

Reconhecimento da forma da mão em acordes de violão utilizando processamento de imagens

Hand shape recognition on guitar chords using image processing

RESUMO

Este trabalho tem como objetivo a utilização e avaliação de métodos de aprendizado de máquina aplicados em processamento de imagens para realizar a classificação de acordes de violão. Inicialmente, foram utilizadas técnicas de pré-processamento, para a eliminação de ruídos. Posteriormente, foi realizada a segmentação separando os objetos de interesse, mão e pestana do violão, do fundo da imagem. Em seguida, dos objetos segmentados foram extraídas as características das imagens. Por fim, foram utilizados os classificadores, e métodos de meta-aprendizagem, para realizar as classificações. Com isso pôde-se perceber acurácias melhores em relação à algoritmos de meta-aprendizagem, como no caso do classificador que utiliza do método 1-contra-todos, atingindo a casa dos 95,9% em alguns casos.

PALAVRAS-CHAVE: Aprendizado de máquina. Meta-aprendizagem. Violão.

Anderson Makoto Shinoda
shinoda@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Claiton de Oliveira
claitonoliveira@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autorial: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

THIS PROJECT'S GOAL IS USE AND EVALUATE MACHINE LEARNING METHODS APPLIED IN IMAGE PROCESSING TO PERFORM THE CLASSIFICATION OF GUITAR CHORDS. FIRSTLY, WERE USED PREPROCESSING TECHNIQUES FOR THE ELIMINATION OF NOISE. POSTERIORLY, THE SEGMENTATION WAS PERFORMED SEPARATING THE OBJECTS OF INTEREST, HAND AND NUT FROM THE GUITAR, FROM THE BACKGROUND OF THE IMAGE. THEN, FROM THE SEGMENTED OBJECTS, THE CHARACTERISTICS OF THE IMAGES WERE EXTRACTED. FINALLY, THE CLASSIFIERS AND META-LEARNING METHODS WERE USED TO PERFORM THE CLASSIFICATIONS. THEREBY, IT WAS POSSIBLE TO NOTICE A BETTER ACURACY USING META-LEARNING ALGORITHMS, AS IN THE CASE OF 1-AGAINST-ALL METHOD, REACHING 95,9% IN SOME CASES.

KEYWORDS: Guitar. Machine learning. Meta-learning.

INTRODUÇÃO

Com a necessidade de se desenvolver algoritmos computacionais que tivessem a capacidade de se adaptar a diferentes problemas, foi concebido o conceito de inteligência artificial.

A inteligência artificial é o estudo que objetiva desenvolver algoritmos que possam simular a percepção e cognição humana (HERGET, 2017). Através de processamento de dados e análise de padrões.

O aprendizado de máquina pode ser aplicado em vários campos, dentre eles, o de processamento de imagens, que consiste em realizar análises e extrair características de objetos da imagem para classificação (ALBUQUERQUE e ALBUQUERQUE, 2000, p. 1).

Este trabalho utilizou processamento de imagens para análise e classificação de acordes de violão.

MATERIAL E MÉTODOS

Esta seção é dividida em duas partes: materiais, onde serão descritas as ferramentas de *hardware* e *software* utilizados e metodologia, onde serão descritos os processos para a realização do trabalho.

MATERIAL

Foi utilizado um *notebook* Lenovo Thinkpad Edge 05782GP para processamento das imagens e um celular Moto G5 Plus para aquisição das fotos.

Foram tiradas 517 fotos na resolução 4032x3024 divididos em 64 imagens do acorde dó maior, 62 do acorde fá maior, 69 do fá# maior (fá suspenso), 64 do lá maior, 62 do mi maior, 70 do ré maior, 64 do si maior e 62 do acorde sol maior.

Para a implementação, foi utilizado a linguagem Python 3.7 com as bibliotecas OpenCV, Numpy e Os para processamento da imagem, manipulação de *arrays* e gerenciamento de pastas, respectivamente.

Para a classificação foi utilizado o *software* Weka3.8 que contém métodos já implementados de aprendizado de máquina (FRANK, HALL e WITTEN, 2016).

MÉTODOS

Os métodos foram divididos em 5 etapas, sendo elas: pré-processamento, segmentação, pós-processamento, extração de características e classificação.

Para o pré-processamento das imagens, inicialmente todas as imagens foram redimensionadas para a resolução de 1344x1008 para reduzir o custo computacional. Após isso, foi definido o espaço de cor YCrCb, que é um espaço de

cor baseado nos valores de azul e vermelho (ARAÚJO e RIBEIRO, 2007, p. 2) para o processamento das imagens.

Para eliminação de ruídos foi aplicado um filtro bilateral utilizando de uma máscara 11x11, essa máscara aplica uma função gaussiana nos *pixels* pertencentes a ela para suavização, mantendo-se a integridade das bordas das imagens.

Em relação ao processo de segmentação, tomou-se como objetos de interesse a mão que forma o acorde e a pestana do violão. Por conta da diversidade de ambientes e iluminação das imagens, não foi alcançado um único parâmetro que realizasse a separação dos objetos de interesse e o fundo de maneira eficaz, por conta disso, foram utilizados diferentes valores de parâmetros para cada imagem, sendo que as médias dos valores para a pestana são: entre 91 e 255 para Y, de 105 a 126 para Cr e 129 a 160 para Cb. Sabendo que os valores variam de 0 a 255 na biblioteca OpenCv.

No caso da mão, os valores médios de Y, Cr e Cb utilizados são: 60 a 138 para Y, 135 a 155 para Cr e 106 a 138 para Cb. Valores estes, obtidos empiricamente.

Para a etapa de pós-processamento, foram utilizadas as operações morfológicas de abertura (erosão seguida de dilatação) e fechamento (dilatação seguida de erosão), utilizando uma máscara 7x7 para a abertura e 5x5 para o fechamento.

Feito isso, também foi necessário remover a parte do objeto da mão referente ao braço, uma vez que essa porção do objeto não possui alguma característica relevante. Para isso foi calculado a razão da largura pela altura do objeto, e caso a razão fosse superior à 2.5, é removido a porção referente ao braço.

Por fim, para remover qualquer outro objeto que tenha sido segmentado com os de interesse, foram calculados os perímetros de todos os objetos segmentados e foram mantidos apenas os objetos cujo valor do perímetro em *pixels* estivessem dentro do intervalo. Com isso, as imagens resultantes podem ser vistas na Figura 1.

Figura 1-Mão(esquerda) e pestana(direita) pós-processadas.



Fonte: elaborada pelo autor (2019).

Na etapa de extração de características, foram utilizadas 11 características pertinentes.

As 8 primeiras são referentes à quantidade de *pixels* que estão em uma determinada direção (norte, sul, leste, oeste, nordeste, noroeste, sudeste e sudoeste) em relação ao pixel central.

As últimas 3 características estão relacionadas à razão entre a altura da mão e sua largura, a dimensão da maior convexidade da figura da mão e a distância entre a mão e a pestana do violão.

Todas as características foram normalizadas, já que deve se levar em conta que a distância na qual a foto foi tirada pode influenciar nos valores absolutos das características. E foram escritas em um arquivo ARFF, arquivo de texto para o *software* Weka realizar a leitura e posterior análise dessas características, assim como mostra a Figura 2.

Figura 2-Arquivo ARFF.

```
@relation characteristics

@attribute cima numeric
@attribute baixo numeric
@attribute esquerda numeric
@attribute direita numeric
@attribute DES numeric
@attribute DEI numeric
@attribute DDS numeric
@attribute DDI numeric
@attribute razao numeric
@attribute convex numeric
@attribute casa numeric
@attribute class {D0, RE, MI, FA, SOL, LA, SI}

@data
0.07973856209150326,0.08496732026143791,0.10588235294117647,0.1111111111111111,0.1594771
0.16265640038498555,0.15976900866217517,0.1693936477382098,0.17131857555341676,0.2839268
0.2281124497991968,0.22971887550200804,0.242570281124498,0.23775100401606425,0.4,0.36385
0.4371029224904701,0.45108005082592123,0.4815756035578145,0.4866581956797967,0.781448538
0.4783118405627198,0.5064478311840562,0.5463071512309496,0.5650644783118406,0.8663540445
0.6048284625158831,0.6315120711562897,0.6950444726810674,0.7306226175349428,1.0991105463
0.578076525336091,0.5987590486039297,0.6752843846949328,0.6938986556359876,1.05997931747
0.7270642201834863,0.75,0.8589449541284404,0.8853211009174312,1.3360091743119267,1.21559
0.9188829787234043,0.9441489361702128,1.1023936170212767,1.1382978723404256,1.7114361702
0.9026128266033254,0.9323040380047506,1.0819477434679334,1.1258907363420427,1.6947743467
0.8514456630109571,0.8753738783649053,1.0099700897308075,1.0488534396809572,1.5922233300
1.3075842696629214,1.327247191011236,1.5140449438202248,1.5646067415730338,2.40168539325
1.6854700854700855,1.6974358974358974,1.9504273504273504,2.005128205128205,3.08888888888
1.1943820224719102,1.1898876404494383,1.3943820224719101,1.4258426966292135,2.2,1.974157
1.118561710398445,1.1117589893100097,1.3206997084548104,1.3440233236151604,2.06219630709
```

Fonte: elaborada pelo autor (2019).

Por fim, para a etapa de classificação foi utilizado o *software* Weka, que já contém implementado alguns métodos de aprendizado de máquina, foram testados 42 classificadores, sendo que foram selecionados para análise de resultados os que obtiveram as 4 melhores acurácias, sendo 1 deles, utilizando de meta-aprendizagem.

O *dataset* foi dividido de duas formas diferentes para treinamento e testes, sendo eles: *k-fold* com o valor de $k=10$ e *split* com 66% selecionados para treinamento e 33% para testes.

Como uma tentativa de melhorar os resultados, foi também utilizado o método de votação, utilizando dos três classificadores e, a partir disso, realizando a classificação dos dados através do classificador que tivesse a maior taxa de certeza na sua classificação.

RESULTADOS E DISCUSSÕES

Nesta seção serão discutidos os resultados obtidos na parte de metodologia e, também, discutir sobre estes resultados.

RESULTADOS

Os resultados podem ser vistos na Tabela 1 e Tabela 2.

Tabela 1-Métricas de avaliação dos classificadores utilizando kfold (k= 10).

Classificador	Acurácia(%)	Precisão	Recall	F-Measure	ROC
Classificador Multiclasse	95.961	0.960	0.960	0.960	0.998
LMT	87.307	0.875	0.873	0.873	0.958
Logística Simples	85.961	0.863	0.860	0.859	0.974
Rede Neural Multicamada	85.961	0.863	0.860	0.860	0.966
Votação	95.192	0.953	0.952	0.952	0.993

Fonte: elaborada pelo autor (2019).

Tabela 2-Métricas de avaliação de classificadores utilizando Split (66% para treinamento).

Classificador	Acurácia(%)	Precisão	Recall	F-Measure	ROC
Classificador Multiclasse	92.655	0.932	0.927	0.927	0.993
LMT	86.440	0.871	0.864	0.864	0.968
Logística Simples	85.310	0.858	0.853	0.851	0.962
Rede Neural Multicamada	83.615	0.861	0.836	0.838	0.958
Votação	94.350	0.947	0.944	0.944	0.985

Fonte: elaborada pelo autor (2019).

DISCUSSÕES

Pode-se observar uma melhora de acurácia quando se utiliza a divisão do *dataset* através do método *k-fold*, isso pode ser explicado pelo fato de que este métodos consiste em alternar a parcela dos dados utilizados no treinamento com os utilizados nos teste, fazendo com que todos os dados sejam usados para o treinamento, o que ajuda na melhora da acurácia.

A votação se mostrou mais eficaz quando utilizado com o método *split*, uma vez que apresentou uma acurácia geral menor do que o classificador multiclasse quando a divisão do *dataset* foi através do método *k-fold*.

Em ambos os casos, o classificador multiclasse se mostrou, entre os outros 3 classificadores, o que possui maior acurácia, isso se deve pelo fato de ser um modelo de meta-aprendizagem que utiliza da especialização de um classificador

binário em um único rótulo (RIVOLLI e CARVALHO, 2015), ou seja, para este caso, usando 8 funções logísticas, uma especializada em cada rótulo do *dataset*.

CONCLUSÃO

Este trabalho teve como finalidade a aplicação e análise de métodos de aprendizado de máquina para classificação de acordes de violão.

Foram realizadas as etapas referentes ao processamento de imagens, sendo elas: pré-processamento, segmentação, pós-processamento, extração de características e classificação. E a partir disso mostrando a melhoria na acurácia quando utilizados com métodos de meta-aprendizagem.

Pôde-se perceber, a partir do desenvolvimento deste trabalho, a importância do estudo no aprendizado de máquina não somente analisando cada método individualmente, mas também, a aplicação a partir das relações entre diferentes métodos.

Como proposta futura, uma sugestão seria obter diferentes descritores para utilização na etapa de classificação. Além de focar em técnicas de segmentação de pele para utilização na área médica. Por fim, avaliar outros métodos para que o processo possa ser realizado, também, em tempo real.

AGRADECIMENTOS

Eu gostaria de agradecer, inicialmente, à minha família por proporcionar todas as oportunidades que, sem elas, não me permitiriam chegar onde estou hoje. E agradecer, também, ao meu orientador Claiton de Oliveira por sempre me auxiliar em tudo que foi necessário ao longo deste trabalho.

REFERÊNCIAS

ALBUQUERQUE, M. P. D.; ALBUQUERQUE, M. P. D. **Processamento de Imagens: Método e Análises**. Centro Brasileiro de Pesquisas Físicas. Rio de Janeiro, p. 12. 2000.

ARAÚJO, C. G. V.; RIBEIRO, J. A. P. **Conversão e Manipulação em Espaços de Côr**. Faculdade de Engenharia da Universidade do Porto. Porto, p. 9. 2007.

FRANK, E.; HALL, M. A.; WITTEN, I. H. **Data Mining: Practical Machine Learning Tools and Techniques**. 4ª. ed. Wakaito: Morgan Kaufmann, 2016.

HERGET, S. O que é inteligência artificial? História, definições e aplicações. **AndroidPIT**, 2017. Disponível em: <<https://www.androidpit.com.br/o-que-e-inteligencia-artificial>>. Acesso em: 10 Julho 2019.

RIVOLLI, A.; CARVALHO, A. C. P. L. F. D. **O uso seletivo de classificadores binários na solução de problemas multirrótulos**. Cornélio Procópio. 2015.