

# Reconhecimento de traços da notação musical baseada em caneta

Hevelyn Sthefany Lima de Carvalho  
Departamento de Ciência da Computação  
Universidade de Brasília  
Brasília, Distrito Federal, Brazil  
Email: hev.sthefany@aluno.unb.br

**Abstract**—A notação musical é um sistema complexo e exige abordagens específicas para o reconhecimento de símbolos musicais, levando à dificuldade de obter precisões precisas. Tal campo de pesquisas ainda é um desafio e não se encontram muitas pesquisas relacionadas, especialmente em relação ao reconhecimento online de símbolos musicais escritos à caneta. Motivado ao desafio e levando em consideração que um símbolo pode ser representado como uma combinação de vários traços básicos, esse trabalho apresenta um experimento de reconhecimento de traços de símbolos musicais: uma parte do processo de reconhecimento do símbolo como um todo. O experimento contribui para a construção de uma base de conhecimento do autor no campo de pesquisa e dessa forma, poder reconhecer desafios a serem solucionados e definir propostas de novas abordagens para essa tarefa. O método é avaliado usando dois conjuntos de dados, incluindo HOMUS, um dos maiores conjuntos de dados de símbolos musicais, e classes de traços encontrados nesse mesmo conjunto, construído por [1].

## I. INTRODUÇÃO

Normalmente os musicistas usam software de composição e edição musical por meio de teclado, mouse ou um instrumento conectado ao computador. Apesar da complexidade e robustez que tais softwares ganharam com o tempo, muitos compositores ainda preferem expressar suas novas composições musicais usando caneta e papel pela simplicidade e praticidade. Se enquadrar a um menu limitado, estilos predefinidos e ações de apontar e clicar pode ser tedioso para os artistas. Consequentemente, percebeu-se a necessidade de reconhecimento automático de símbolos musicais escritos à mão. Surgiu então a área de pesquisa de Reconhecimento Musical Óptico (OMR) com a tarefa de converter páginas de música digitalizadas em um formato legível por máquina.

A análise automática de imagens de notação musical não é uma tarefa nova, como mostram Blostein e Baird em seu artigo [2], apresentando uma visão geral e análise crítica dos problemas e soluções propostas no processamento de imagens musicais no período de 1966 a 1990. Mas até hoje, OMR se mostra como uma área desafiadora, sendo estudada por muitos pesquisadores, como Malakar et al. [3] que usou o descritor de recurso baseado em textura Daisy para extrair as características das imagens de partituras musicais. No entanto, o desempenho dos sistemas OMR está longe de ser o ideal, especialmente no caso da notação manuscrita [4].

O reconhecimento de símbolos musicais, com processamento de imagens, é normalmente definido como um método

offline pela comunidade científica. No entanto, também temos o método online, se referindo ao reconhecimento de símbolos musicais escritos a caneta. Nesse método, os símbolos musicais são desenhados em uma superfície digital através de uma caneta eletrônica, e são processados pelo sistema para reconhecimento da notação e apresentação do símbolo correspondente em formato digital. Podemos dizer que esse processo é como uma digitalização contínua da escrita para um arquivo de música estruturado e legível para o computador. Essa tarefa é muito semelhante à tarefa Reconhecimento Óptico de Caracteres (OCR), para a qual a abordagem de processamento de caracteres baseados em caneta foi amplamente proposta [5], [6], [6], [7]. No entanto, a notação musical é um sistema complexo e exige abordagens específicas para seu reconhecimento, levando à dificuldade de obter precisões precisas. Consequentemente, existem poucas interfaces baseadas em caneta para composição e edição de documentos digitais e, em particular, para notações musicais. Alguns estudos anteriores foram realizados, mas este campo ainda permanece pouco explorado. Por tais razões, fui motivada a conhecer e entender melhor esse campo, para então conseguir contribuir com novas abordagens no futuro.

Portanto, esse trabalho apresenta um experimento de reconhecimento de traços de símbolos musicais, uma parte do processo de reconhecimento do símbolo como um todo. Ou seja, o trabalho contribui para construir uma base de conhecimento do autor no campo de pesquisa e dessa forma, poder reconhecer desafios a serem solucionados e definir propostas de novas abordagens para a OMR online.

O documento está organizado da seguinte forma. Seção II apresenta alguns trabalhos relacionados ao reconhecimento online de símbolos musicais. Seção III detalha o método proposto e suas etapas de constituição. Seção IV descreve os experimentos para validar o método proposto e discute os resultados. E por fim, Seção V conclui o trabalho e apresenta propostas para trabalhos futuros.

## II. REVISÃO DE LITERATURA

Miyao et al. [8] lida com os traços individuais de um símbolo musical em vez do próprio símbolo, sem se importar com a ordem de escrita. Em seguida, eles são combinados para formar um símbolo. Os autores usaram Código em Cadeia de Freeman de 8 direções e características da imagem do traço.

Com diferentes cenários de teste, o método rendeu taxas de classificação acima de 98%.

Pela falta de conjuntos de dados, impossibilitando a comparação de técnicas, Calvo-Zaragoza et al. [9] construiu HOMUS<sup>1</sup>, atualmente o maior e mais conhecido conjunto de dados com notações musicais, em especial, a caneta. HOMUS foi construído por 100 músicos, constituindo em 15200 arquivos de texto contendo rótulo da amostra e uma ou mais linhas, cada uma com um único traço contendo um conjunto de pontos 2D. O artigo também apresenta algumas técnicas de classificação tanto para modo online e offline (imagens) sem pretensão de boas precisões, apresentando apenas possibilidades do que pode ser feito. Para o modo online, eles restringiram ao uso da técnica do Vizinho Mais Próximo (NN), adotando como medida de similaridade a distância entre vetores com Códigos de Freeman (FFC) e Time warping dinâmico, e dos Modelos de Markov Escondidos (HMM).

No cenário onde o estilo do usuário não era conhecido, ou seja, o método aprende com amostras de muitos músicos diferentes (experimento independente do usuário), todas as taxas de erro foram superiores a 15%. Já no cenário onde o classificador sabe o estilo do usuário (experimento dependente do usuário), FCC relatou a melhor taxa de erro, em média 7%. Além disso, viu-se que os algoritmos que trabalham com os dados online, exceto o HMM, obtêm resultados significativamente melhores do que os outros.

Mais tarde, os mesmo autores propoaram uma abordagem para resolver os problemas envolvidos na tarefa de reconhecimento online de música [10], utilizando máquinas de estados finitos e medidas de dissimilaridade entre 55 tipos de traços, das quais um símbolo musical é composto. Os traços recebidos como entrada são mapeados na probabilidade de representar cada uma das primitivas consideradas. A abordagem também usa um modelo semântico que descreve quais sequências musicais são formalmente aceitáveis.

O reconhecimento é adaptável a qualquer tipo de escrita à mão desde que sejam fornecidos dados de treinamento. Ou seja, o usuário precisa passar por uma fase de treinamento antes de usar a ferramenta. No experimento dependente do usuário, 90% das sequências são perfeitamente reconhecidas. Já no experimento independente do usuário, apenas 64% das sequências foram perfeitamente reconhecidas.

No mesmo ano, Jiyong Oh et al. [1] propõe uma abordagem composta por duas partes: reconhecimento dos traços que compõe o símbolo e reconhecimento do símbolo baseada na classificação de seus traços. Tal método tem uma estrutura semelhante a [8], mas os usuários do método proposto não precisam estar cientes de nenhuma regra de escrita. São extraídos três tipos de recursos dos traços: a informação de tamanho (SI), o histograma de ângulos de movimento direcional (HDA) e o histograma de ângulos de movimento não direcionados (HUA) para lidar com os casos em que os recursos HDA não são discriminantes o suficiente. O vetor representante de um

símbolo usa, além da classificação de seus traços, a localização central de cada traço bem como sua largura e altura.

Para o experimento, os autores consideraram o conjunto HOMUS e SNU<sup>2</sup>. Para possibilitar a realização do método, os pesquisadores analisaram 31.768 traços do conjunto de dados HOMUS em todas as amostras e escolheram vinte e três traços básicos. Ou seja, criaram um conjunto de dados que estende HOMUS, contendo rótulos de vinte e três traços básicos. As taxas máximas de reconhecimento foram de 92,40% na ordem de escrita (Tempo) e 93,65% na ordem lexicográfica (Lexi) das amostras de símbolos para o conjunto de dados HOMUS e SNU, respectivamente. Além disso, apesar do tempo de execução ser maior que em [8], os autores consideraram aceitável devido a melhoria da precisão de mais de 10%. Este presente trabalho tem como base a primeira parte do método deste artigo.

### III. METODOLOGIA

Um símbolo pode ser composto de um ou mais traços, composto por um conjunto de pontos 2D, referentes às posições pelas quais a caneta percorreu em uma certa taxa de amostragem. O vetor de características de cada traço é composto pela informação de tamanho (*size information* - SI), pelo histograma de ângulos de movimento direcional (*histogram of directional movement angles* - HDA) e pelo histograma de ângulos de movimento não direcionados (*histogram of undirected movement angles* - HUA), conforme [1].

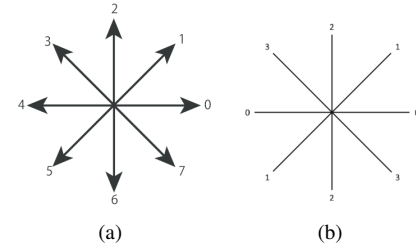


Fig. 1. **a** atribuição de códigos a ângulos de movimento direcional para recursos HDA. **b** atribuição de códigos a ângulos de movimento não direcional para recursos HUA.

#### A. Informação de tamanho (SI)

As informações de tamanho constitui três recursos:

$$w_s = \max_i(x_i) - \min_i(x_i), \quad (1a)$$

$$h_s = \max_i(y_i) - \min_i(y_i), \quad (1b)$$

$$l_s = \sum_{i=1}^{S-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2}, \quad (1c)$$

onde  $w_s$  e  $h_s$  são a largura e a altura da caixa ao redor do traço, respectivamente, e  $l_s$  corresponde ao comprimento do traço. Partindo do pressuposto de que as linhas da pauta são fornecidas, os três valores são divididos pela altura da pauta, o

<sup>1</sup><https://grfia.dlsi.ua.es/homus/>

<sup>2</sup>[http://mipal.snu.ac.kr/index.php/Labels\\_of\\_Strokes\\_in\\_HOMUS\\_Dataset](http://mipal.snu.ac.kr/index.php/Labels_of_Strokes_in_HOMUS_Dataset)

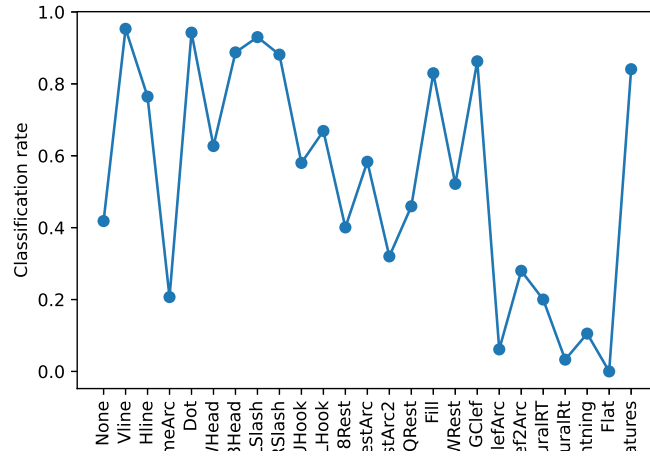


Fig. 2. Taxas de classificação por classe.

que significa a distância vertical entre as linhas da pauta mais alta e mais baixa, porque a maioria das pessoas geralmente escreve símbolos musicais dependendo do tamanho da pauta.

#### B. Histograma de ângulos de movimento direcional (HDA)

Ao invés de armazenar as coordenadas absolutas dos pixels, pode-se usar a posição relativa entre pontos consecutivos de um traço. Para isso, faz-se uso do código de cadeia de Freeman, o qual atribui um valor inteiro variando de 0 a 7 à direção formada entre cada ponto e seu vizinho, como na Figura 1(a). No trabalho, após computar o código, um histograma de cada ponto é computado, resultado num vetor de oito posições.

#### C. Histograma de ângulos de movimento não direcionados (HUA)

Alguns símbolos apresentam o mesmo resultado quando escritos no sentido horário ou anti-horário. Para o método tratar desse caso, são calculados recursos HUA. Tais recursos são calculados como os de HDA, com a diferença de considerar direções opostas equivalentes, como na Figura 1(b).

#### D. Treinamento do classificador

Os 3 tipos de recursos são concatenados num só vetor de recursos de tamanho 19. Tais recursos são dados como entrada para um classificador Máquinas de vetor de suporte baseada em libsvm. Os hiperparâmetros do classificador foram definidos de acordo com os resultados fornecidos pela abordagem de otimização *RandomizedSearchCV*<sup>3</sup>.

### IV. RESULTADOS EXPERIMENTAIS

Nesta seção, realizamos experimentos e avaliamos a taxa de classificação e tempo de computação. O método proposto

foi codificado em Python 3.8. Além disso, foram classificados traços dos símbolos do conjunto de dados HOMUS, considerando as classes definidas por [1].

A Figura apresenta as taxas de classificação para cada classe de traços, com média de precisão de 70%. Apesar da classificação dos traços em [1] ser melhor, o vetor de recursos possui tamanho 195, sobrecarregando o processamento. Levando em consideração a extração de recursos e a segunda parte do reconhecimento dos símbolos, o tempo de classificação numa situação real poderia ser incômodo. Essa observação também é apresentada pelos autores.

Esse problema pode ser selecionado aplicando um método de classificação mais robusto e rápido reconhecimento após o treinamento, sem precisar de um extrator de recursos com muitas etapas. Além disso, seria interessante considerar a ordem de escrita dos traços, e a representatividade de cada um para o todo.

A tabela I exhibe o tempo médio de computação, que foram maiores que em [1]. No entanto são tempos aceitáveis levando em consideração o fato de que o código foi executado em um notebok Jupyter.

Extração de características	Classificação
1.0	2.1

TABLE I  
TEMPO MÉDIO DE COMPUTAÇÃO (EM MILISSEGUNDOS)

### V. CONCLUSÃO

Esse trabalho realiza um experimento de reconhecimento de traços baseado no método proposto em [1]. Viu-se que o processamento dos dados e tamanho do vetor de recursos pode sobrecarregar a computação e aumentar o tempo de classificação final.

Proponho como pesquisa futura a classificação dos símbolos musicais com *Transformes encoder only*, um modelo de aprendizado profundo que usa o mecanismo **attention** e que ganhou

<sup>3</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.RandomizedSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html)

muita notoriedade. Atualmente, usada geralmente para tarefas de processamento de linguagem natural (NLP), rendendo altas precisões, a técnica pode classificar os traços de um símbolo de forma análoga às palavras de uma frase. Além disso, o modelo Transformers irá considerar os traços escritos antes e depois e suas importâncias ou contribuições na classificação de um traço em questão.

#### REFERENCES

- [1] J. Oh, S. J. Son, S. Lee, J.-W. Kwon, and N. Kwak, "Online recognition of handwritten music symbols," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 20, no. 2, pp. 79–89, 2017.
- [2] D. Blostein and H. S. Baird, "A critical survey of music image analysis," in *Structured document image analysis*. Springer, 1992, pp. 405–434.
- [3] S. Malakar, M. Ghosh, A. Chatterjee, S. Bhowmik, and R. Sarkar, "Offline music symbol recognition using daisy feature and quantum grey wolf optimization based feature selection," *Multimedia Tools and Applications*, vol. 79, no. 43, pp. 32 011–32 036, 2020.
- [4] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. Marcal, C. Guedes, and J. S. Cardoso, "Optical music recognition: state-of-the-art and open issues," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 3, pp. 173–190, 2012.
- [5] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 63–84, 2000.
- [6] H. Swethalakshmi, A. Jayaraman, V. S. Chakravarthy, and C. C. Sekhar, "Online handwritten character recognition of devanagari and telugu characters using support vector machines," in *Tenth international workshop on Frontiers in handwriting recognition*. Suvisoft, 2006.
- [7] X.-Y. Zhang, Y. Bengio, and C.-L. Liu, "Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark," *Pattern Recognition*, vol. 61, pp. 348–360, 2017.
- [8] H. Miyao and M. Maruyama, "An online handwritten music symbol recognition system," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, no. 1, pp. 49–58, 2007.
- [9] J. Calvo-Zaragoza and J. Oncina, "Recognition of pen-based music notation: The homus dataset," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 3038–3043.
- [10] —, "Recognition of pen-based music notation with finite-state machines," *Expert Systems with Applications*, vol. 72, pp. 395–406, 2017.