Detecção de Movimentos Auxiliares de Clarinetistas Através de Visão Computacional

Rodrigo A . Seger[1*], Alessandro L. Koerich[12], Marcelo M. Wanderley[3]

[1]Departamento de Engenharia Elétrica
Universidade Federal do Paraná
81531-970 Curitiba, PR, Brasil

rodrigo.seger@gmail.com, alessandro.koerich@ufpr.br

[2]Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná
80215-901 Curitiba, PR, Brasil

alekoe@ppgia.pucpr.br

[3]Centre for Interdisciplinary Research in Music Media and Technology
Schulich School of Music
McGill University
H3A 1E3 Montreal, QC, Canada

marcelo.wanderley@mcgill.ca

Abstract. *In this paper we discuss the use of computer vision techniques as an alternative method for the detection of ancillary movements produced by clarinetists during performances. Ancillary gestures are movements spontaneously made by musicians and do not produce sound, but they help in the creation of music. In order to optimize the current analysis procedures, we propose the use of computer vision algorithms to detect movements of clarinet, knees and backs, three characteristic ancillary gestures of these musicians. The techniques were applied to nine videos of clarinetists in a controlled studio and the three proposed gestures were detected successfully.*

Resumo. *Neste artigo discute-se o uso de técnicas de visão computacional como um método alternativo para a detecção de movimentos auxiliares produzidos por clarinetistas durante apresentações musicais. Movimentos auxiliares são gestos produzidos espontaneamente pelos instrumentistas e que não geram som, mas que acompanham a execução musical. Para otimizar os atuais procedimentos de análise propôs-se o uso de algoritmos de visão computacional para detectarem movimentos de clarinete, joelhos e costas, três gestos auxiliares característicos destes músicos e que ocorrem com frequência. As técnicas foram aplicadas em nove vídeos de clarinetistas em ambiente controlado e os três movimentos foram detectados com sucesso.*

1. Introdução

Há cerca de 30 anos estuda-se como o ser humano se interage com outras pessoas e com máquinas. Esta interação advém basicamente da fala e da produção de ges-

*Esta pesquisa contou com suporte parcial da CAPES/Fulbright processo BEX1770/712-9, da Fundação Araucária, processo 203/12 e do CNPq processos 306.703/2010-6 e 472.238/2011-6.

tos. Pesquisas em áreas como engenharia, computação, música e psicologia procuram compreender qual o significado real de cada gesto produzido e como reproduzi-lo ou interpretá-lo [Aggarwal and Cai, 1999] [Gavrila, 1999] [Mitra and Acharya, 2007] [Dahl and Friberg, 2004] [Davidson, 1993] [Jensenius et al., 2010].

Centrado no estudo dos movimentos executados por instrumentistas, estes se dividem em três grandes grupos: os gestos efetivos, os gestos de acompanhamento e os gestos meramente figurativos [Cadoz, 1998]. Em uma analogia com um clarinetista, os gestos efetivos são os movimentos que efetivamente produzem som, como, no caso, os dedos pressionando as teclas do clarinete. Os gestos de acompanhamento, também conhecidos como gestos auxiliares, espontâneos ou não óbvios [Wanderley, 1999], são os movimentos que acompanham o processo de produção musical, como movimentos dos ombros e da cabeça. Por fim, os gestos figurativos são todos aqueles que são percebidos pelo som em si, e não pelos movimentos - normalmente gerados por mudanças na melodia ou da entonação.

Os movimentos auxiliares, justamente por não possuírem significado lógico ou motivo de existência, são o principal foco de estudos atuais [Wanderley, 1999] [Wanderley, 2002]. No universo dos clarinetistas, estes gestos são basicamente o movimento completo cima/baixo do clarinete, movimento circular completo do clarinete, movimento cima/baixo de cabeça, movimento cima/baixo de ombros, movimento de ondulação das costas, movimento de abertura/fechamento de braços (”bater asas”), movimento de inclinação da cintura, movimento de dobra de joelhos, movimento de inclinação e batida de pés, deslocamento de peso (balanço do corpo) para a esquerda ou para a direita [Wanderley et al., 2005]. Dentre esta listagem, alguns movimentos aparecem com mais frequência. É o caso dos gestos de cima/baixo do clarinete, ondulação das costas, dobra dos joelhos e deslocamento de peso. Estes são movimentos com alto grau de distinção entre si e amplamente utilizados em estudos desta natureza [Wanderley, 2002] [Wanderley et al., 2005].

Atualmente, a avaliação destes movimentos é feita através de observações de vídeos gravados em ambientes controlados e da análise de informações obtidas via sensores de movimento presos aos músicos. Estes dados normalmente são de grande quantidade, onerando o processo de análise, ao passo que as observações dos vídeos são entediantes e subjetivas (a detecção de movimentos é extremamente dependente do avaliador) [Wanderley, 2002] [Wanderley et al., 2005] [Teixeira, 2010] [Verfaille et al., 2006].

Baseado nesta problemática que se propôs o uso de técnicas de visão computacional aplicadas a vídeos de clarinetistas para criar uma rotulação para os mesmos, ou seja, indicar - via esforço computacional - instantes nos vídeos onde há a execução dos movimentos auxiliares de cima/baixo do clarinete, ondulação das costas e dobra dos joelhos. Dessa forma, seria possível substituir as observações dos vídeos e poder concentrar os posteriores estudos de dados dos marcadores apenas nos intervalos sugeridos pelo algoritmo, otimizando o processo.

A base de dados de estudo é uma compilação de nove vídeos de quatro clarinetistas executando um trecho de cerca de 1 minuto do Segundo Movimento das Três Peças de Stravinsky para Clarinete Solo, fornecidos pelo *Input Devices and Music Interaction Laboratory* do Setor de Tecnologia Musical da Universidade McGill, Canadá. Os vídeos foram gerados por uma simples câmera fixa, em um estúdio com iluminação constante.

Este artigo é composto inicialmente por uma breve contextualização e os objetivos da pesquisa, já mencionados. Dando prosseguimento, será apresentado o método de trabalho proposto, com os resultados obtidos e suas respectivas avaliações. Por fim, as conclusões obtidas e todas as referências utilizadas.

2. Método Proposto

Para poder analisar a base de vídeos e a partir destes mensurar os três movimentos não óbvios neles contidos, foi criado um método composto por três partes fundamentais: pré-processamento dos vídeos, detecção de movimentos e avaliação dos resultados.

Qualquer detecção de movimentos a partir dos vídeos originais é muito complexa. Por isso, inicialmente fez-se necessário adequar a base para, então, viabilizar a detecção de movimentos. Esta etapa foi denominada pré-processamento e tem como principal objetivo segmentar o músico, ou seja, eliminar a influência do cenário no vídeo e corrigir problemas de iluminação durante as filmagens.

A técnica de visão computacional escolhida para a eliminação do cenário foi a de subtração de fundo [Gonzalez and Woods, 2007]. Este procedimento foi escolhido à iluminação ser constante - não há mudança brusca de quantidade de luz - e o ambiente ser estático, porque o cenário é fixo, apenas o músico se desloca, enquanto que a câmera e o cenário em si permanecem imóveis.

Para a utilização da técnica, faz-se necessário o armazenamento de um quadro de referência que será subtraído dos quadros em análise. Como o intuito é eliminar o cenário, ou seja, tudo além do músico, o quadro do vídeo contendo apenas o ambiente, sem o músico, é a referência ideal. O método aplicado encontra-se resumido na Figura 1.

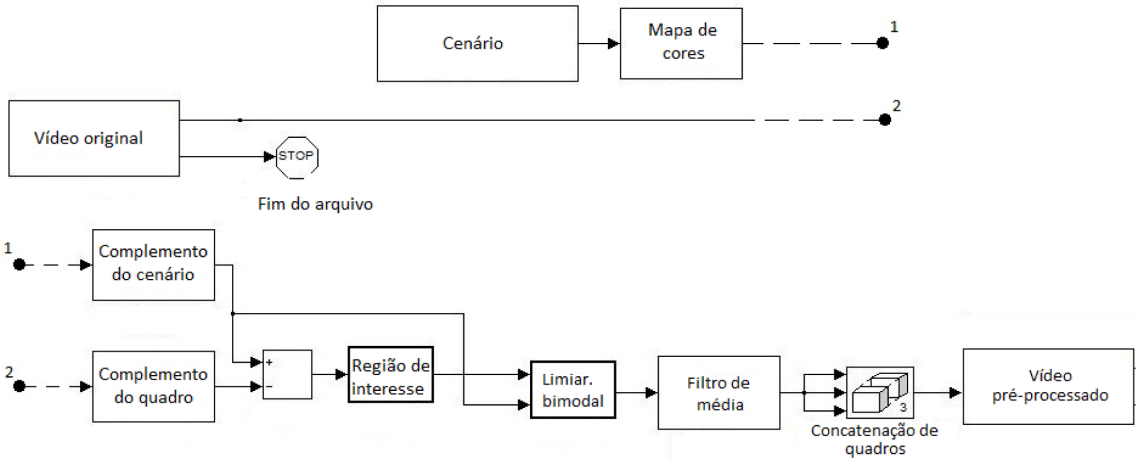

Figura 1: Etapas do pré-processamento.

A Figura 2 ilustra o pré-processamento aplicado. Primeiramente suprimiram-se as informações de cores transformando os vídeos em níveis de cinza com oito bits de profundidade (Figura 2c). Em seguida, fez-se a subtração do quadro-referência de cenário (Figura 2b) de todos os demais quadros dos vídeos (Figura 2d). A próxima etapa foi delimitar a região de interesse (Figura 2e). Como o músico passa quase que a totalidade de duração da gravação na região central, dividiu-se o vídeo em quatro partes iguais (no sentido da largura da imagem), e eliminaram-se as duas extremidades. Realizado o corte, aplicou-se uma limiarização da imagem, ou seja, através de um simples algoritmo condicional, níveis medianos de cinza foram reduzidos para preto e os demais elevados a branco, tornando a imagem binária: o músico e seu clarinete em preto e o cenário inteiramente branco. Um filtro de média foi ainda utilizado no intuito de suavizar os contornos da imagem resultante (Figura 2f). A Figura 3 ilustra o resultado obtido para vídeos dos quatro clarinetistas.

Os três movimentos espontâneos propostos tiveram procedimentos particulares para sua detecção. O movimento auxiliar de cima/baixo de clarinete foi o mais evidente

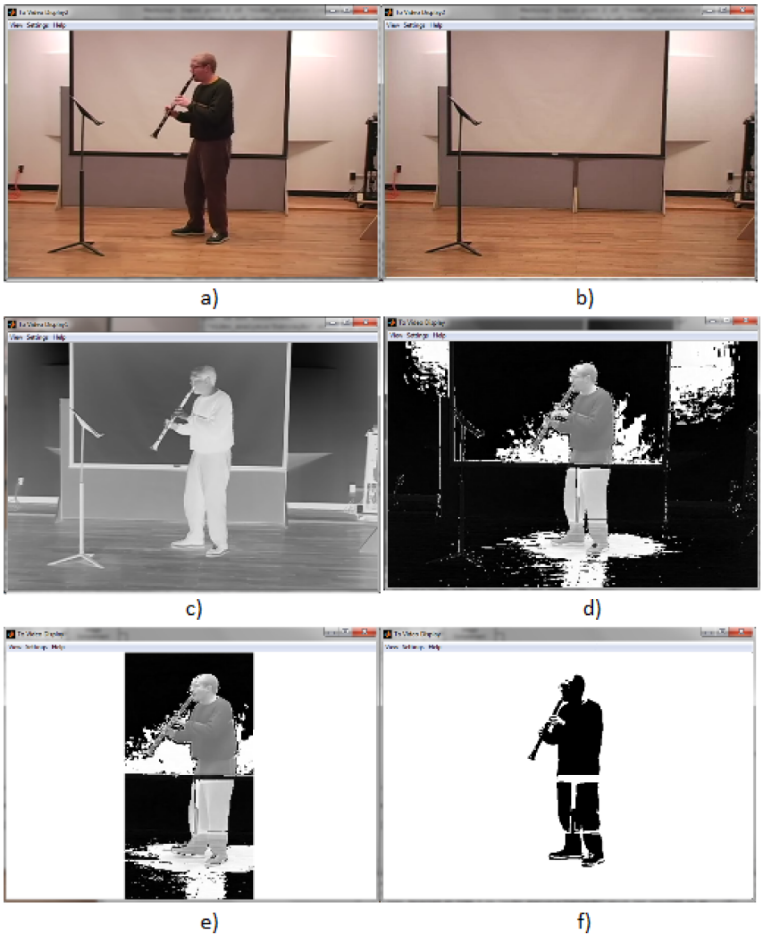

Figura 2: Resultados das etapas do pré-processamento.

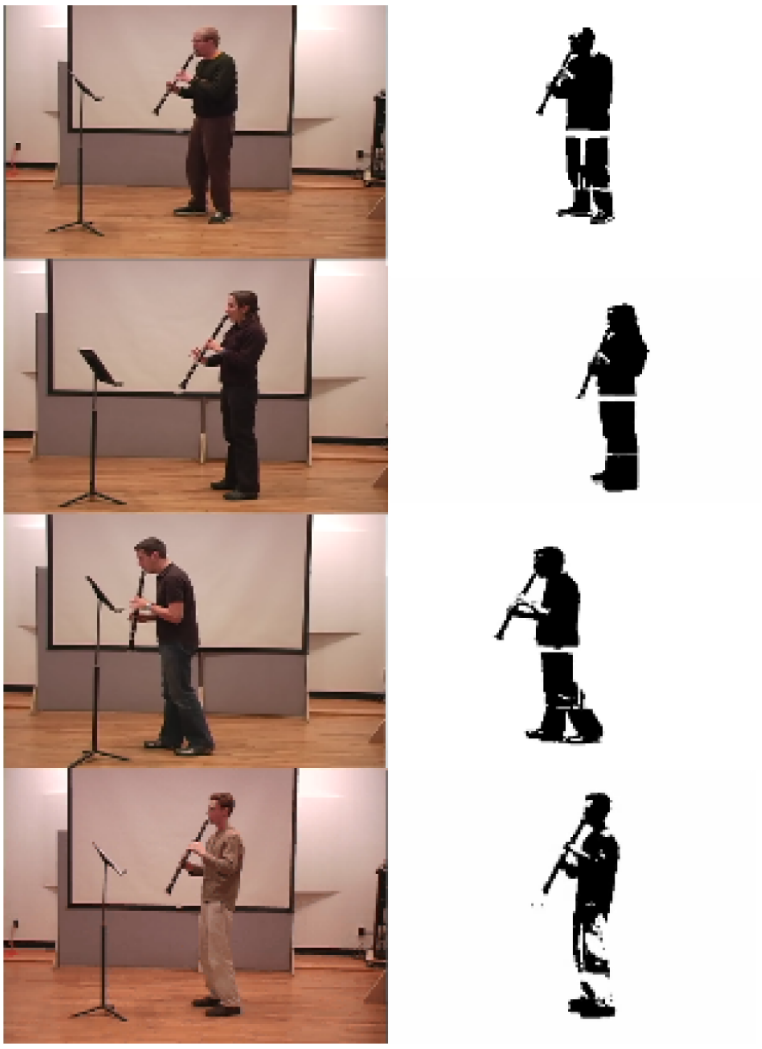

Figura 3: Resultados finais do pré-processamento.

por se tratar de um instrumento anexo ao corpo do clarinetista. A detecção de movimentos baseou-se no rastreamento da trajetória de um marcador virtual fixado na extremidade do instrumento. Este rastreamento foi viabilizado através de um simples algoritmo em linguagem C, com o auxílio da biblioteca OpenCV [Bradski and Kaehler, 2008]. No entanto, ter o valor puro das coordenadas do marcador não foi o suficiente para indicar ou não a presença do movimento. Fez-se necessário relativizar a medição, baseado em um referencial interno ao músico. Era preciso que fosse interno, pois, caso não fosse, os deslocamentos do instrumentista atrapalhariam na manutenção da confiabilidade do referencial. Por isso, determinou-se o centro de massa de cada músico, em cada vídeo. Agora, referenciando a trajetória do marcador virtual ao centro de massa (quadro a quadro), pôde-se gerar a curva de deslocamento. Limiarizando esta curva, foi possível inferir os instantes dos vídeos onde ocorreram movimentos auxiliares de clarinete. Estes resultados estão exemplificados na Figura 4.

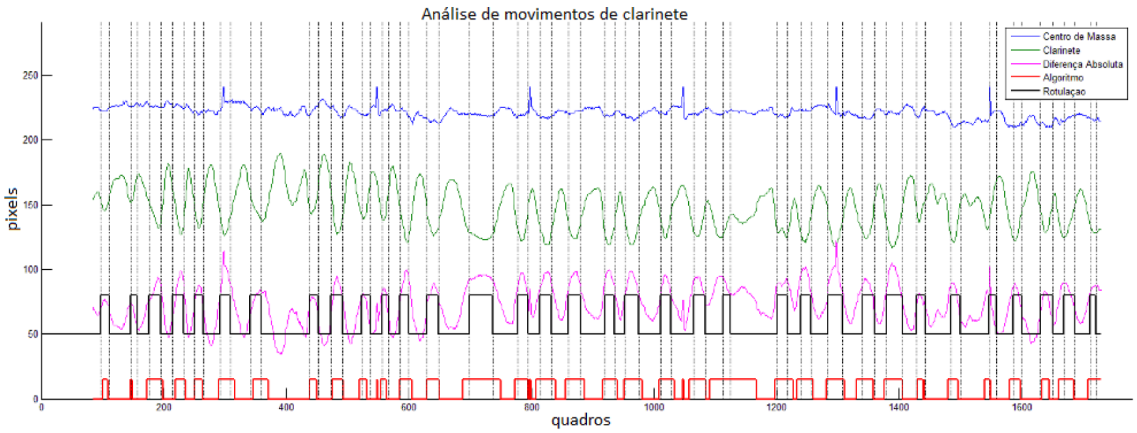

Figura 4: Análise de movimentos auxiliares de clarinete.

Para a detecção de gestos de dobra de joelhos o procedimento foi um pouco mais trabalhoso. Não se tem como rastrear pontos neste caso, uma vez que os joelhos se sobrepõem e que não há região destacada para que se possa rastrear coerentemente. Por isso, optou-se pelo emprego da técnica de projeção de perfil [14]. Gerou-se o perfil horizontal nos nove vídeos e analisou-se o comportamento da contagem de pixels na região, conforme ilustra a Figura 5a. Redução no valor de contagem de pixels indicaria os intervalos dos vídeos onde houve movimentos de dobra de joelhos (Figura 5b). Para relativizar esta medida, usou-se novamente o centro de massa, porém não sua coordenada original, mas sim um deslocamento. Originalmente o centro de massa localiza-se atrás dos joelhos, ou seja, a redução da contagem de pixels nunca se confundiria com esta referência. Assim, calculou-se um percentual do centro de massa, que seria análogo a um deslocamento à frente deste ponto. Agora, todas as vezes em que a região dos joelhos na projeção de perfil horizontal atingisse ou ultrapassasse o percentual de centro de massa, indicaria os intervalos de tempo contendo movimento. A Figura 6 exemplifica este resultado. Cada linha do gráfico representa a linha da imagem, enquanto que cada coluna é equivalente ao frame analisado. Quando a contagem de pixels é inferior ao limitador, gera-se o aspecto escuro no gráfico, ou seja, o movimento.

Por fim, programou-se a detecção de movimentos de ondulações nas costas. Similar à metodologia dos gestos de dobra de joelhos utilizou-se a técnica de projeção de perfil [Zramdini and Ingold, 1993], porém agora na vertical. Desta vez, a técnica foi aplicada apenas nos trechos da imagem onde o músico se fez presente. Isso foi possível através da determinação de bounding boxes. A bounding box nada mais é que um retângulo gerado pelas extremidades da imagem. Neste caso as extremidades são os braços, a cabeça e os pés do músico, ou seja, a pessoa sempre estará circunscrita à bounding box. Assim,

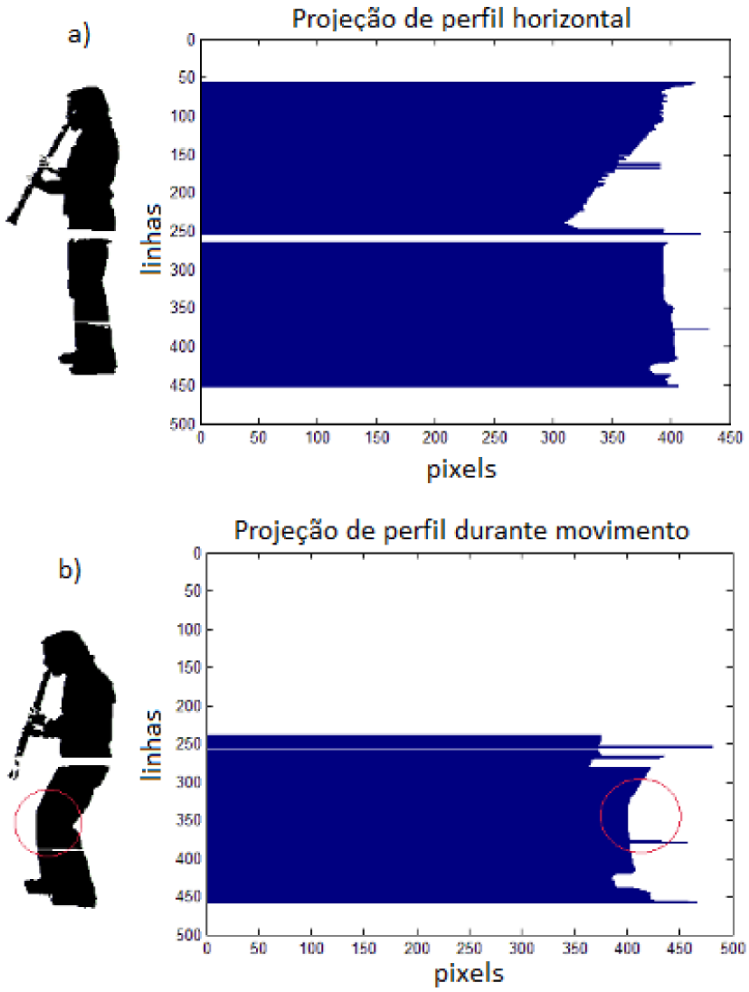

Figura 5: Projeção de perfil horizontal para detecção de movimentos auxiliares de dobra de joelhos.

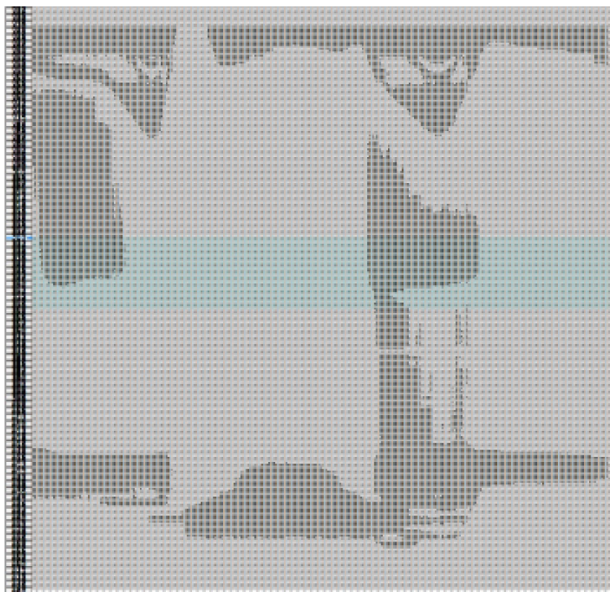

Figura 6: Análise de movimentos auxiliares de dobra de joelhos.

aplicando a projeção de perfil vertical dentro destes retângulos, gera-se o perfil vertical do músico em todos os quadros dos vídeos, conforme a Figura 7a. O movimento de ondulação nas costas pode ser resumido a variações nas barras de contagem de pixels limitadas pelo retângulo e pela cabeça do músico. Transições de pequenas para grandes áreas sugerem movimentos. Para determinar essas variações, calcularam-se dois pontos de interesse no perfil: um na cabeça e um relativo à última barra da contagem de pixels (Figura 7b). Calculando a área formada por esses dois pontos e um terceiro co-linear a ambos, pôde-se inferir a área caracterizadora do movimento (Figura 7c). Análogo ao movimento de clarinete, limiarizando os valores de área obtidos, teve-se como indicar os intervalos de ocorrência do movimento auxiliar de ondulações nas costas, conforme ilustra a Figura 8.

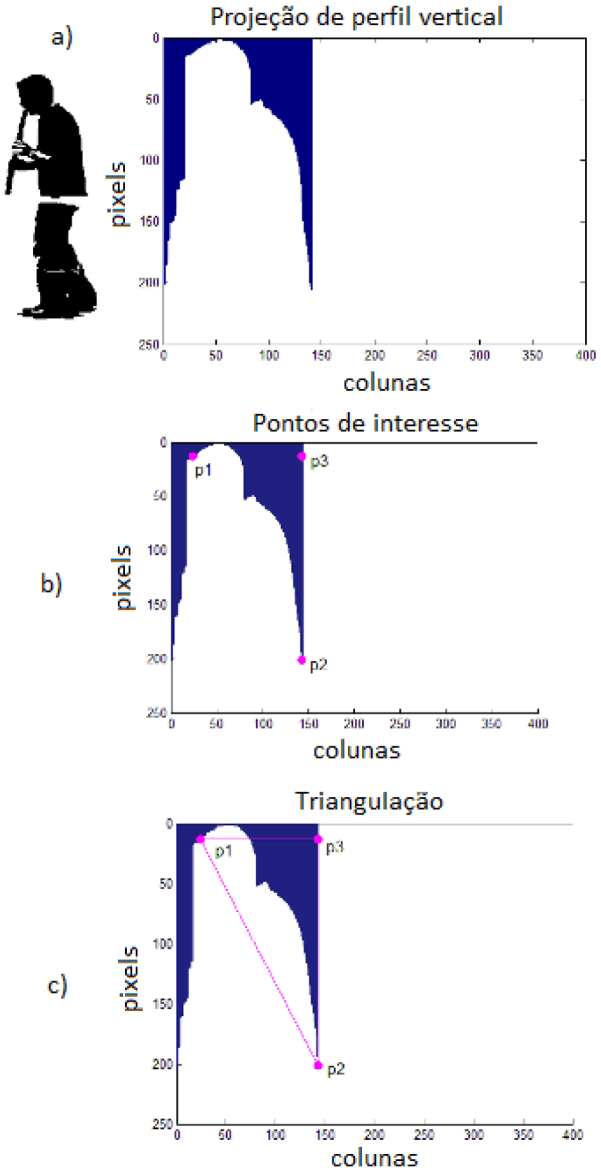

Figura 7: Projeção de perfil vertical para detecção de movimentos auxiliares de ondulações nas costas.

3. RESULTADOS

O método proposto foi avaliado em nove sequências de vídeos de quatro clarinetistas executando o Segundo Movimento das Três Peças de Stravinsky para Clarinete Solo que

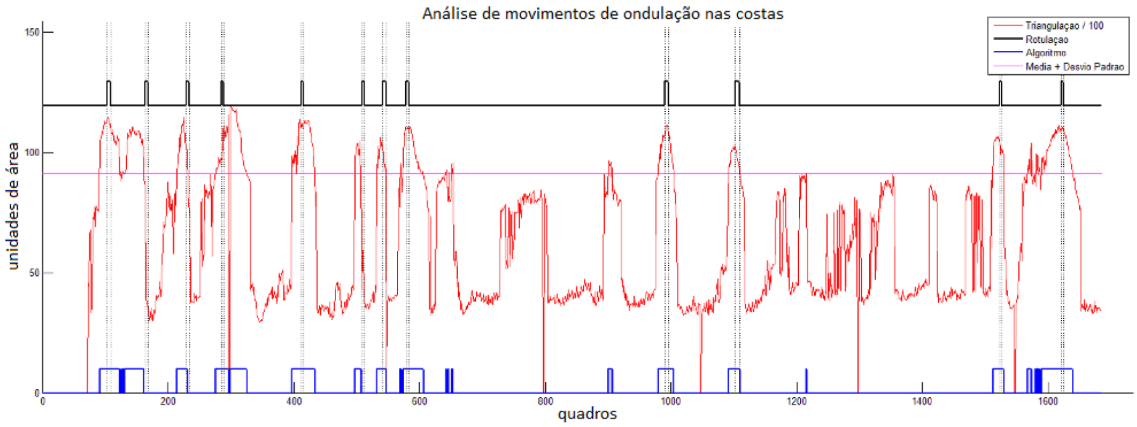

Figura 8: Análise de movimentos auxiliares de ondulações nas costas.

foram gravados no Input Devices and Music Interaction Laboratory da McGill University, Canada. Os vídeos foram gravados com uma única câmera em um estúdio com uma iluminação constante. Em particular, esta peça foi escolhida em razão da sua estrutura não-métrica; a ausência de uma métrica consistente foi utilizada para filtrar alguns movimentos rítmicos que são omnipresentes em música métrica [Wanderley et al., 2005]. Esta peça também é ideal para este estudo por que é sem acompanhamento; a representação mental do músico para o trabalho não inclui outros eventos sonoros, senão aqueles presentes em seu próprio som. Finalmente, esta peça faz parte do repertório padrão de clarinetistas avançados, o que permite a replicação ou extensão de pesquisas prévias com músicos de diferentes escolas.

Os vídeos foram capturados em um estúdio usando uma câmera de vídeo digital instalada em paralelo com o músico. A iluminação foi a do próprio ambiente, assim como o fundo. A resolução dos vídeos variam de 720x480 à 400x300 pixels e a taxa de quadros varia entre 15 e 25 quadros por segundo. Cada vídeo possui entre 1.300 e 1.700 quadros. Além disso, cada vídeo foi rotulado de acordo com a presença ou ausência dos três movimentos auxiliares de interesse. Esta rotulação é necessária para avaliar posteriormente a precisão do método proposto. Os vídeos rotulados totalizam 335 movimentos auxiliares, sendo 151 movimentos de clarinete, 54 movimentos de dobra de joelhos e 130 movimentos de ondulações nas costas. Já os algoritmos aqui propostos totalizaram 486 detecções de movimento, sendo 245 movimentos auxiliares de clarinete, 75 movimentos não óbvios de dobra de joelhos e 166 movimentos de acompanhamento de ondulações nas costas.

Dentre os 486 intervalos de tempo contendo movimento, 82 foram falsos positivos, ou seja, o algoritmo indicou os quadros, no entanto não existiam movimentos de fato. Além disso, outros 39 movimentos que ocorreram não foram detectados, ou seja, falsos negativos.

4. CONCLUSÃO

É possível reparar que o método de detecção proposto identificou cerca de 20% a mais de movimentos auxiliares que a tradicional observação dos vídeos. Além disso, aproximadamente 75% dos movimentos de clarinete, 70% dos movimentos de dobra de joelhos e mais de 80% dos movimentos de ondulações nas costas foram detectados corretamente.

A maioria das falhas de detecção (falsos negativos + falsos positivos) foi de falsos positivos e devido a falhas de pré-processamento e, principalmente, de erros de tomada de decisão baseada na limiarização das curvas - cerca de 70% dos erros. Este resultado é tolerável, uma vez o método proposta visa a substituir apenas as observações. Como os

resultados dos algoritmos são uma espécie de filtragem dos vídeos para análises posteriores de dados obtidos via marcadores, os falsos positivos implicariam apenas em análises desnecessárias. Já os falsos negativos sim são problemáticos, uma vez que representam perda de dados.

Enfim, os resultados aqui expostos são condicionados à base estudada, não sendo possível sua generalização, entretanto, foram conclusivos e comprovam que é possível detectar movimentos auxiliares de clarinete, de dobra de joelhos e de ondulações nas costas através de simples arranjos de algoritmos de visão computacional, simplificando o método de avaliação.

Para se melhorar estes resultados, faz-se necessário aprimorar as técnicas de pré-processamento dos vídeos e estudar formas alternativas de tomada de decisão, ou seja, substituir a proposta de limiarização das curvas por outros arranjos menos suscetíveis a erros, para então compilar os métodos em um único software detetor dos três gestos auxiliares propostos.

Referências

Aggarwal, J. K. and Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding*, 73:428–440.

Bradski, G. and Kaehler, A. (2008). Learning opencv: Computer vision with the opencv library. *O'Reilly*, 1:556.

Cadoz, C. (1998). Instrumental gesture and music composition. *Proceedings of the International Computer Music Conference*, 1:1–12.

Dahl, S. and Friberg, A. (2004). Expressiveness of musicians' body movements in performances on marimba. *Gesture-Based Communication in Human-Computer Interaction*, 1:479–486.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21:103–113.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73:82–98.

Gonzalez, R. C. and Woods, R. E. (2007). Digital image processing. *Prentice Hall*, 3:976.

Jensenius, A. R., Wanderley, M. M., Godoy, R. I., and Leman, M. (2010). Musical gesture: Concepts and methods in research. *Musical Gestures: Sound, Movement and Meaning*, 1:12–35.

Mitra, S. and Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man and Cybernetics*, 37(3):311–324.

Teixeira, E. C. F. (2010). Análise da expressividade musical com base em medidas acústicas e do gesto físico. *Dissertação de Mestrado*, 1:46.

Verfaille, V., Quek, O., and Wanderley, M. M. (2006). Sonification of musicians' ancillary gestures. *Proceedings of the 12th International Conference on Auditory Display*, 1:20–23.

Wanderley, M. M. (1999). Non-obvious performer gestures in instrumental music. *Gesture-based Communication in Human-Computer Interaction*, 1:37–48.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. *Gesture and Sign Language in Human-Computer Interaction*, 1:241–253.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34(1):97–113.

Zramdini, A. and Ingold, R. (1993). Optical font recognition from projection profiles. *Electronic Publishing*, 6(3):249–260.