

Remoção de compasso usando combinações de operadores heurísticos e treinados automaticamente

Rafael Reis da Silva, Igor S. Montagner, Nina S. T. Hirata
University of São Paulo,
São Paulo, Brazil

rafael.reis.silva@usp.br, [igordsm,nina]@ime.usp.br

Resumo—A remoção de compasso é um pré-processamento comumente utilizado em sistemas de reconhecimento automático de partituras. Diferentes imperfeições e deformações dificultam o tratamento destas imagens. Diversos métodos heurísticos para este fim podem ser encontrados na literatura, mas cada um funciona para conjuntos diferentes de deformações. Neste trabalho propomos uma técnica híbrida que utiliza aprendizagem de máquina para combinar métodos heurísticos de remoção de compasso, permitindo criar um sistema que pode escolher automaticamente quais métodos usar dependendo das imagens de entrada disponíveis. Este tipo de combinação obtém resultados bons em diversas deformações, frequentemente superando cada um dos métodos heurísticos.

Keywords—Remoção de compasso, Aprendizado de W-operadores, Processamento de imagens, Aprendizado de Máquina

Resumo—Staff removal is a very common preprocessing in Optical Music Recognition. Different imperfections and deformations present in the images make this task a hard one. Although many methods for staff removal are found in the literature, each one works best in different deformation contexts. We propose a hybrid, parameter free approach that combines the output of many heuristic methods using machine learning. This approach allows us to train operators that can choose which method to use depending on the deformations present in the training images. Our approach achieved good performance in all deformation contexts, frequently surpassing all individual methods.

Keywords—Staff removal, W-operator learning, Image processing, Machine Learning.

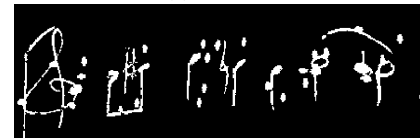
I. INTRODUÇÃO

A remoção de compasso de partituras é um pré-processamento muito comum em sistemas de reconhecimento ótico de músicas (*Optical Music Recognition*) [1]. Ao segmentar a imagem em símbolos e fundo, facilita o reconhecimento dos símbolos musicais. Tipicamente, métodos de remoção de compasso exploram, mesmo que indiretamente, alguma característica intrínseca de imagens de partituras, como por exemplo o fato do compasso ser um conjunto de linhas paralelas horizontais e longas.

Porém, devido a diversas imperfeições nas imagens e em suas condições de digitalização, estas suposições são frequentemente falsas. Desta maneira, o desenvolvimento de métodos robustos é necessário. Dalitz *et al* [1] examinam o desempenho de vários métodos heurísticos utilizando diversas deformações



(a) Entrada



(b) Saída esperada

Figura 1: Par de imagens entrada-saída para ilustrar a remoção de compasso [7]

artificiais e conclui que não existe um método que obtém bom desempenho em todas as situações.

Trabalhos anteriores [2], [3] utilizam Aprendizado de Máquina para estimar operadores baseados no modelo dois níveis [4] de operadores que sejam robustos às condições apresentadas em [1]. Nessa abordagem, a tarefa de processamento de imagens é representada por pares de imagens entrada-saída como os da Figura 1. Os operadores estimados são chamados de W -operadores [5] e consistem, essencialmente, em um classificador de pixels que determina a cor de cada pixel da imagem de saída baseado em uma janela ao seu redor na imagem de entrada. O modelo dois níveis combina o resultado de diversos W -operadores treinados com janelas diferentes, obtendo um operador com desempenho superior. Como mostrado em [3] e [6], o modelo dois níveis exige determinação cuidadosa de seus parâmetros para otimizar seu desempenho.

Neste trabalho exploramos uma abordagem híbrida que combina operadores heurísticos usando o modelo dois níveis. Como mostrado em [1], cada método heurístico obtém melhores resultados em determinadas deformações. Logo, podemos utilizar o modelo dois níveis para combinar os resultados dos métodos heurísticos de modo a obter um operador que possa se adaptar a quaisquer deformações. A combinação de operadores heurísticos também evita a determinação de parâmetros para o modelo dois níveis.

Experimentos realizados utilizando os dados publicados em [1] mostram que combinações de métodos heurísticos superam o desempenho dos métodos individuais na maioria das deformações testadas e que combinações de métodos heurísticos e treinados automaticamente podem ter, em algumas situações, desempenho ainda melhor.

Este artigo está organizado da seguinte maneira. Na Seção II estão descritos os fundamentos do aprendizado de W -operadores e do modelo de dois níveis. Na Seção III apresentamos a abordagem proposta e discutimos suas vantagens e desvantagens. Na Seção IV descrevemos os experimentos realizados e seus resultados. Finalmente, apresentamos as conclusões do trabalho na Seção V.

II. FUNDAMENTOS DO TREINAMENTO DE W -OPERADORES

Seja $E = \mathbb{Z}^2$ o espaço de definição das imagens. Uma imagem binária pode ser representada como um subconjunto $I \subset E$. Um operador de imagens $\Psi : \{0, 1\}^E \rightarrow \{0, 1\}^E$ é uma função que mapeia imagens em imagens. W -operadores são operadores localmente definidos e invariantes a translação, isto é, existe uma janela $W \subset E$ e uma função $\psi : \{0, 1\}^W \rightarrow \{0, 1\}$ tal que, para todo ponto $x \in E$

$$\Psi(I)[x] = \psi(I_{-x} \cap W),$$

onde I_x é a translação da imagem I por $x \in E$. Desta maneira, Ψ é totalmente determinado por W e ψ .

Podemos estimar ψ a partir de um conjunto de pares de imagens como os da Figura 1 posicionando a janela W em cada ponto i das imagens de treinamento e coletando as co-ocorrências $(x_i, y_i), x_i \in \{0, 1\}^W, y_i \in \{0, 1\}$. Então, treinamos um classificador ψ tomando como vetores de características os vetores x_i e como rótulos os valores y_i . Neste trabalho estimamos ψ usando árvores de decisão, porém trabalhos anteriores [2], [3] utilizaram o algoritmo ISI [8], que minimiza o Erro Absoluto Médio no conjunto de treinamento [9].

A janela usada define quais características serão usadas no classificador ψ . Logo, a definição de uma janela adequada para cada tarefa é essencial para obtermos operadores com bom desempenho. Janelas pequenas não são capazes de diferenciar padrões complexos, resultando em operadores com desempenho ruim. Por outro lado, uma grande quantidade de dados é necessária para treinar operadores com janelas grandes, resultando em operadores mais suscetíveis a erros de generalização.

O modelo dois níveis [4] foi desenvolvido para mitigar este problema. Ao invés de usar somente um operador treinado com uma janela grande, é feita a combinação de diversos operadores treinados com janelas de tamanho moderado. A saída de cada operador é concatenada em um padrão de segundo nível, que é novamente classificado para a obtenção do valor de saída do operador. Uma ilustração deste processo pode ser vista na Figura 2.

O trabalho [2] propõe um conjunto de janelas, mostradas na Figura 3, criadas especificamente para o processamento de imagens de partituras.

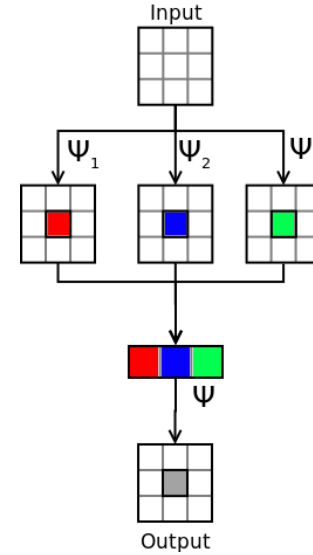


Figura 2: Operador dois níveis: cada pixel na imagem de entrada é processado por três operadores de primeiro nível, ψ_1 , ψ_2 e ψ_3 , e os três valores resultantes são concatenados para formar um novo padrão. Este é então classificado pelo operador de segundo nível ψ , que gera a classificação final.

III. COMBINAÇÃO DE OPERADORES HEURÍSTICOS E TREINADOS AUTOMATICAMENTE

Trabalhos anteriores [4], [2], [10] mostram que os operadores dois níveis descritos acima tem em geral desempenho melhor que os operadores individuais. Uma vez que o treinamento do operador de segundo nível (aquele que realiza a combinação) não utiliza o fato de que os operadores de primeiro nível são W -operadores treinados automaticamente, a ideia central deste trabalho é a utilização de operadores heurísticos no primeiro nível ao invés dos operadores treinados. Desta forma, o modelo de operador dois níveis pode ser utilizado para combinar operadores heurísticos.

Como mostrado em [1], não existe um único método dentre os operadores heurísticos propostos na literatura que, individualmente, seja robusto em qualquer situação na remoção de compasso. Portanto, o procedimento de combinação de operadores heurísticos pode ser especialmente vantajoso, pois ele poderia considerar prioritariamente os melhores operadores heurísticos no conjunto de treinamento de interesse. Isto é, o processo de combinação seria capaz de “escolher” quais métodos heurísticos são melhores para cada mistura de deformações em particular.

A. Operadores Heurísticos

Foram escolhidos 6 métodos heurísticos para serem combinados: Skeleton, Line Tracking Height(Lth), Line Tracking Chord(Ltc) [1], Roach and Tatem [11], Carter [12] uma versão simplificada de [13] (RL Simple). Estes algoritmos foram escolhidos por terem características diferentes entre si e, portanto, terem desempenhos distintos em relação a diferentes tipos de deformações.

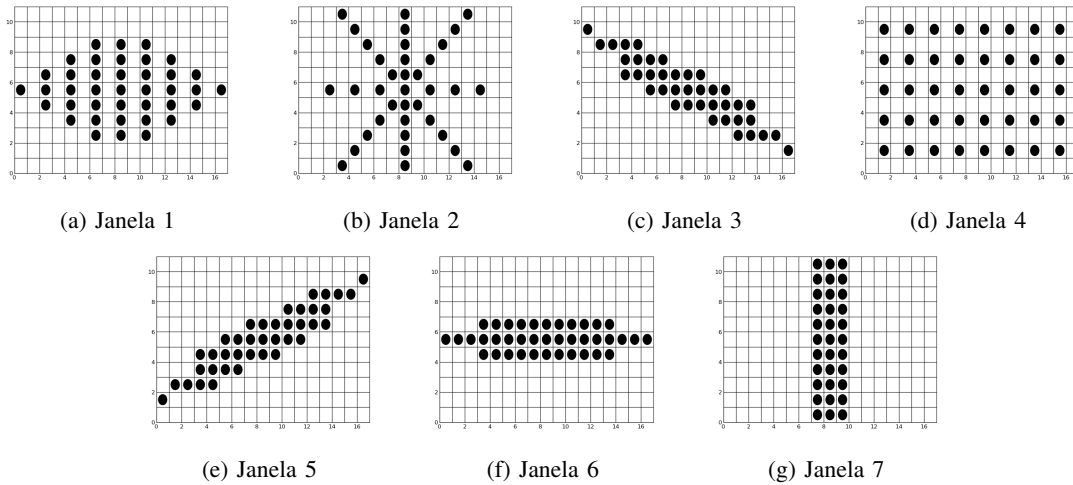


Figura 3: Janelas usadas nos experimentos.

IV. EXPERIMENTOS

As imagens, assim como os códigos para simular as deformações, usados nos experimentos foram propostos por Dalitz *et al* [1] e encontram-se em <http://gamera.informatik.hsnr.de/addons/musicstaves/>. Usamos o mesmo conjunto de deformações sobre 32 imagens sintéticas, que foi usado em [2]. Exemplos do efeito de cada uma das deformações consideradas podem ser vistos na Figura 4. Dividimos o conjunto de imagens em duas partes, a primeira, contendo 16 imagens, para o treinamento dos operadores e o restante para teste. Das 16 imagens para treinamento, as 12 primeiras foram usadas no primeiro nível, enquanto as 4 últimas foram usadas no treinamento dos operadores de segundo nível.

Apresentamos na Tabela I o resultado da avaliação individual de cada um dos métodos utilizados. A primeira coluna indica o tipo de deformação, as colunas de 2 a 7 indicam cada um dos seis métodos heurísticos, e as últimas 7 colunas indicam os operadores treinados individualmente para cada janela da Figura 3. Os valores correspondem à taxa de erro (pixels classificados incorretamente), calculados sobre as imagens de teste. Consistentemente ao reportado em [1], o método Skeleton foi o melhor em 5 de 10 deformações. Os operadores treinados obtiveram o melhor desempenho nas 5 deformações restantes, sendo que a janela 4 obteve o melhor desempenho em 3 testes. Claramente, não existe um método que é mais robusto em todas as situações.

Os resultados da combinação de operadores são apresentados na Tabela II. As colunas de 2 a 5 mostram o desempenho respectivamente da combinação dos métodos heurísticos, do melhor método heurístico, da combinação dos operadores treinados, e do melhor operador treinado individual. Para os operadores treinados, o modelo dois níveis melhora os resultados em relação ao melhor operador individual em 8 testes, sendo que nos 2 testes restantes o desempenho é apenas ligeiramente inferior. Já para os operadores heurísticos, existe melhora em 5 casos, em um caso o desempenho é o mesmo e em 4 casos houve piora. Porém, a combinação obtém

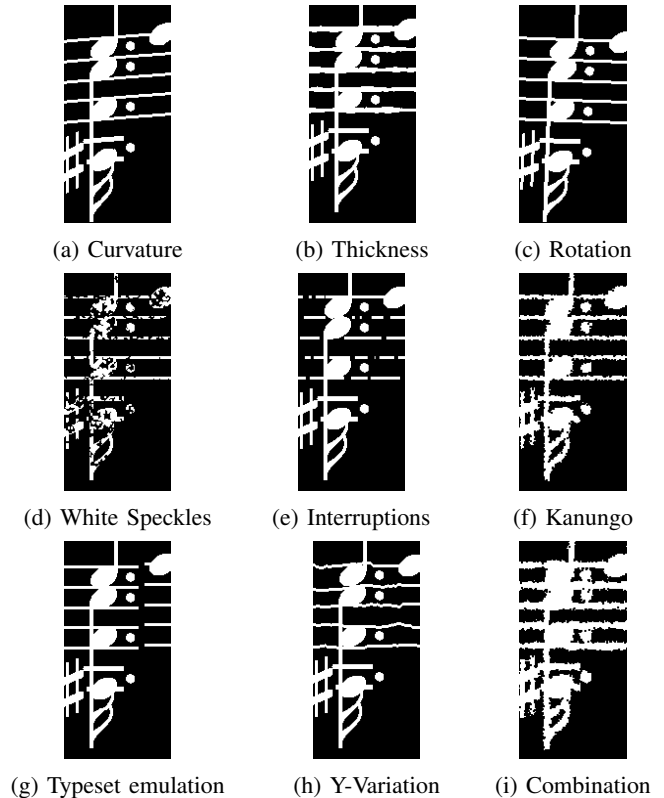


Figura 4: Deformações usadas nos experimentos.

desempenho melhor que o segundo melhor método heurístico em 9 dos 10 casos. Isto mostra que a combinação, apesar de não ser sempre superior ao melhor método individual, pode se adaptar a diferentes deformações nas imagens.

V. CONCLUSÃO

Existem diversos métodos heurísticos para a remoção do compasso de partituras musicais, mas nenhum deles produz resultados satisfatórios em todos os contextos de deformação.

	Skeleton	RI Simple	RI Roach	RI Carter	Ltc	Lth	W1	W2	W3	W4	W5	W6	W7
Undeformed	0.016	0.046	0.071	0.020	0.061	0.049	0.130	0.130	0.137	0.030	0.133	0.120	0.132
Curvature	0.046	0.216	0.239	0.172	0.250	0.242	0.150	0.140	0.114	0.082	0.110	0.162	0.135
Kanungo	0.123	0.118	0.125	0.129	0.110	0.107	0.121	0.115	0.162	0.104	0.154	0.196	0.138
Thickness	0.046	0.082	0.096	0.060	0.085	0.070	0.081	0.077	0.115	0.073	0.109	0.139	0.086
Interruptions	0.018	0.087	0.070	0.055	0.059	0.048	0.126	0.131	0.136	0.035	0.134	0.151	0.131
Mixed-Deform	0.273	0.242	0.356	0.470	0.268	0.262	0.175	0.158	0.222	0.157	0.214	0.222	0.171
Rotation	0.017	0.057	0.088	0.045	0.091	0.079	0.162	0.152	0.154	0.068	0.092	0.156	0.141
Speckles	0.122	0.353	0.258	0.414	0.204	0.197	0.174	0.154	0.180	0.087	0.175	0.200	0.080
Typeset	0.206	0.076	0.097	0.125	0.083	0.072	0.134	0.134	0.141	0.044	0.137	0.153	0.132
Y-Variation	0.020	0.096	0.103	0.063	0.098	0.086	0.015	0.018	0.050	0.046	0.046	0.169	0.017

Tabela I: Desempenho dos métodos heurísticos e dos operadores treinados com as janelas de referência para cada deformação testada.

	Heurísticos Combinados	Melhor heurístico Individualmente	Treinados em dois níveis	Melhor janela Individualmente
Undeformed	0.039	0.016	0.024	0.030
Curvature	0.030	0.046	0.088	0.082
Kanungo	0.063	0.107	0.085	0.104
Thickness	0.066	0.046	0.057	0.073
Interruptions	0.018	0.018	0.022	0.035
Mixed-Deform	0.177	0.242	0.133	0.157
Rotation	0.038	0.017	0.070	0.068
Speckles	0.163	0.122	0.055	0.080
Typeset	0.025	0.072	0.025	0.044
Y-Variation	0.017	0.020	0.012	0.015

Tabela II: Comparação dos erros das combinações de operadores heurísticos e treinados automaticamente.

Neste trabalho mostramos que o modelo dois níveis, para treinamento de operadores de imagens, pode ser aplicado na remoção de compasso. Ele permite combinar tanto operadores heurísticos quanto operadores treinados automaticamente. Combinações de métodos heurísticos geraram resultados melhores que a média dos operadores individuais, frequentemente sendo melhores que o melhor método individual. A combinação de operadores heurísticos pôde, portanto, se adaptar a diferentes deformações nas imagens, superando as fraquezas dos métodos heurísticos quando esses são considerados isoladamente.

AGRADECIMENTOS

Este trabalho foi apoiado pela FAPESP (2015/17741-9, 2011/50761-2, 2015/01587-0) e CNPq (484572/2013-0). Rafael R. Silva e Nina S. T. Hirata São apoiados pelo CNPq. Igor S. Montagner é apoiado pela FAPESP (2014/21692-0, 2011/23310-0).

REFERÊNCIAS

- [1] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga, "A comparative study of staff removal algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 753–766, 2008.
- [2] I. Montagner, R. Hirata Jr., and N. S. T. Hirata, "Learning to remove staff lines from music score images," in *IEEE International Conference on Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 2614 – 2618.
- [3] —, "A machine learning based method for staff removal," in *22nd International Conference on Pattern Recognition (ICPR)*, 2014, pp. 3162 – 3167.
- [4] N. S. T. Hirata, "Multilevel training of binary morphological operators," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 707–720, April 2009.
- [5] H. J. A. M. Heijmans, *Morphological Image Operators*. Boston: Academic Press, 1994.

- [6] I. S. Montagner, N. S. T. Hirata, R. Hirata Jr., and S. Canu, "NILC: a two level learning algorithm with operator selection," in *To appear at IEEE International Conference on Image Processing 2016 (ICIP 2016)*.
- [7] M. Visaniy, V. Kieu, A. Fornes, and N. Journet, "ICDAR 2013 music scores competition: Staff removal," in *12th International Conference on Document Analysis and Recognition (ICDAR)*, 2013, pp. 1407–1411.
- [8] N. S. T. Hirata, R. Hirata Jr., and J. Barrera, "Basis computation algorithms," in *Mathematical Morphology and its Applications to Signal and Image Processing (Proceedings of the 8th International Symposium on Mathematical Morphology)*, 2007, pp. 15–26.
- [9] J. Barrera, R. Terada, R. J. Hirata, and N. Hirata, "Automatic programming of morphological machines by pac learning," *Annales Societatis Mathematicae Polonae. Series 4: Fundamenta Informaticae*, vol. Vol. 41, Nr 1,2, pp. 229–258, 2000.
- [10] M. Dornelles and N. Hirata, "Selection of windows for w-operator combination from entropy based ranking," in *28th Conference on Graphics, Patterns and Images (SIBGRAPI)*, Aug 2015, pp. 64–71.
- [11] J. W. Roach and J. E. Tatem, "Using domain knowledge in low-level visual processing to interpret handwritten music: An experiment," *Pattern Recogn.*, vol. 21, no. 1, pp. 33–44, Jan. 1988. [Online]. Available: [http://dx.doi.org/10.1016/0031-3203\(88\)90069-6](http://dx.doi.org/10.1016/0031-3203(88)90069-6)
- [12] N. P. Carter and R. A. Bacon, "Automatic recognition of printed music," in *Structured Document Image Analysis*. Springer, 1992, pp. 456–465.
- [13] I. Fujinaga, "Staff detection and removal," *Visual perception of music notation: on-line and off-line recognition*, pp. 1–39, 2004.