Desta forma, podemos iniciar o processo de diferenciar o objeto de interesse do restante da imagem. Após Black Hat, vem o método Binarização de Otsu. Ambas as operações já estão prontas para uso na biblioteca de processamento de imagem OpenCV.

A Binarização de Otsu é a primeira tentativa significativa de segmentar a mão do restante da imagem. Como o próprio nome indica, esse método vai binarizar a imagem, ou seja, por o seu padrão de cor em um espaço binário, preto sendo a parte a não ser analisada e branco sendo o objeto segmentado. Essa binarização, segundo Barelli (2018), diferente de algumas outras presentes na biblioteca OpenCV, funciona de forma automática, não necessitando da definição de parâmetros que deem alguma indicação sobre o objeto de interesse. Esse método busca por objetos que se diferenciam bastante, em termos de cor, de outros elementos da imagem, então faz com que somente estes permaneçam na imagem. Porém, como ele não está buscando especificamente pelo objeto de interesse, ao utilizá-lo, se acaba segmentando outros objetos e elementos de forma involuntária. Então mais métodos precisam ser aplicados para corrigir esse tipo de problema.

Detector de contorno de Canny é utilizado para se extrair o contorno dos objetos segmentados com base em Yeo, Lee e Lim (2015). Durante o processo de segmentação, partes do fundo ou de outros objetos podem acabar grudadas no objeto de interesse. A separação de objetos “grudados” pode se dar ao se capturar o contorno de ambos com o detector, e então utilizar o contorno capturado para eliminar a região em volta dos objetos. Com os objetos devidamente separados um do outro, e indo com base na hipótese de que a mão seja o maior objeto ainda presente na imagem, realiza-se então uma segmentação por tamanho. Essa segmentação vai identificar o objeto com maior área presente na imagem, e irá eliminar todos os outros com menor. Até esse ponto assume-se que a mão seja o único objeto a ir para a próxima fase.

**Etapas 4: Pós-processamento** - Vem após a fase de segmentação com o intuito de corrigir falhas pelo corpo do objeto. A segmentação em si não é um processo muito preciso, e no ato de retirar elementos da imagem que não foram definidos como sendo pertinentes, partes do próprio objeto de interesse podem ser erroneamente removidos, apesar dos esforços para o contrário. Então, após a



segmentação, operações morfológicas da biblioteca OpenCV são aplicadas para eliminar ruídos, fechar “buracos” no objeto, ou até mesmo eliminar sobras de objetos não pertinentes. Não é sempre que tais defeitos acontecem, porém essas operações são aplicadas como forma de prevenção.

O pós-processamento marca o fim da fase de segmentar o objeto de interesse. Com o objeto de interesse já segmentado, se inicia a fase de extração de características. Ainda segundo Barelli (2018), essa fase fica encarregada de extrair informações do objeto de interesse que permitam classificá-lo ou dar-lhe uma identidade. Para o presente trabalho, a principal fonte de informação é o contorno da mão. A partir do contorno, é possível obter informações como a área, o perímetro e o formato do objeto. O processo descrito para segmentação e extração é realizado em cerca de 7800 imagens do repositório do github de Marie (2018), subdivididas em 26 gestos estáticos, cerca de 300 imagens para cada gesto.

**Etapas 5: Extração de Características** - Os métodos utilizados para extrair a maioria das informações durante esta etapa se encontram implementados na biblioteca de processamento de imagens OpenCV. A primeira informação a ser extraída é o centro do objeto obtido através dos momentos de uma imagem, sendo estes segundo Barelli (2018), funções matemáticas que retornam valores que representam um determinado objeto matematicamente. Momentos não caem necessariamente na classe de características de baixo nível, mas sim em características inerciais que fornecem informações geométricas a respeito do objeto de interesse. Depois as informações relacionadas à palma da mão, que é o maior círculo que pode ser encontrado dentro do contorno da mão segundo Yeo, Lee e Lim (2015), e o círculo que cobre completamente o contorno da mão usando a menor área possível segundo Barelli (2018). De ambos são extraídos o centro ( $x$ ,  $y$ ), o raio ( $R_a$ ) e a área (área). Essas informações foram escolhidas com base na hipótese de que cada gesto possui um tamanho diferente, ou ocupa áreas de tamanhos

diferentes. Isso pode ser observado ao se comparar o gesto estático A e B. Em seguida, é extraída a quantidade de dedos presente na imagem, visto que, determinados gestos utilizam uma quantidade diferente de dedos. Essa extração se dá com base nas descrições de Yeo, Lee e Lim (2015) e Bakar et. al. (2015) que se inicia com a captura dos pontos de início e fim de cada convexidade para serem analisados, visto que, na teoria, eles poderiam ser a ponta de um dedo. Para verificar essa hipótese, o ângulo de cada convexidade presente na imagem é calculado, assim como se verifica se o ponto está em uma região acima do raio do círculo da palma da mão e se permanece dentro do raio do círculo que abrange a região da mão. Por último o ângulo de cada convexidade é avaliado, se considera uma convexidade pertencente a um dedo quando o ângulo da mesma é igual ou menor que 60 graus. Então conta-se a quantidade de convexidades que passaram pela validação para se obter o possível número de dedos na imagem.

O último grupo de informações a serem extraídas é a massa central do objeto calculada a partir dos momentos da imagem e os sete momentos invariantes descritos por Hu (1962), que podem fornecer informações como a área do objeto e até ser usado como um vetor de recurso. Essas informações buscam descrever o formato de um dado objeto de forma numérica, e elas podem ser capazes de discernir um gesto estático do outro. Diferente das outras informações extraídas, esses momentos invariantes são características de alto nível visto que eles buscam descrever o objeto de forma específica. Eles foram escolhidos a princípio para principalmente ajudar a reduzir a similaridade de um gesto em relação ao outro. Esses momentos podem ser extraídos com a ajuda da biblioteca OpenCV.

Ao fim da etapa de extração, colhe-se a classe do gesto, que no caso da base de imagens utilizada no presente trabalho, está na descrição da imagem. Com essas informações, um banco de dados com as informações extraídas de cada classe é gerado para treinar os algoritmos classificadores na próxima

etapa. Porém, antes de prosseguir, uma seleção de características é realizada a fim de selecionar as informações mais importantes para a classificação dentro da base de características geradas. Tal seleção, segundo Guyon e Elisseeff (2003), contribui para a redução de recursos necessários tanto para o armazenamento dos dados em memória, quanto na redução do tempo de treinamento de cada algoritmo, que tende a aumentar à medida que mais dados são inseridos na base.

Para realizar a seleção, primeiro cria-se o algoritmo classificador Árvore de Decisão, e depois passa o algoritmo como *kernel* para o selecionador *SelectFromModel* onde se define um limiar de seleção que representa um ponto específico dentro da árvore de nós gerada durante o treinamento. Toda característica, ou informação, que se posicione acima desse ponto será considerada como relevante e então selecionada. Ambas as implementações do algoritmo classificador e selecionador utilizados são da biblioteca SciKit-Learn.

Com a seleção realizada, outra base é gerada contendo somente as características avaliadas como pertinentes, que também será utilizada na próxima etapa. Com ambas as bases em mão, será possível verificar na prática se todos os dados coletados são de fato pertinentes.

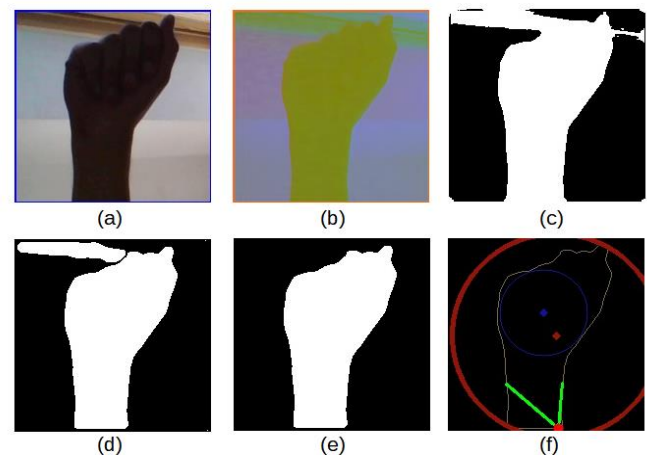
**Etapa 6: Reconhecimento de Padrões** - Os algoritmos classificadores de aprendizado supervisionado utilizados neste trabalho têm o propósito exclusivo de permitir a avaliação das características extraídas a partir da segmentação proposta. Para isso, eles, precisam saber a identidade dos objetos em fase de treinamento para que possam classificar um objeto desconhecido. Os algoritmos classificadores são treinados e passam por testes que empregam a técnica de validação cruzada, onde a base da vez é subdividida em quantidades de 2 a 10 grupos que serão empregados para treino e teste de forma intercalada. Tal técnica é utilizada a fim de evitar o viciamento dos algoritmos classificadores, que ocorre quando ele é testado pelos mesmos dados dos quais ele foi

treinado. Classificadores quando “viciados” tendem a retornar uma taxa de acerto enganosamente alta ou baixa, que de fato não correspondem à realidade.

Na próxima seção serão exibidos os resultados da metodologia (**Etapa 7: Resultados**), assim como formas de melhorar tais resultados e possíveis soluções para problemas encontrados durante a pesquisa.

### 3 RESULTADOS E DISCUSSÃO

A fase de segmentação é uma das principais fases de um sistema de Visão Computacional, onde seu resultado irá influenciar diretamente no desempenho das próximas fases, principalmente a fase da extração das características onde defeitos provocados por má segmentação passam a oferecer informações não relacionadas ao objeto em seu estado original. A figura 2 mostra os principais resultados obtidos durante a fase de segmentação e uma ilustração dos dados que foram segmentados.

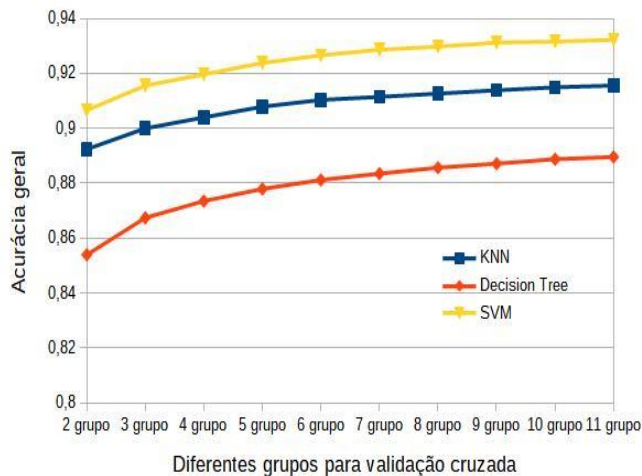


**Figura 2:** Resultados da segmentação e extração de características.

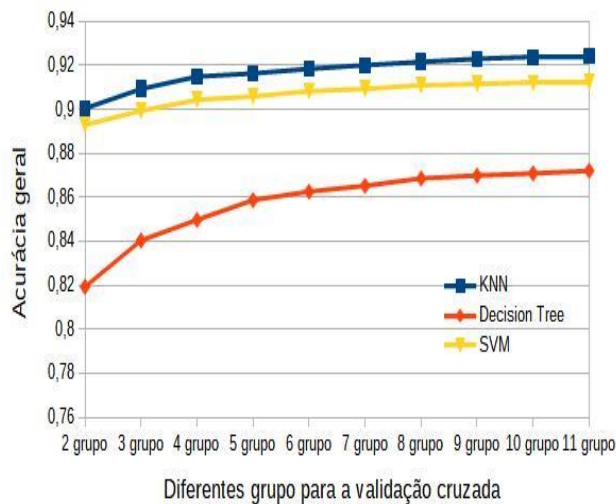
É possível observar que na imagem (c) da figura 2, há uma falha na primeira fase da segmentação. Esse tipo de falha é muito comum principalmente em situações de uso real. Embora a falha possa ser resolvida com a eliminação da área ao redor dos objetos e com a segmentação por área, da imagem (d) para (e), nem sempre essas técnicas vão ser capazes de segmentar a mão de forma satisfatória. Por isso, as informações colhidas na etapa de

extração devem ser robustas a esses problemas na medida do possível.

Indo da segmentação para extração e seleção de características, os resultados anteriores à seleção e posterior a seleção, são apresentados respectivamente nas Figuras 3 e 4.



**Figura 3:** Resultados de antes da seleção de características.



**Figura 4:** Resultados após a seleção de características.

Na figura 3 se tem os resultados dos algoritmos classificadores de antes da seleção de características. Observa-se que o algoritmo com maior acurácia é o SVM, com cerca de 93%, em

seguida vem o KNN, com cerca de 91%, e pôr fim, a árvore de decisão gerada com C4.5, com cerca de 88%. Observa-se que todos os algoritmos obtiveram altas taxas de acurácia geral, principalmente KNN e SVM. Já com esses resultados pode-se observar a relevância das características extraídas. Porém, ainda assim é aplicado uma seleção nessas características a fim de aumentar a eficiência tanto da base quanto dos algoritmos classificadores.

Ao contrário do que se esperava no início deste estudo, os cinco primeiros momentos invariantes de Hu, a quantia de dedos presente nos gestos e o centro da circunferência da palma da mão e de todo o objeto em si não têm um grande impacto na classificação geral. Os resultados presente nas Figuras 3 e 4 comprovam essa afirmação. Resultados com as características restantes após a seleção, oito ao todo, foram cerca de 92% para KNN, 91% para o SVM e 86% para a Árvore de decisão C4.5. A princípio não se pode observar grande diferença em comparação aos resultados anteriores à seleção, porém, segundo Guyon e Elisseeff (2003), a retirada das informações consideradas desnecessárias pelos métodos de seleção escolhidos contribuiu tanto para a redução do custo necessário para o armazenamento da base, tempo de carregamento da base pelo algoritmo e o tempo de treinamento de cada algoritmo.

## 4 CONCLUSÃO

O presente trabalho descreve uma metodologia que busca traduzir gestos estáticos de linguagens de sinais. Embora não tendo apresentado avanços em fases como a de segmentação e classificação, pois os métodos apresentados nessas fases são comuns e vastamente utilizados, o presente trabalho avança ao obter altas taxas de acerto ao utilizar características que tornam a metodologia em si agnóstica a uma determinada linguagem de sinais, ou a outros fins que também envolvem a classificação de gestos estáticos.

O principal problema que a fase de segmentação enfrenta é quando se tem na imagem um objeto semelhante ao objeto de interesse. No caso da base

de imagens utilizada, se possui no fundo um objeto de cor semelhante ao da cor da mão. O ideal seria que somente a região envolta da mão fosse para a segmentação, e não a imagem inteira, porque, quanto mais elemento se possui na imagem, mais difícil se torna a eliminação dos mesmos. Então, para reduzir à probabilidade de objetos indesejados irem para a fase de segmentação, a região envolta da mão deve ser localizada e extraída por meio de técnicas de reconhecimento com as de Redes Neurais Convolucionais. O contratempo dessa solução seria que, tais técnicas requerem bastantes recursos computacionais, reduzindo assim a possibilidade de se implementar um sistema mais acessível. Porém, existem técnicas intermediárias que, embora não tão eficientes e mais laborais quanto às Redes Neurais Convolucionais, conseguem desempenhar um reconhecimento de um dado objeto em uma imagem de modo eficaz, como é o caso do método Haar-Cascade e HOG + SVM. Próximos estudos abordaram essas duas técnicas para melhorar a etapa de segmentação da metodologia proposta pelo presente trabalho.

Quanto aos resultados obtidos pelos classificadores, observam-se diferentes resultados em testes anteriores e pós-seleção, embora não tão discrepantes. Em especial, nota-se que o algoritmo KNN obteve uma melhora após a seleção e o SVM teve uma queda. Esse fato pode estar diretamente relacionado a tanto à eficácia da fase de segmentação, quanto à natureza características extraídas, onde cada algoritmo utilizado no presente trabalho lida com elas de forma diferente. Como trabalho futuro é sugerido o estudo das causas deste comportamento a fim de verificar possíveis melhorias no método proposto.

Por fim, conclui-se com base nos resultados alcançados que a metodologia desenvolvida é relevante para o problema da classificação de gestos estáticos de línguas de sinais. Contudo, pelo fato da maior parte das características extraídas descreverem um aspecto geral a respeito do gesto estático e da metodologia não tentar mapear gestos específicos, é possível afirmar que a metodologia

pode contribuir também para outros tipos de sistemas que também empregam o reconhecimento de gestos, como os de jogos por exemplo. A classificação de um dado gesto estática, por fim, depende de fato da base de imagens e ao significado atribuído a ela.

## REFERÊNCIAS

- IBGE. Escolaridade e rendimento aumentam e cai mortalidade infantil. Disponível em: <<https://censo2010.ibge.gov.br/noticiascenso?view=noticia&id=1&idnoticia=2125/>>. Acesso em: 5 nov 2018.
- OLIVEIRA, R. A. G. G. NOTA SAJ Nº 4138/05-RG. Revista Jurídica da Presidência, v. 7, n. 76, 2006.
- TOSCANO, L. C.; DIZEU, B.; CAPORALI, S. A. A língua de sinais constituindo o surdo como sujeito. Educ. Soc, v. 26, n. 91, p. 583-597, 2005.
- BARELLI, F. Introdução a Visão Computacional: Uma abordagem prática com Python e OpenCV. ed. Casa do Código, 2018.
- NIXON, M.; AGUADO, A. S. Feature extraction and image processing for computer vision. Academic Press, 2012.
- YEO, H.; LEE, B.; LIM, H. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. Multimedia Tools and Applications, v. 74, n. 8, p. 2687-2715, 2015.
- BASTOS, I. L.; ANGELO, M. F.; LOULA, A. C. Recognition of static gestures applied to brazilian sign language (libras). In: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, 2015. p. 305-312.
- IBRAHIM, N. B.; SELIM, M. M.; ZAYED, H. H. An automatic arabic sign language recognition system (arslrs). Journal of King Saud University-Computer and Information Sciences, v. 30, n. 4, p. 470-477, 2018.
- MURTHY, G. R. S.; JADON, R. S. A review of vision based hand gestures recognition. International Journal of Information Technology



and Knowledge Management, v. 2, n. 2, p. 405-410, 2009.

ANACONDA. About Anaconda. Disponível em: <<https://www.anaconda.com/about-us/>> Acesso em: 15 ago. 2018.

OPENCV. About. Disponível em: <<https://opencv.org/about/>> Acesso em: 20 jul. 2018.

PANDAS. The pandas project. Disponível em: <<https://pandas.pydata.org/about.html>> Acesso em: 20 abr. 2019.

PYTHON. About. Disponível em: <<https://www.python.org/about/>> Acesso em: 20 jul. 2018.

MARIE, L. Simple sign language alphabet recognizer using Python, openCV and tensorflow for training Inception model (CNN classifier). Disponível em: <<https://github.com/loicmarie/sign-language-alphabet-recognizer>> Acesso em: 09 out. 2018.

OPENCV. **Contour Features**. Disponível em: <[https://docs.opencv.org/3.4/dd/d49/tutorial\\_py\\_contour\\_features.html](https://docs.opencv.org/3.4/dd/d49/tutorial_py_contour_features.html)> Acesso em: 09 mar. 2019.

BAKAR, M. Zabri Abu; SAMAD, Rosdiyana; PEBRIANTI, Dwi; MUSTAFA, Mahfuzah; ABDULLAH; Nor Rul Hasma. Finger application using K-Curvature method and Kinect sensor in real-time. In: 2015 International Symposium on Technology Management and Emerging Technologies (ISTMET). IEEE, 2015. p. 218-222.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. Journal of machine learning research, v. 3, n. Mar, p. 1157-1182, 2003.

HU, M. Visual pattern recognition by moment invariants. IRE transactions on information theory, v. 8, n. 2, p. 179-187, 1962.