Wellingson da Silva Moraes

Detecção e Identificação de Teclas de Pianos Eletrônicos Usando Técnicas de Processamento de Imagens

São Luís - MA

2020

Wellingson da Silva Moraes

Detecção e Identificação de Teclas de Pianos Eletrônicos Usando Técnicas de Processamento de Imagens

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Msc. Giovanni Lucca França da Silva

São Luís - MA

2020

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Moraes, Wellingson da Silva.
Detecção e Identificação de Teclas de Pianos Eletrônicos Usando Técnicas de Processamento de Imagens / Wellingson da Silva Moraes. – 2020.
49 f.

Orientador(a): Giovanni Lucca França da Silva.
Monografia (Graduação) – Curso de Ciência da Computação, Universidade Federal do Maranhão, São Luís – MA, 2020.

1. Piano Eletrônico. 2. Processamento de Imagens. 3. Segmentação de Imagens. I. Silva, Giovanni Lucca França da. II. Título.

Wellingson da Silva Moraes

Detecção e Identificação de Teclas de Pianos Eletrônicos Usando Técnicas de Processamento de Imagens

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho Aprovado. 08 de Janeiro de 2020:

Prof. Msc. Giovanni Lucca França da Silva
Orientador
Universidade Federal do Maranhão

Prof. Msc. Carlos Eduardo Portela Serra de Castro
Examinador
Universidade Federal do Maranhão

Prof. Msc. João Otávio Bandeira Diniz
Examinador
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão

São Luís - MA
2020

Agradecimentos

A Deus, por abrir as portas durante minha caminhada e permitir que eu desse mais esse passo.

Ao meus pais, Pedro e Raimunda, pelos sacrifícios e dedicação que tiveram na minha criaçao.

Aos meus irmãos, Wellington e Eliane, por serem minhas bússolas e exemplos de vida.

À Emmile, por estar ao meu lado, trazendo alegria para os meus dias, esperança e força para concluir este trabalho.

Ao professor Giovanni, por ter aceitado ser meu orientador, pela sua paciência e sabedoria na condução deste trabalho.

Aos membros da banca examinadora, Portela e João Otávio pela disponibilidade de participar e pelas contribuições acerca da monografia.

Aos amigos de graduação do Departamento Acadêmico de Ciência da Computação, Robherson, Leonardo, Flávio e Mateus, compartilhamos e superamos momentos difíceis.

À Gabi, monumento do Centro de Ciências Humanas pelos seus conselhos e motivação diária para continuar.

Aos professores que contribuíram na minha formação acadêmica e aos demais amigos e colegas que fiz durante o curso.

Dedico este trabalho a todos que, assim como eu, reaprenderam a sonhar.

"Amar muitas coisas, por aí que reside a verdadeira força, e quem ama muito executa muito, e pode realizar muito mais, e aquilo que é feito com amor é bem feito."

(Vincent Van Gogh - Carta 143)

Resumo

O reconhecimento de objetos é um dos principais processos que integram o Processamento de Imagens pois envolve os problemas de identificação dos objetos, definindo quais fazem parte da cena, e a localização espacial deles. Esses desafios impõem a identificação de tarefas isoladas, solucionando separadamente os problemas através da aplicação de técnicas de Processamento de Imagens e integrando-os, criando metodologias que reúnam as informações resultantes destes módulos independentes. A metodologia proposta neste trabalho aplica as técnicas de extração de contornos e comparação de histogramas para a segmentação de teclado de pianos eletrônicos e identifica individualmente as teclas com uma busca de padrões de disposição de teclas pretas e brancas ao longo do teclado, através da aplicação da detecção de linhas de Hough e da extração de componentes conexos. A metodologia foi aplicada em dois modelos comuns de pianos eletrônicos, de 61 e 76 teclas, rotulando as teclas de acordo com a escala musical.

Palavras-chaves: Processamento de Imagens, Segmentação de Imagens, Piano Eletrônico.

Abstract

Object detection is one of the most important processes of Image Processing as it involves problems of identifying objects, defining what objects are in the scene, and their spatial localization. These challenges require the identification of individual tasks, solving problems separately by using Image Processing techniques and combining the results of these independent modules by creating new methodologies. The proposed methodology in this work applies techniques of contour extraction and histogram comparison for electronic piano's keyboard segmentation and identifies individually black and white keys by searching for patterns repeated across the entire keyboard through application of Hough line detection and connected components extraction. The methodology was applied in two popular models of electronic pianos (61 and 76 keys) labeling the different keys according to the musical scale.

Keywords: Image Processing, Image Segmentation, Electronic Piano.

Lista de ilustrações

Lista de abreviaturas e siglas

2D	Duas Dimensões
3D	Três Dimensões
CLAHE	Equalização Adaptativa de Histograma com Limitação de Contraste
PI	Processamento de Omagens
RO	Reconhecimento de Objetos
ROI	Região de Interesse
VC	Visão Computacional

Sumário

1 Introdução

A visão humana nos proporciona a habilidade do tomar decisões de acordo com a observação de informações, a Visão Computacional busca imitar esse processo. O Processamento de Imagens é uma tarefa essencial da Visão Computacional que utiliza técnicas para a obtenção de descrições das imagens que contenham informações suficientes para diferenciar os objetos de forma confiável (PEDRINI; SCHWARTZ, 2008).

No passado, as restrições das tecnologias dos sistemas computacionais, tanto em velocidade quanto em custo, dificultavam a implementação das mais simples técnicas de processamento. Hoje, essas restrições têm se tornado cada vez menos a preocupação dos que trabalham na área (JAMES, 1987), o crescente avanço da tecnologia digital associado ao desenvolvimento de novos algorítmos permite que computadores pessoais consigam ser utilizados para implementar e testar as técnicas disponíveis em bibliotecas, como a Open Source Computer Vision Library (OpenCV), que é de uso livre e voltada à elaboração de aplicações da área de Visão Computacional e Processamento de Imagens.

O Processamento de Imagens (PI) é definido como o conjunto de técnicas que capturam, transformam e representam imagens com o auxílio de computadores. Operações típicas das etapas iniciais de PI como a redução de ruído, aumento de contraste, extração de bordas e compressão de imagens utilizam pouco conhecimento sobre a semântica das imagens e, por isso, técnicas de análise de imagens foram desenvolvidas. Essas técnicas utilizam algorítmos que recebem imagens previamente submetidas a um tipo de processamento e buscam interpretá-las para extrair informações, envolvendo tarefas como a segmentação da imagem em regiões ou objetos de interesse baseando-se na forma, textura, níveis de cinza ou nas cores dos objetos presentes nas imagens (PEDRINI; SCHWARTZ, 2008).

De acordo com Gonzalez e Woods (2018), os passos relevantes que envolvem o Processamento de Imagens são sequenciados por um operador humano que detém o conhecimento ou a experiência sobre o domínio da aplicação. A combinação do avanço técnologico com o desenvolvimento de metodologias que utilizam as técnicas de PI tem permitido resolver problemas em diferentes áreas (PEDRINI; SCHWARTZ, 2008). Exemplos de domínios de conhecimento que são aplicados incluem medicina, indústria, militar, sensoriamento remoto, artes, segurança e, neste trabalho, pretende-se aplicar uma combinação satisfatória dessas técnicas em um contexto musical.

1.1 Justificativa

Formalmente, aprender a tocar uma música em um piano eletrônico exige do aluno o aprendizado de leitura de partitura, sistema de escrita que associa símbolos às notas musicais. Cada símbolo indica qual tecla deve ser pressionada, a duração e o tempo. Enquanto a maioria dos intrumentos exigem apenas a leitura de uma linha da partitura, um piano exige duas, o que adiciona mais dificuldade no processo de aprendizagem.

É imposta ao aluno que quer aprender a tocar um piano eletrônico uma dificuldade formal, seja pelo tempo necessário para o aprendizado da leitura dos diferentes símbolos que uma partitura pode conter e o entendimento dos seus significados, seja pela dificuldade em traduzi-los em movimentos mecânicos das mãos.

O uso de cores é uma das mais importantes experiências visuais nos seres humanos (ADAMS; OSGOOD, 1973). Um auxílio que indicasse a tecla a ser pressionada através de uma sugestão visual aceleraria o processo de leitura, tornando mais simples o processo de aprendizado do instrumento musical. No entanto, não existe na literatura trabalhos que identifiquem as teclas em pianos eletrônicos, primeiro passo e de fundamental importância para qualquer aplicação que utilize o instrumento como objeto de interesse.

1.2 Objetivo

O objetivo deste trabalho é desenvolver uma metodologia automática para a segmentação de teclado e a identificação das teclas pretas e brancas de pianos eletrônicos utilizando as técnicas de PI.

1.2.1 Objetivos Específicos

Os objetivos específicos pretendidos para este trabalho são:

- Adquirir a base de vídeos para validação da metodologia.
- Desenvolver e aplicar técnicas de segmentação de teclas.
- Realizar o reconhecimento das teclas pretas e brancas presentes em teclados de Pianos Eletrônicos.

1.3 Organização do trabalho

Este trabalho está dividido em capítulos na seguinte estrutura: no Capítulo 2 será apresentada a fundamentação teórica necessária para o entendimento da metodologia proposta. Esse capítulo está dividido em três seções: a primeira apresenta o piano eletrônico,

descrevendo sua estrutura e os padrões encontrados em seu teclado, a segunda seção apresenta fundamentos de imagens digitais e a terceira aborda as técnicas de processamento de imagens aplicadas neste trabalho. O Capítulo 3 descreve e detalha todas as etapas do método proposto. No Capítulo 4 são mostrados e discutidos os resultados obtidos pela metodologia proposta. Por fim, o Capítulo 5 apresenta as considerações finais e sugestões de trabalhos futuros.

2 Fundamentação Teórica

O conhecimento de alguns conceitos e técnicas se faz necessário para o entendimento deste trabalho. Esses conceitos e técnicas são apresentados neste capítulo. É feita uma exposição da estrutura de um Piano Eletrônico e os padrões encontrados em seu teclado, um resumo de fundamentos de imagens digitais e apresentação das técnicas computacionais de PI que são utilizadas na metodologia proposta por este trabalho.

2.1 Piano Eletrônico

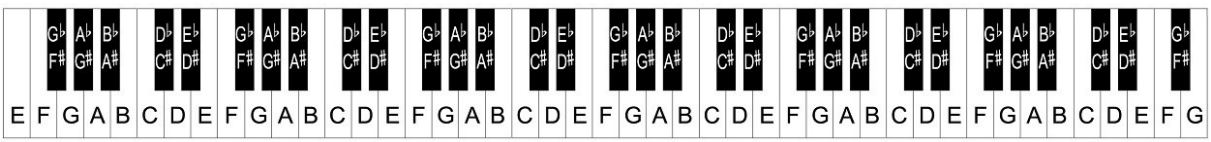
Figura 1 – Localização do teclado de um piano eletrônico.

Fonte: Autor

Um piano eletrônico é um instrumento composto por um conjunto de teclas adjacentes pretas e brancas que quando pressionadas produzem os sons. As teclas geralmente são agrupadas em 12 teclas arranjadas de acordo com a escala musical anglo-saxônica, sendo elas: C (Dó), C# (Dó Sustenido), D (Ré), D# (Ré Sustenido), E (Mi), F (Fá), F# (Fá Sustenido), G (Sol), G# (Sol Sustenido), A (Lá), A# (Lá Sustenido) e B (Si). Piano eletrônicos vêm em diferentes tamanhos e formatos, dependendo do fabricante a quantidade de teclas variam, com modelos de 61, 76 ou 88 teclas (UNSOUND, 2019).

A Figura 1 destaca a localização do teclado em um piano eletrônico e a identificação das teclas é ilustrada na Figura 2.

Figura 2 – O teclado de um piano eletrônico moderno com os valores das teclas.



Fonte: Adaptado de Yamaha Keyboard Guide (2019)

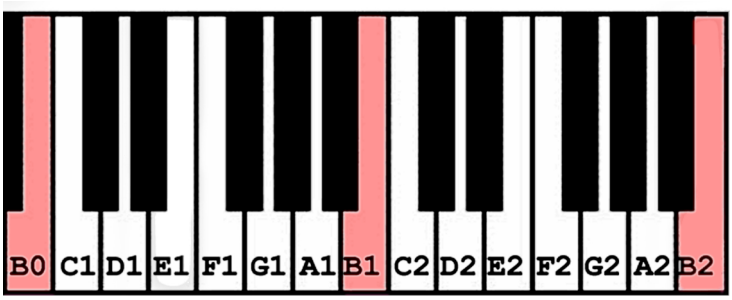
2.1.1 Padrões do teclado de um Piano Eletrônico

Alguns padrões podem ser encontrados na disposição das teclas ao longo do teclado de um piano eletrônico, dois importantes são apresentados a seguir.

2.1.1.1 Oitava

A distância entre qualquer tecla e a próxima tecla do mesmo valor é conhecida como uma oitava. Técnicamente, a frequência entre uma nota e sua próxima oitava é a multiplicação por dois (DUNNE, 2019). Por exemplo, a frequência da nota Lá Maior (A4) é 440hz, a posição do próximo Lá Maior (A5) é 880hz. Na Figura 3, a distância entre o primeiro Si à esquerda (B0) e o Si ao meio (B1) é uma oitava. Já a distância entre o Si mais à esquerda (B0) e o Si mais à direita (B2) são duas oitavas.

Figura 3 – Ilustração da distância entre oitavas.

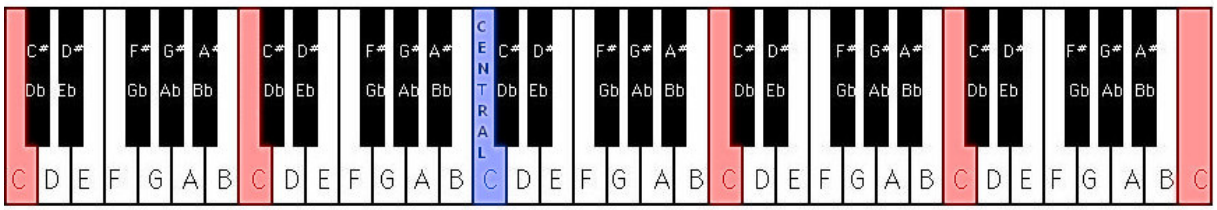


Fonte: Autor

2.1.1.2 Dó Médio

Um padrão que serve como posição fundamental para a localização no teclado e que divide os tons em transcrições musicais é a do *Dó Médio*. A tecla pode ser localizada contando a quantidade k de Dó's (C) presentes ao longo do teclado e identificando o Dó (C) que ocupa a posição média (PIANO WORKSHOP, 2013). A localização do Dó Médio em um teclado é ilustrada na Figura 4 e a dedução da sua posição no Algorítmo 1.

Figura 4 – Localização do Dó Médio.



Fonte: Autor

Algoritmo 1: Localização do Dó Médio

Input: Número k de Dós presentes no teclado

Output: Posição do Dó Médio

if *Número k de Dó's é par* **then**

  localização do Dó Médio é na posição $\frac{k}{2}$

else

  localização do Dó Médio é na posição $\left\lceil \frac{k}{2} \right\rceil$

end

2.2 Imagens Digitais

Uma imagem, formada por um único canal de cor, pode ser descrita como uma função bidimensional *f(x,y)*, onde x e y indicam as coordenadas espaciais e o valor de f representa a intensidade da imagem no ponto (x,y) (BALLARD; BROWN, 1982; FACON, 2002). A Figura 5 mostra uma imagem e a orientação do sistema de coordenadas. Por convenção, a origem da imagem é localizada no canto superior esquerdo dela.

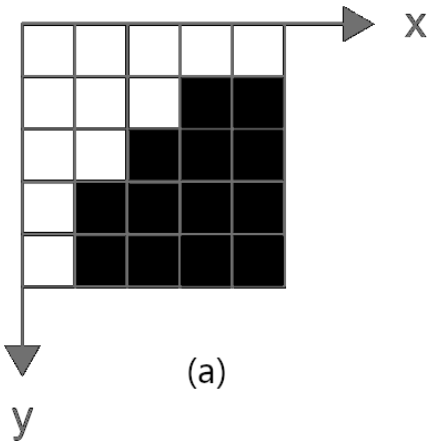
Figura 5 – Convenção do sistemas de coordenadas na representação de imagens digitais.

Fonte: Autor

A maioria das técnicas de análise de imagens é realizada por meio de processamento computacional, de forma numérica, logo, a função *f(x,y)* deve ser discretizada. Quando as coordenadas espaciais e a intensidade assumem valores finitos e discretos, chamamos a imagem de imagem digital. Neste aspecto prático, as imagens digitais são representadas por matrizes em vez de funções (FACON, 2002) (Figura 6(a)). Cada elemento da imagem digital é chamado de pixel (*picture element*) e associa um ponto *(x,y)* do domínio espacial a uma intensidade.

O processo de converter uma imagem em um array de números é conhecido como digitalização, que envolve dois passos, a amostragem e a quantização. A amostragem discretiza o domínio de definição da imagem nas direções x e y, gerando uma matriz de M

Figura 6 – Representações de uma imagem digital.



$$f = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

(b)

Fonte: Autor

x N amostras, onde cada amostra é um pixel. O valor inteiro de cada pixel é obtido pela quantização.

Neste trabalho, são utilizadas imagens digitais binárias, que assumem intensidades no intervalo [0,1], branco (0) e preto (1). Uma imagem binária representada na forma matricial está ilustrada na Figura 6(b).

2.2.1 Relacionamento entre pixels

Esta seção apresenta relacionamentos básicos entre elementos de uma imagem.

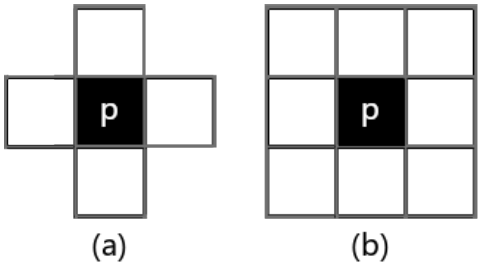
2.2.1.1 Vizinhança de pixel

Um importante conceito que define a proximidade dos pixels é a vizinhança. Um pixel p de coordenadas *(x,y)* possui 4 vizinhos horizontais e verticais. As coordenadas desses vizinhos são *(x+1,y)*, *(x-1,y)*, *(x,y+1)* e *(x,y-1)*. Esses pixels formam a chamada "*4-vizinhança*"de p, também denotada por N4(p) e ilustrada na Figura 7(a). O pixel p também possui 4 vizinhos diagonais, eles são os de coordenadas *(x-1,y-1)*, *(x-1,y+1)*, *(x+1,y-1)* e *(x+1,y+1)*, que formam o conjunto Nd(p). A união dos conjuntos N4(p) e Nd(p) formam a "*8-vizinhança* "de p, denotada por N8(p) e ilustrada na Figura 7(b).

2.2.1.2 Conectividade

Dois pixels são ditos conexos se existe uma sequência de pixels que os liga, tal que os dois pixels consecutivos dessa sequência satisfazem uma condição de conectividade estabelecida. A condição pode ser características comuns como intensidade de cor ou textura ou que eles são vizinhos. A conectividade então é condicionada pela noção de vizinhança (*4-* ou *8-vizinhança*) (2.2.1.1 e é um conceito importante utilizado para estabelecer limites de

Figura 7 – Tipos de vizinhança.

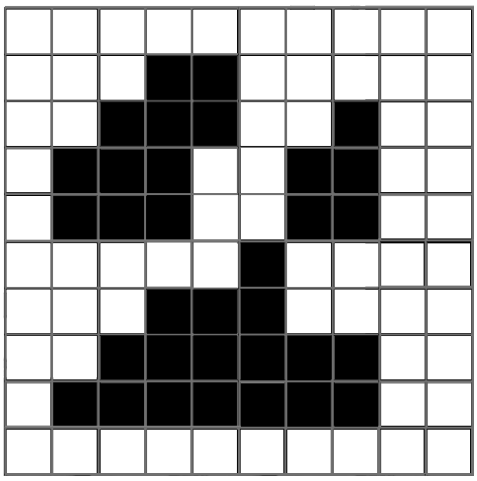


Fonte: Adaptado de Pedrini e Schwartz (2008)

objetos e componentes de regiões em uma imagem (FACON, 2002; PEDRINI; SCHWARTZ, 2008).

A Figura 8 mostra uma imagem bidimensional contendo três componentes conexos, caso seja considerada a *4-vizinhança* ou, então, dois componentes conexos se considerada a *8-vizinhança*.

Figura 8 – Conectividade de objetos em uma imagem bidimensional.

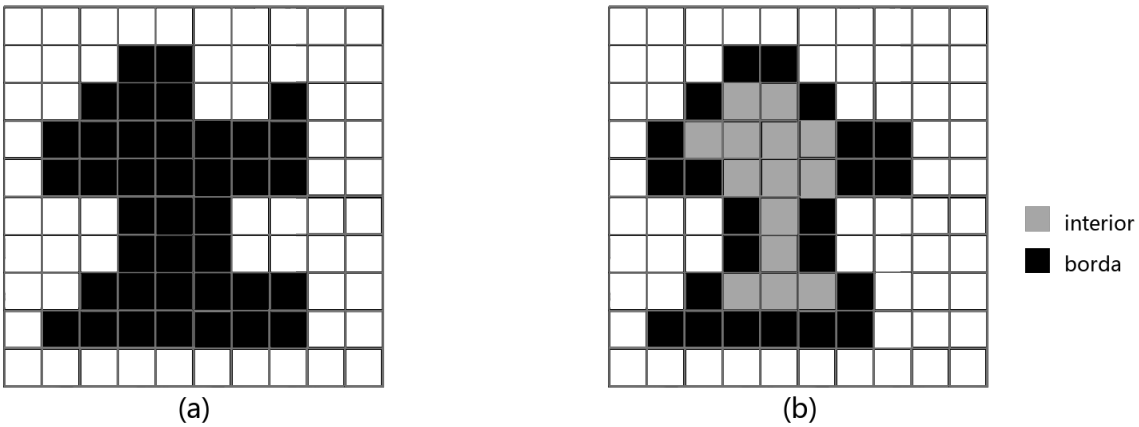


Fonte: Adaptado de Pedrini e Schwartz (2008)

2.2.1.3 Borda e Interior

A *borda* de um componente conexo C em uma imagem bidimensional é o conjunto de pixels que pertencem ao objeto e que possuem 4-vizinhança com um ou mais pixels externos a C. A *borda* corresponde ao conjunto de pontos no contorno do componente conexo. O *interior* é o conjunto de pixels que não estão na sua borda. A Figura 9 ilustra um exemplo onde uma imagem binária tem sua borda e o interior identificados (PEDRINI; SCHWARTZ, 2008).

Figura 9 – Borda e interior de um componente.



Fonte: Adaptado de Pedrini e Schwartz (2008)

2.3 Processamento Digital de Imagens

Entende-se por processamento digital de imagens a manipulação de uma imagem por computador de modo que a entrada e a saída do processo sejam imagens (MARIA, 2000). Esse estágio é usado para pré-processar a imagem, melhorando o aspecto visual de certas feições estruturais, e convertê-la em um formato conveniente para realização de diversas análises adicionais (PEDRINI; SCHWARTZ, 2008).

Segundo Szeliski (2010), o processamento digital de imagens é o primeiro passo de qualquer aplicação na área de VC. No entanto, não existem limites claros se for considerada uma linha contínua com o processamento digital de imagens em um extremo e VC no outro. Por isso, um paradigma útil adotado é dividir as técnicas de análise de imagem em três tipos: processsamento nível baixo, médio e alto (GONZALEZ; WOODS, 2018).

Técnicas de baixo nível envolvem operações primitivas, a *aquisição* de imagens e as tarefas de *pré-processamento*, sendo caracterizadas por tanto a entrada quanto a saída do processo serem imagens (GONZALEZ; WOODS, 2018).

A etapa de *aquisição* captura a imagem por meio de um dispositivo ou sensor e converte-a em uma representação adequada para o processamento digital subsequente. A aquisição da imagem depende muito do domínio do problema, elas podem ser geradas a partir de diversas fontes (i.e. raios gama, raios-x, sonar, ondas de rádio, espectro visível, infravremelho, dentre outras) e meios (sensores, câmeras, satélites, etc). Dentre os aspectos envolvidos nesta etapa estão a escolha do tipo de sensor, as condições de iluminação da cena, a resolução e o número de níveis de cinza ou cores da imagem digitalizada (PEDRINI; SCHWARTZ, 2008).

A imagem resultante do processo de aquisição pode apresentar imperfeições ou degradações decorrentes, por exemplo, das condições de iluminação ou características dos

dispositivos. A etapa de *pré-processamento* visa melhorar a qualidade da imagem por meio da aplicação de técnicas para atenuação de ruído, correção de contraste ou brilho e suavização de determinadas propriedades da imagem (PEDRINI; SCHWARTZ, 2008).

Tarefas como a *segmentação* e *descrição* são consideradas como processamento de imagens de nível médio por Gonzalez e Woods (2018).

A etapa de *segmentação* realiza a extração e identificação de áreas de interesse contidas na imagem. Essa etapa é geralmente baseada na detecção de descontinuidades (bordas) ou de similaridades (regiões) na imagem (PEDRINI; SCHWARTZ, 2008; GONZALEZ; WOODS, 2018).

Estruturas adequadas de representação devem ser utilizadas para armazenar e manipular os objetos de interesse extraídos da imagem. Segundo Pedrini e Schwartz (2008), o processo de *descrição* visa a extração de características ou propriedades que possam ser utilizadas na discriminação entre classes de objetos. Essas características, são em geral, descritas por atributos numéricos armazenados em forma de vetor.

A principal característica desses processos de nível médio é que suas entradas, em geral, são imagens, e as saídas são atributos extraídos dessas imagens (GONZALEZ; WOODS, 2018).

Por fim, o processamento de nível alto envolve as etapas de *reconhecimento* e *interpretação* dos componentes de uma imagem. *Reconhecimento* (ou classificação) é o processo que atribui um identificador ou rótulo aos objetos da imagem, baseado nas características providas pelos seus descritores. O processo de *interpretação* consiste em atribuir um significado ao conjunto de objetos reconhecidos (GONZALEZ; WOODS, 2018; PEDRINI; SCHWARTZ, 2008).

A essência do processamento de imagens está em utilizar diferentes propriedades de uma imagem como cor, correlação entre diferentes pixels, posicionamento de objetos e outros detalhes para extrair características da imagem, como bordas, objetos e seus contornos (PEDRINI; SCHWARTZ, 2008).

2.3.1 Realce de Imagens

Técnicas de realce de imagens buscam acentuar ou melhorar a aparência de determinadas características da imagem manipulando os valores dos pixels, tornando-a mais adequada à interpretação das informações para a aplicação em questão.

2.3.1.1 Brilho e Contraste

Como visto em 2.2, uma imagem digital pode ser representada por uma função bidimensional $f(x,y)$ onde f representa o valor da intensidade do ponto naquela coordenada

espacial.

Associado à sensação visual de uma imagem estar muito escura ou clara demais, o brilho de uma imagem é um descritor subjetivo de percepção de luz. O brilho pode ser caracterizado como a média dos tons de intensidade de uma imagem (SMITH, 1997).

O contraste pode ser descrito como a diferença de brilho entre objetos e o fundo de imagens (SMITH, 1997). É ele que controla a intensidade das cores, aumentando a intensidade de pontos escuros, de modo que os pontos claros se destacam, ficando mais visíveis.

2.3.1.2 Filtragem Gaussiana

Filtros são técnicas computacionais utilizadas para corrigir tipos de ruídos específicos ou realçar características em uma imagem, como por exemplo bordas (GONZALEZ; WOODS, 2018).

O filtro gaussiano é um dos diversos filtros que são encontrados na literatura, ele utiliza a função gaussiana para calcular os coeficientes das máscaras a serem aplicadas na imagem. A sua utilização é mais apropriada quando é necessária uma suavização das imagens, tendo um efeito de desfocagem e, com isso, reduzindo ruídos (PEDRINI; SCHWARTZ, 2008). O filtro é calculado pela equação:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{2.1}$$

Onde:

σ é o desvio padrão gaussiano;

2.3.1.3 Equalização de Histograma

O histograma de uma imagem corresponde à distribuição dos níveis de cinza da imagem, é uma ferramenta que serve de base para algorítmos de segmentação (PEDRINI; SCHWARTZ, 2008). Ele pode ser representado por um gráfico indicando o número de pixels na imagem para cada nível de cinza, possuindo diversos modelos, como o unimodal, bimodal e multimodal (Figura 10).

O histograma $H(p)$ de valores discretos pode ser denotado por:

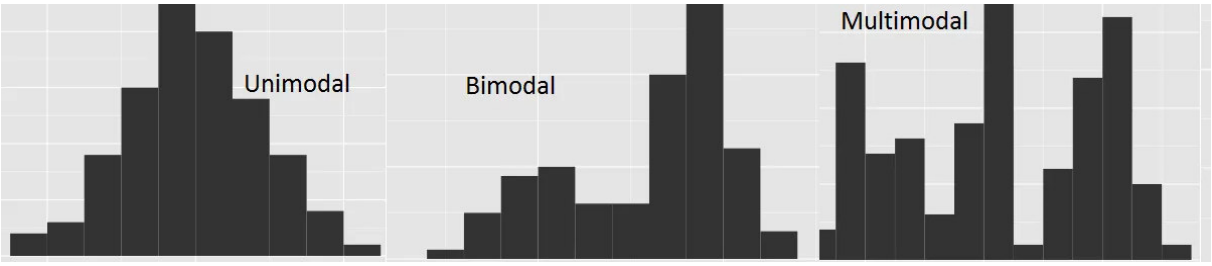
$$H(p) = \frac{N_{pn}}{N_{pt}}, n = 0, 1, \cdots, k \tag{2.2}$$

Onde:

N_{pn} é o número de pixels com intensidade n;

N_{pt} é o número total de pixels;

Figura 10 – Modelos de histogramas unimodal, bimodal e multimodal.



Fonte: Adaptado de Make Me Analyst (2019)

k é a quantidade de níveis de intensidade;

$$img_{i,j} = \left\lfloor k \sum_{n=0}^{f_{i,j}} H(p) \right\rfloor \tag{2.3}$$

Imagens img obtidas após o processo de equalização do histograma (2.3) possuem uma distribuição mais uniforme dos seus níveis de cinza, apresentando melhor contraste, o que possibilita uma melhor detecção de características da imagem.

2.3.1.3.1 Equalização Adaptativa de Histograma com Limitação de Contraste

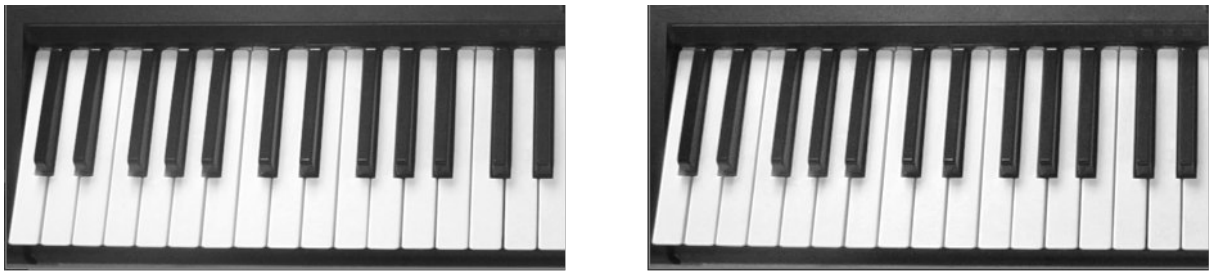
A Equalização Adaptativa de Histograma com Limitação de Contraste (CLAHE) é uma das variações de equalização de histograma existentes. O algoritmo separa a imagem em regiões contextuais e aplica a equalização de histograma em cada uma delas (PIZER et al., 1987), ele é dado pela equação 2.4:

$$p_n = (p_{max} - p_{min})H(p) + p_{min} \tag{2.4}$$

onde p_n é novo valor do pixel, p o valor inicial do pixel, p_{max} e p_{min} são os valores máximos e mínimos de pixel respectivamente, inseridos através da faixa máxima de valores de tons de cinza da imagem. A função $H(p)$ é o valor do histograma usado na região de interesse.

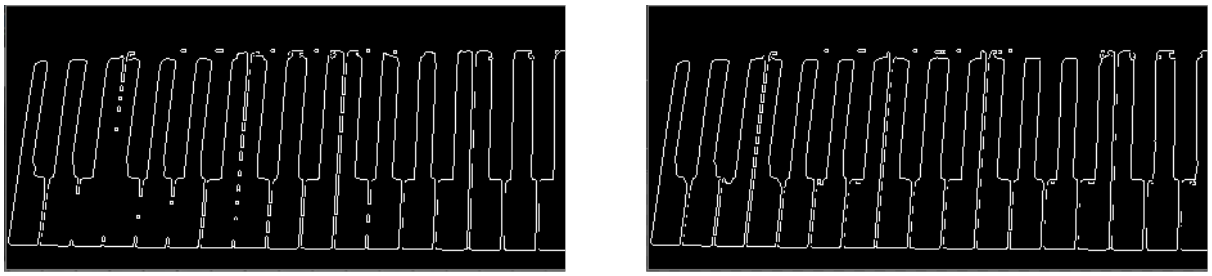
A Figura 11 ilustra a equalização de histograma por CLAHE. Embora a diferença de iluminação seja pouco perceptível, ela foi decisiva na utilização da filtragem de Canny, técnica que será apresentada na seção (2.3.3.1). As regiões que limitam as teclas ficaram mais visíveis na imagem que teve o histograma equalizado por CLAHE, como ilustra a Figura 12.

Figura 11 – Comparação entre uma imagem original e uma imagem com histograma equalizado por CLAHE.



Fonte: Autor

Figura 12 – Aplicação do filtro de Canny nas imagens da Figura 11.



Fonte: Autor

2.3.2 Morfologia Matemática

A morfologia matemática é baseada na Teoria dos Conjuntos e se refere ao estudo e análise de imagens usando operadores não lineares. Os estudos foram inicializados por Georges Matheron e Jean Serra na década de 60, sendo originalmente desenvolvida para manipular imagens binárias (SERRA, 1983).

A idéia base da morfologia é percorrer uma imagem com um objeto de forma conhecida, denominado elemento estruturante, comparando-o com o objeto que queremos analisar. O resultado dessa comparação é quantificado de acordo com a maneira que este se encaixa na imagem.

A área mais prática de PI tem se ocupado com o processamento de imagens binárias (JAMES, 1987). Por convenção, objetos são representados por pixels pretos enquanto o fundo é representado por pixels brancos. Além de ocupar pouca memória para armazenamento, imagens binárias são mais favoráveis ao aparecimento bem definido de propriedades como limite, área e forma (PEDRINI; SCHWARTZ, 2008).

As principais atividades que a morfologia executam em PI são: remoção de imperfeições, aumento de qualidade, detecção de falhas, restauração e obtenção de informações

a respeito da forma e estrutura de imagens (GONZALEZ; WOODS, 2018).

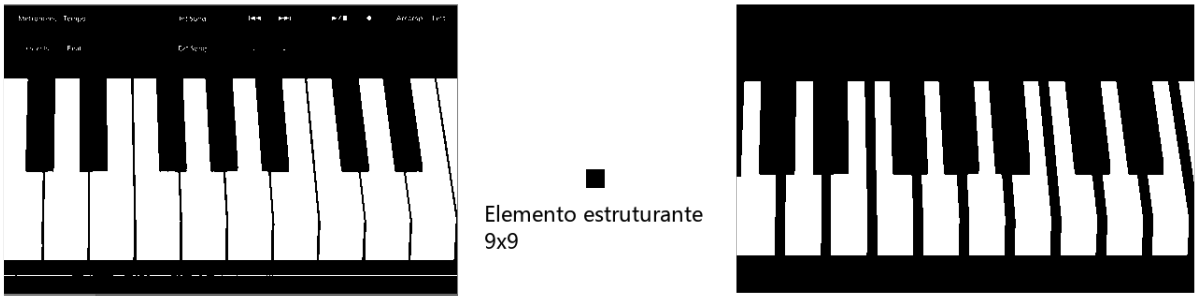
2.3.2.1 Erosão e Dilatação

Há duas operações básicas morfológicas, a erosão e a dilatação, as outras operações são elaboradas a partir delas.

A erosão é aplicada quando pretende-se desaparecer os objetos de tamanho inferior ao do elemento estruturante. Ela pode diminuir o tamanho dos objetos, aumentar buracos ou separar os que antes eram conexos. Esses resultados são obtidos a partir da interseção de todos os conjuntos obtidos a partir de translações dos pixels da imagem binária pela reflexão do elemento estruturante (SZELISKI, 2010; PEDRINI; SCHWARTZ, 2008).

Por exemplo, na Figura 13 as regiões de pixels pretos aumentaram enquanto as regiões de pixels brancos com tamanho inferior ao do elemento estruturante desapareceram.

Figura 13 – Uma iteração da erosão por um elemento estruturante 9x9 todo preto.



Fonte: Autor

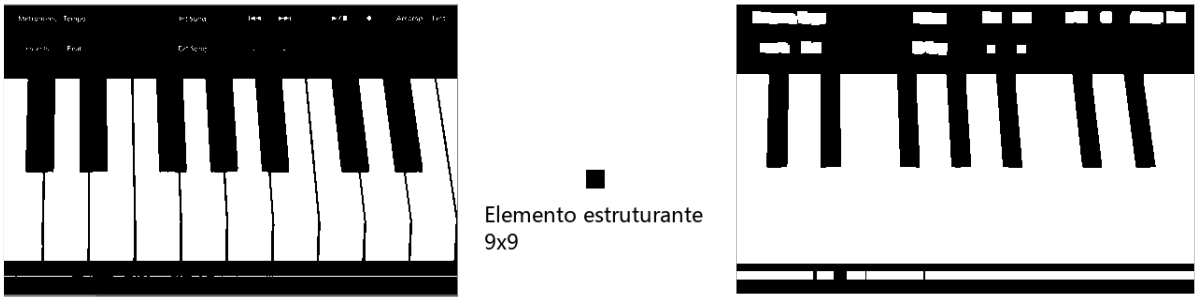
Considera-se como dilatação a união de todos os conjuntos obtidos a partir de translações dos pixels da imagem binária pelo elemento estruturante. A dilatação é usada quando o efeito de aumentar objetos é desejado, preenchendo buracos ou permitindo a conexão de componentes próximos (SZELISKI, 2010; GONZALEZ; WOODS, 2018).

Na Figura 14, regiões que possuem pixels brancos aumentaram enquanto regiões que possuem elementos pretos diminuíram.

2.3.2.2 Abertura e Fechamento

Como pode ser observado nas Figuras 13 e 14, as imagens processadas pelas operações morfológicas básicas não mantém o mesmo tamanho. A partir de propriedades da erosão e da dilatação, novos conjuntos de operações que preservam as características de forma e tamanho da imagem foram derivados: a *abertura* e o *fechamento*.

Figura 14 – Uma iteração da dilatação por um elemento estruturante 9x9 todo preto.

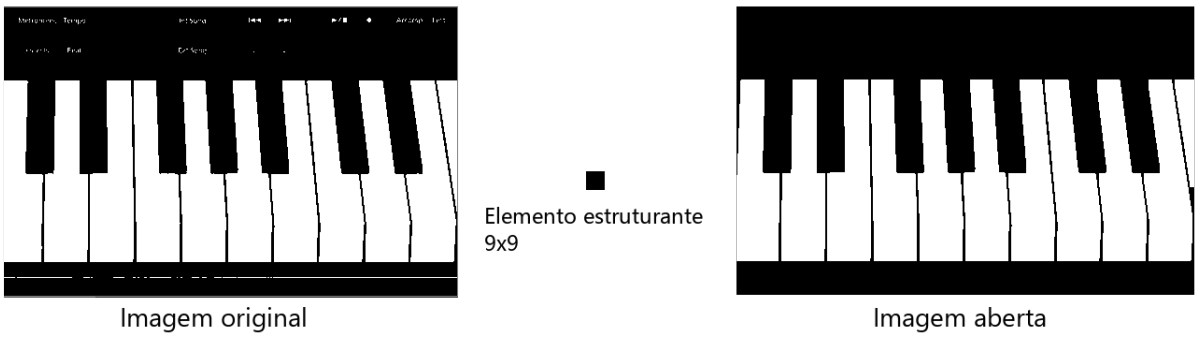


Fonte: Autor

A operação de *abertura* (Figura 15) tem como objetivo eliminar componentes indesejáveis de uma imagem, suavizando o contorno (GONZALEZ; WOODS, 2018; SZELISKI, 2010). Ela consiste em erodir e depois dilatar o resultado da erosão. Um conjunto aberto é mais regular e menos rico em detalhes que o conjunto original.

Os principais efeitos da abertura são: nivelação dos contornos pelo interior, separação de partículas e eliminação de partículas inferiores em tamanho em relação ao elemento estruturante (FACON, 2002; SZELISKI, 2010).

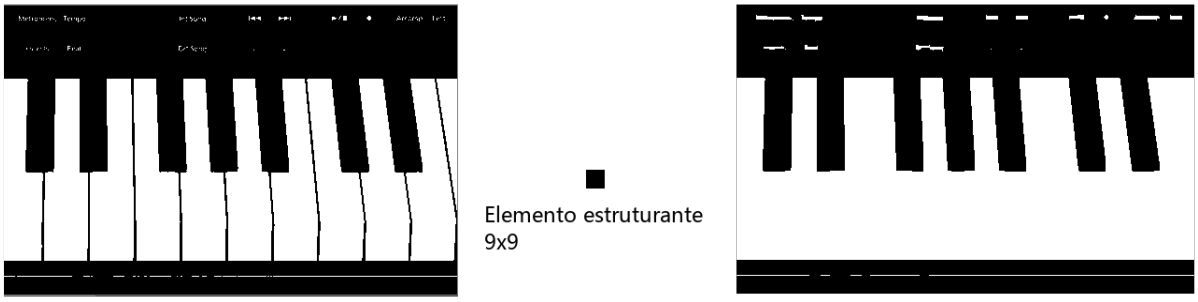
Figura 15 – Abertura.



Fonte: Autor

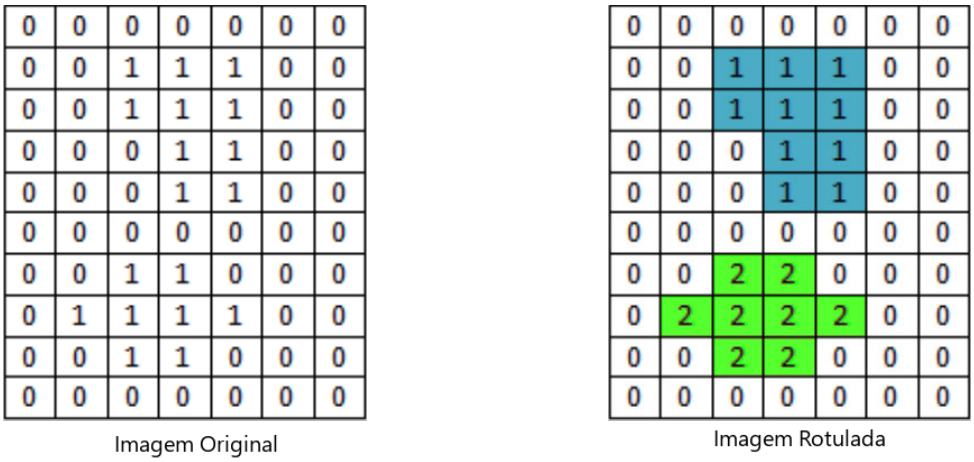
O *fechamento* (Figura 16) é uma operação que consiste em dilatar um conjunto e depois erodir o resultado da dilatação. O processo suaviza as fronteiras pelo exterior e o conjunto fechado é mais regular mas menos rico em detalhes que o conjunto original (SZELISKI, 2010). Aplica-se o fechamento quando se quer preencher buracos inferiores em tamanho em relação ao elemento estruturante ou emendar conexões próximas sem modificar radicalmente o tamanho e a forma dos conjuntos iniciais (GONZALEZ; WOODS, 2018).

Figura 16 – Fechamento.



Fonte: Autor

Figura 17 – Processo de rotulação de componentes conexos.



Fonte: Adaptado de Pedrini e Schwartz (2008)

2.3.2.3 Extração de Componentes Conexos

Dois conceitos são importantes para o entendimento do processo de extração de componentes: *conectividade*, apresentado na seção 2.2.1.2 e *vizinhança*, seção 2.2.1.1. A extração de componentes é uma importante ferramenta de auxílio na identificação de objetos de uma imagem (GONZALEZ; WOODS, 2018; SZELISKI, 2010). O objetivo é identificar e numerar na imagem binária os agrupamentos de pixels conexos, considerando um critério de vizinhança.

Componentes conexos podem ser rotulados de acordo com atributos como tamanho, formato, valor de intensidade e similaridade com outros objetos. O processo de rotulação de componentes conectados atribui etiquetas para cada componente conexo encontrado, possibilitando a identificação e separação dos objetos em diferentes grupos. Um processo de extração de componentes conexos está ilustrado na Figura 17.

2.3.3 Segmentação de Imagens

A segmentação em imagens procura separar o conjunto de dados de entrada em estruturas com conteúdo relevante semântico para a aplicação em questão. Após um processo de segmentação, cada objeto pode ser descrito por meio de suas propriedades geométricas ou topológicas. Atributos como localização, centróide, área, forma e textura podem ser extraídos dos objetos, características de fundamental importância para que as informações resultantes da análise da imagem sejam confiáveis (FACON, 2002).

Em imagens binárias, duas abordagens de segmentação são possíveis baseadas em suas propriedades básicas de valores: descontinuidade e similaridade (FACON, 2002).

As técnicas que envolvem a descontinuidade visam particionar a imagem com base em uma mudança abrupta no valor de intensidade, caracterizando a presença de pontos isolados, linhas ou contornos na imagem (PEDRINI; SCHWARTZ, 2008). Nas seções 2.3.3.1, 2.3.3.2 e 2.3.3.3 são apresentados as técnicas de detecção de descontinuidade utilizadas no trabalho.

Técnicas de similaridade procuram agrupar pontos da imagem que apresentam valores similares para um determinado conjunto de características, produzindo um conjunto de regiões homogêneas (PEDRINI; SCHWARTZ, 2008). Uma técnica comum é a limiarização, ela estabelece um ou mais limiares que separam os pontos da imagem entre pontos de objeto e pontos de fundo. A técnica utilizada neste trabalho para destacar o piano do fundo foi a limiarização global de Otsu, descrita na seção 2.3.3.4.

2.3.3.1 Detecção de bordas Canny

A detecção de bordas é, essencialmente, a operação de identificação de mudanças locais significativas nos níveis de cinza da imagem.

John F. Canny em 1986 desenvolveu um algorítmo de detecção de bordas de objetos dentro de uma imagem utilizando gradientes. Gradientes são vetores que indicam os locais nos quais os níveis de cinza sofrem maior variação (PEDRINI; SCHWARTZ, 2008). A entrada do algorítmo recebe uma imagem em tons de cinza e sua saída é uma outra imagem que possui a posição das descontinuidades detectadas.

Canny (1986) estabeleceu três critérios principais para a detecção de bordas. A primeira é que a imagem deve ter uma qualidade mediana. Assim, o operador tem uma baixa probabilidade de que não consiga detectar bordas ou pior, marque bordas falsas por conta de ruídos na imagem. A boa localização foi o segundo critério estabelecido. Segundo ele, as bordas marcadas devem estar o mais próximo possível da borda verdadeira. Por fim, o operador deve apresentar apenas uma resposta para cada borda, marcando as que forem localizadas na imagem apenas uma vez.

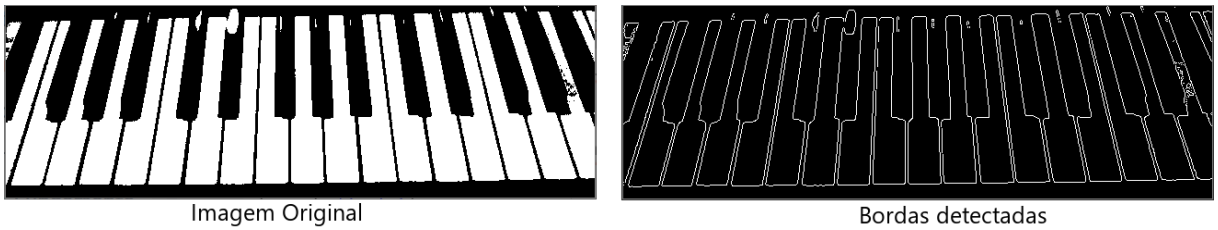
Por conta desses critérios o operador funciona em vários processos, o primeiro

estágio melhora a qualidade da imagem através da redução de ruídos, suavizando a imagem com a filtragem Gaussiana (2.3.1.2). Gradientes que tem grandes índices de intensidade tem mais probabilidade de formarem bordas, por isso, em seguida, a magnitude e a direção dos gradientes da imagem são calculadas. A borda é localizada utilizando apenas os pontos cuja magnitude seja localmente máxima na direção do gradiente. Essa operação é chamada de *supressão não-máxima* e reduz a espessura das bordas.

Para evitar que a borda fique fragmentada em múltiplos segmentos, utiliza-se limiares, T1 e T2, com T2 > T1, durante a etapa de *supressão não-máxima*. Essa operação é conhecida como *limiarização com histeresse*. Pontos da borda que possuem gradiente maior que T2 são mantidos como pontos de borda. Qualquer outro ponto conectado a esses pontos da borda é considerado como pertencente à borda somente se a magnitude do seu gradiente estiver acima de T1. A escolha dos limiares é feita com base em uma estimativa da relação sinal-ruído (PEDRINI; SCHWARTZ, 2008).

A Figura 18 ilustra em sequência a aplicação do filtro de Canny em uma imagem com limiares T1:180 e T2: 255.

Figura 18 – Detecção de bordas com Canny.



Fonte: Autor

2.3.3.2 Detecção de linhas com Hough

A transformada de Hough foi desenvolvida por Paul Hough em 1962 e detecta a presença de grupos de pontos colineares ou quase colineares em imagens digitais sem nenhum conhecimento prévio sobre tais características (FISHER et al., 1997). Originalmente desenvolvida para detectar linhas e curvas em imagens digitais, teve seu método expandido para a detecção de outras formas geométricas (círculos, elipses, entre outras) (GONZALEZ; WOODS, 2018).

Segundo Low (1991), a transformada de Hough prevê que a imagem a ser analisada possua fronteiras bem definidas, ou seja, necessita que os pontos seja previamente definidos através de uma operação de reconhecimento de bordas, como a Canny (2.3.3.1). Imagens binárias são ideais para a aplicação desta técnica por possuírem níveis de contraste bem definidos.

A detecção de linhas com Hough é uma ferramenta com grande eficiência computacional e não necessita que o modelo a ser detectado apresente continuidade, ou seja, as linhas podem estar segmentadas ou fracionadas. A solução consiste em encontrar segmentos de retas concorrentes, analisando todas as possibilidades de existência de segmentos de retas que cortam cada ponto da imagem. Por conta disso, uma filtragem de linhas é necessária quando a técnica é utilizada.

A Figura 19 ilustra o resultado da aplicação da detecção de Hough na versão não binarizada da Figura 18.

Figura 19 – Detecção de linhas com Hough.

Fonte: Autor

2.3.3.3 Contornos de Suzuki

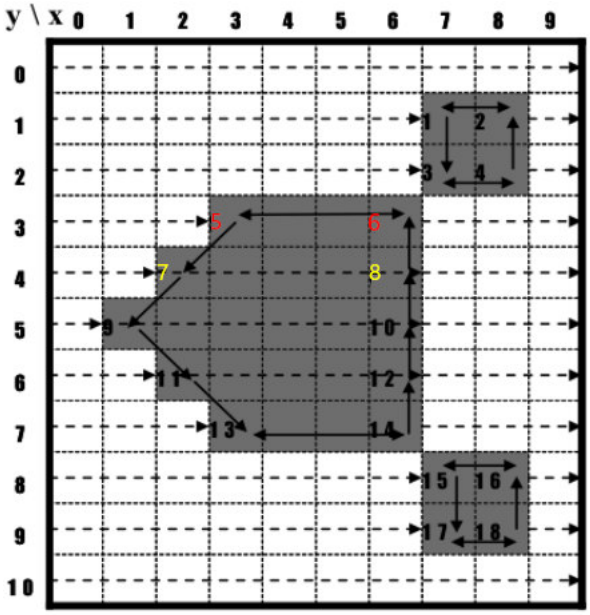
Em imagens digitais, contornos são definidos como pixels onde o brilho muda abruptamente, essas mudanças denotam locais na imagem que correspondem aos limites de objetos. Eles são usados para separar objetos do fundo, calcular tamanho, classificar formas e outras características importantes usando o comprimento e a forma dos pixels do contorno (SONS, 2007).

O algorítmo de Contornos de Suzuki utiliza o paradigma de varredura completa de imagens. Levando em consideração que objetos podem possuir dois tipos de limites, os limites de borda na direita e limites de borda na esquerda, a varredura na imagem se dá de forma horizontal, linha por linha e tem orientação esquerda-direita. Objetos também pode ter limites de contorno, os interiores e os exteriores.

A varredura de Suzuki faz a análise de duas linhas ao mesmo tempo, achando os limites de bordas dos objetos, analisando-os e estabelecendo relacionamento entre as bordas encontradas (SUZUKI; BE, 1985).

Para ilustrar o funcionamento, na Figura 20 a leitura da linha 3 detecta (borda esquerda 5, borda direita 6) e a leitura da linha 4 detecta (borda esquerda 7, borda direita 8). Como a leitura é feita em pares, sequencialmente, os relacionamentos entre (5,7) e (8 , 6) são gerados.

Figura 20 – Varredura de Suzuki.



Fonte: Autor

2.3.3.4 Limiarização Global de Otsu

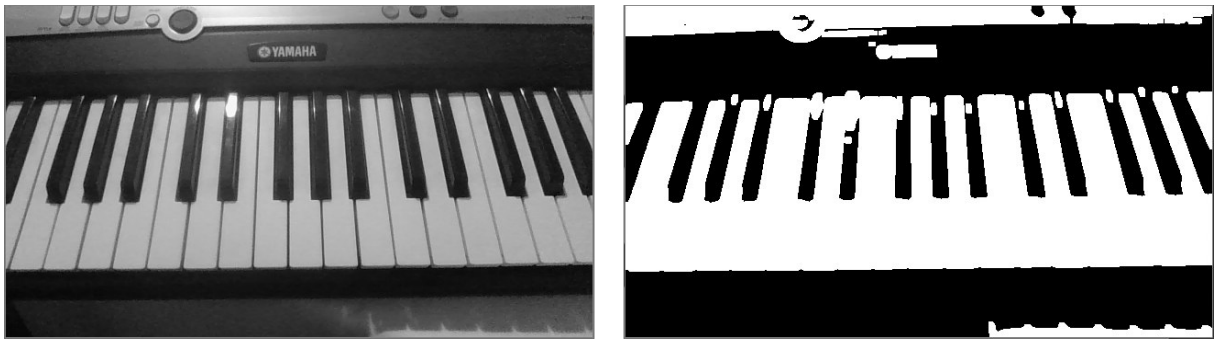
A limiarização é uma das técnicas mais simples de segmentação e consiste na classificação dos pixels de uma imagem entre pixels de objeto e de fundo de acordo com a especificação de um ou mais limiares (PEDRINI; SCHWARTZ, 2008).

A maneira mais direta de selecionar um valor para o limiar é a partir da distribuição de intensidades dos pixels na imagem, o que nem sempre é o ideal, já que podem existir imagens em que a intensidade dos objetos e do fundo não são bem distintas em virtude, por exemplo, da ocorrência de baixo contraste ou ruído (PEDRINI; SCHWARTZ, 2008). Métodos foram desenvolvidos tentando contornar esse problema, determinando o limiar por meio da otimização de certas medidas de separação entre as classes de objetos na imagem.

O método de Otsu (1979) é baseado no fato de que do histograma de uma imagem pode-se identificar duas classes, os pixels de objeto e pixels de fundo, e que elas possuem características de média e desvio padrão (GONZALEZ; WOODS, 2018). O método então, divide a imagem aumentando a diferença de intensidade dos pixels das classes diferentes, maximizando a variância entre classes distintas e diminui a diferença de intensidade de pixels que pertencem a mesma classe, minimizando a variância interna deles (PEDRINI; SCHWARTZ, 2008).

A Figura 21 ilustra a sequência de uma aplicação da limiarização global por Otsu, onde a região do objeto que contém as teclas é aumentada..

Figura 21 – Limiarização de Otsu.

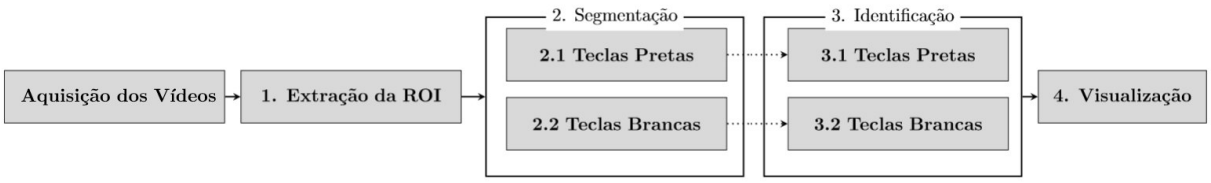


Fonte: Autor

3 Metodologia

Esta seção define os procedimentos da metodologia proposta por este trabalho para a segmentação de teclado e rotulamento de teclas de pianos eletrônicos presentes em vídeo. A Figura 22 representa as etapas da metodologia, sendo estas: Extração da ROI, Segmentação, Identificação e Visualização.

Figura 22 – Etapas gerais da metodologia proposta



Fonte: Autor

3.1 Aquisição dos Vídeos

A metodologia utiliza filmagens reais de cenas que contenham um piano eletrônico. Todo o teclado do piano eletrônico deve estar enquadrado durante a filmagem, de modo que todas as divisões entre as teclas brancas devem estar visíveis. Para este fim, sugere-se utilizar câmeras que gravem em, no mínimo, resolução de 720. Sugere-se também utilizar pedestais estabilizadores para não apresentar alterações de ângulo de filmagem. O esquema de aquisição dos vídeos está ilustrado na Figura 23.

3.2 Extração da ROI

O primeiro passo compreende a localização do teclado de um piano eletrônico em um vídeo, que é a entrada. A etapa pode ser divida em duas partes, explicadas a seguir.

3.2.1 Pré-processamento

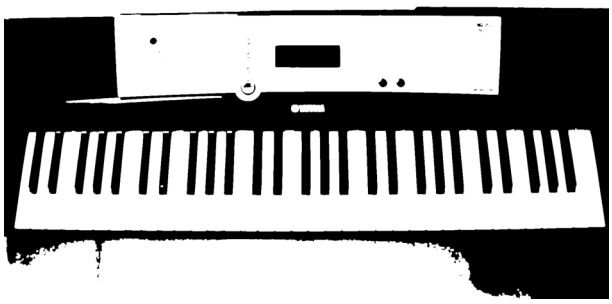
A fase de pré-processamento da extração da ROI tem como objetivo melhorar a qualidade do frame adquirido e achar uma melhor representação da imagem. Uma filtragem Gaussiana (seção 2.3.1.2) é aplicada, fazendo que os pixels apresentem mais uniformidade de cor. Pela estrutura do teclado de um piano eletrônico pode-se perceber que as teclas são pretas ou brancas, sendo desnecessária a representação colorida da imagem. Uma imagem que tem suas cores binarizadas tem uma representação mais adequada, para isso será

Figura 23 – Estrutura da aquisição do vídeo.

Fonte: Autor

Figura 24 – Etapa de Pré-processamento

(a) Frame original



(b) Frame pré-processado

Fonte: Autor

necessária a escolha de um limiar, que foi feita automaticamente pela limiarização de Otsu (seção 2.3.3.4).

A região do teclado de um piano é majoritariamente composta por teclas brancas, atributo que fica mais evidenciado depois a aplicação de uma operação morfológica de dilatação. A Figura 24 demonstra o frame antes e depois da etapa de pré-processamento.

3.2.2 Extração de contornos

Uma outra característica da região das teclas de um piano fundamental é seu formato retangular. Objetos com essa característica podem ser procurados na imagem a

Figura 25 – Contornos de Suzuki.

(a) Todos os contornos encontrados (b) Contornos filtrados

Fonte: Autor

partir do uso de contornos de Suzuki (seção 2.3.3.3). O resultado da busca é ilustrado na Figura 25(a), onde são identificados todos os contornos de objetos no frame. Dos candidatos a ter a ROI encontrados, alguns são ruídos e outros são objetos de fundo, esses podem ser ignorados com a utilização de dois limiares estabelecidos através de observações empíricas, um que descarta objetos que possuem menos de 5% de área do frame original e os que tem mais de 80%. O resultado dessa filtragem inicial é ilustrado na Figura 25(b).

Com os contornos restantes é feita uma análise do conteúdo das segmentações a partir do histograma, ferramenta útil na busca de similaridade entre imagens. Uma comparação de cada um dos histogramas dos contornos candidatos a ROI com um *template* é feita. O método de comparação de histogramas utilizado foi o de correlação, onde o contorno que teve seu histograma com a maior nota no índice foi escolhido como contendo a ROI. Um corte baseado na posição inicial do contorno que contém a ROI e com as dimensões dele é realizado, o template utilizado e o resultado dessa etapa é ilustrado na Figura 26.

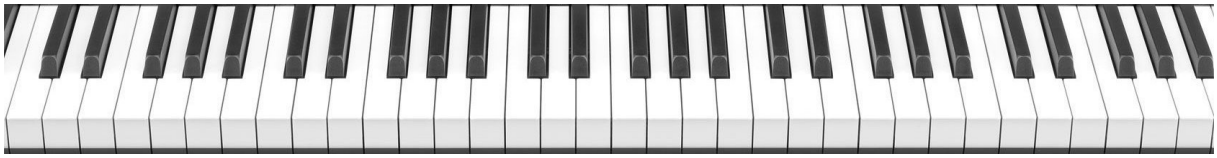
3.3 Segmentação

O segundo passo consiste na segmentação das teclas e a extração de componentes conexos.

Com a ROI localizada, uma detecção de linhas horizontais por Hough foi feita para dividir a imagem em duas sub-regiões: uma sub-região superior que possui a sua maior parte ocupada por teclas pretas e uma sub-região inferior que tem sua área ocupada por teclas brancas (Figura 28).

Para o uso de Hough, no entanto, uma delimitação de bordas é necessária. O algorítmo de detecção de bordas aplicado foi o Canny (seção 2.3.3.1). Uma das condições para o seu bom funcionamento é ter uma imagem de entrada com qualidade mediana. A

Figura 26 – Resultado da comparação de histogramas.



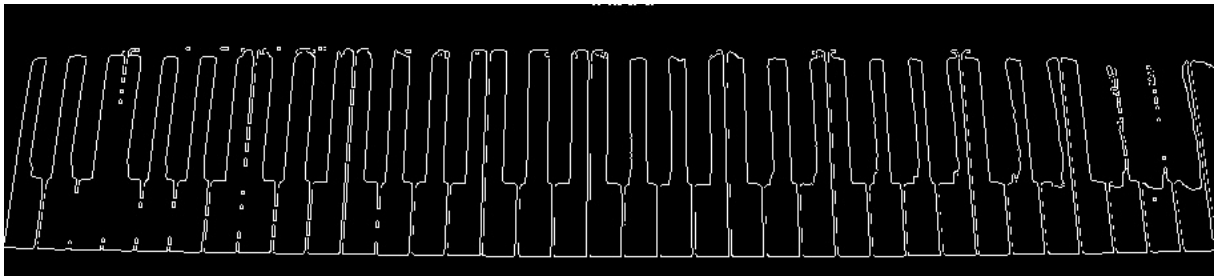
Template utilizado para comparação de histograma

ROI encontrada

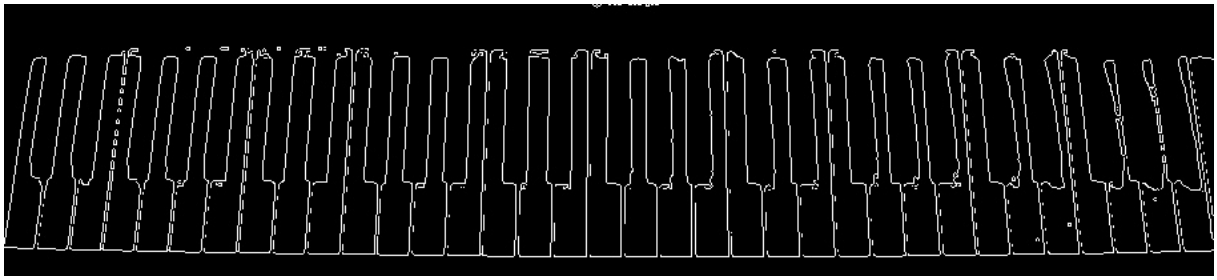
Fonte: Autor

equalização de histograma com CLAHE foi aplicada, normalizando a iluminação da ROI e aumentando as chances das bordas serem bem definidas pelo filtro. Pode ser notada na Figura 27 a diferença de uso da detecção de bordas pelo filtro de Canny no frame original e no frame com sua iluminação normalizada.

Figura 27 – Aplicação do filtro de Canny



Canny no frame original



Canny no frame normalizado.

Fonte: Autor

Figura 28 – Resultado da detecção de linhas horizontais por Hough.

Fonte: Autor

3.3.1 Extração de componentes conexos

Com os limites de cada tecla bem definidos pelo filtro de Canny é possível identificar cada uma utilizando o algoritmo de componentes conexos.

Como já foi mencionado, as teclas pretas estão localizadas na sub-região superior delimitada pelas linhas de Hough (Figura 28), para aumentar a área de cada uma delas uma dilatação é aplicada. Já a sub-região inferior que contém as teclas brancas tem a divisão entre as teclas mais evidenciada após a aplicação de uma erosão. Após esse processo, as delimitações foram utilizadas na busca dos componentes conexos, procurando os componentes conexos das teclas pretas na sub-região superior e os componentes conexos das teclas brancas na sub-região inferior.

A busca de componentes conexos retorna as informações espaciais (posição x, posição y, altura, largura, centróide e área) de cada objeto na imagem, uma filtragem baseada na área e na distância relativa entre os centróides das teclas é feita. O resultado da extração de componentes conexos dos dois tipos de teclas é ilustrado na Figura 29.

Figura 29 – Extração de componentes conexos da ROI.

Componentes conexos das teclas pretas

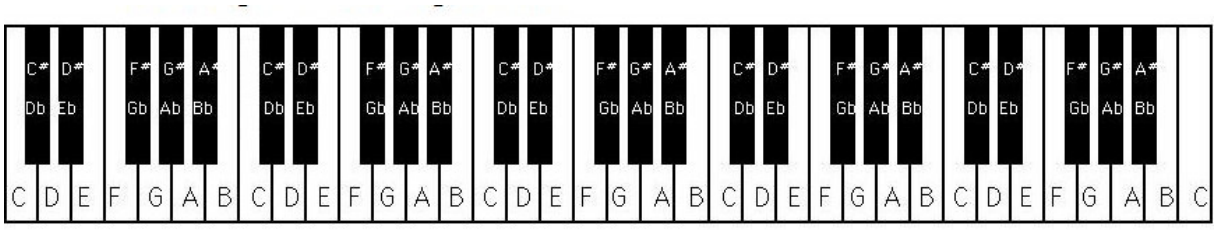
Componentes conexos das teclas brancas

Fonte: Autor

3.4 Identificação

Com base em estudos da disposição das teclas em um teclado de um piano em uma oitava, uma observação empírica é feita: entre duas teclas pretas sempre existe, no mínimo, uma tecla branca. As exceções são entre as teclas que correspondem às notas A#/Bb e C#/Db, ocupada por duas teclas brancas B e C, e entre as notas D#/Eb e F#/Gb, ocupada por duas teclas teclas brancas E e F, como pode ser visto na Figura 30.

Figura 30 – Teclado moderno de 61 teclas com os valores de cada tecla.



Fonte: Adaptado de Yamaha Keyboard Guide (2019)

Para achar as exceções, foi feito um cálculo da distância média entre as teclas pretas, baseando-se nos centróides de cada uma delas. Então, fazendo uma varredura da esquerda pra direita, se a distância entre a tecla preta atual e a tecla preta próxima for maior que a distância média, então duas teclas brancas estão entre essas teclas pretas e uma linha vertical no ponto médio (equação 3.1) delas é adicionada para segregá-las. O processo é repetido até o fim do teclado.

$$pontoMédio = (\frac{cPa_x + cPa_y}{2}, \frac{cPp_x + cPp_Y}{2}) \tag{3.1}$$

Onde:

cPa é o centróide da tecla Preta atual;

cPa é o centróide da tecla Preta próxima;

Concorrente a esse processo de busca, uma representação da disposição das teclas foi criada: a cada tecla preta encontrada é escrito um valor "Pb", indicando uma tecla preta seguido de uma tecla branca e a cada exceção encontrada é escrito um valor "b", referente a uma tecla branca. O resultado é a disposição das oitavas no teclado encontrado, ilustrado na Figura 31.

Como a tecla Dó segue um padrão durante todo o teclado de um piano é possível encontrar a quantidade de Dós presentes no teclado para inferir a posição do Dó Médio (seção 2.1.1.2) através da busca pelo padrão "bbPbPbb"na representação da disposição das teclas. A Figura 32 ilustra o padrão utilizado e a localização do Dó Médio com o uso do Algorítmo 1.

Figura 31 – Resultado da busca ao longo do teclado por exceções.

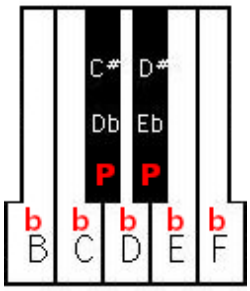
Linha vertical inferida que divide duas teclas brancas

```
Padrão do Teclado = bPbPbbPbPbPbbPbPbbPbPbbPbPbbPbPbbPbPbPbbPbPbbPbPbbPbPbPbbPbPbbPbPbPbb
```

Disposição de teclas encontrada

Fonte: Autor

Figura 32 – Resultado da busca ao longo do teclado por exceções.



Padrão utilizado para localizar as teclas Dó (C)

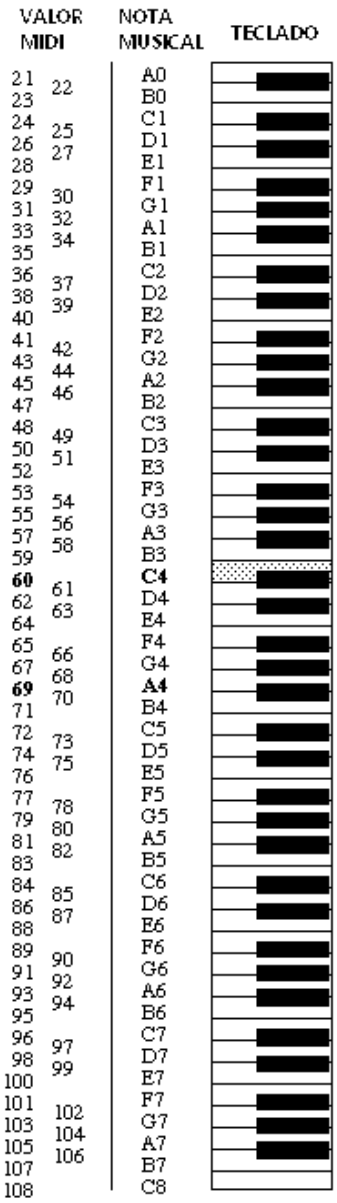
Resultado da busca do Dó Médio.

Fonte: Autor

3.4.1 Associando valores às teclas

O MIDI (Musical Instrument Digital Interface) é um protocolo que estabelece valores à cada nota presente em um teclado de um piano e foi utilizado como referência para a identificação das teclas encontradas. Como pode ser visto na Figura 33, o valor do Dó Médio corresponde ao valor 60. Associando o valor do Dó Médio da tabela MIDI com a nota encontrada é possível fazer duas iterações: uma de incremento a partir dessa posição, associando as teclas encontradas aos valores da tabela até o fim do teclado e outra de decremento, com a mesma finalidade de identificação, só que a partir de uma posição anterior à do Dó Médio até o começo do teclado.

Figura 33 – Tabela MIDI com os valores das notas musicais correspondentes.



Fonte: Adaptado de Joe Wolfe (2019)

3.5 Visualização

O último passo envolve a visualização das associações feitas na etapa anterior.

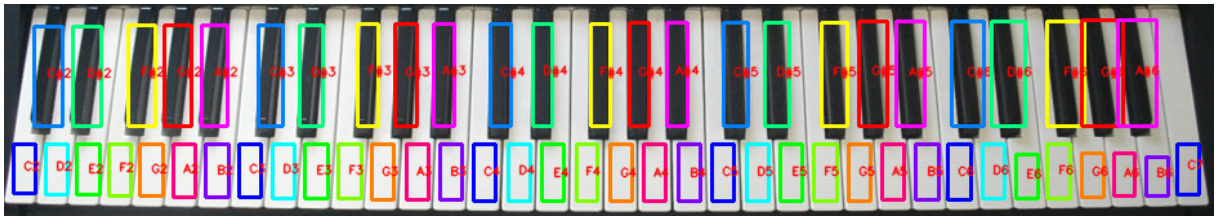
Figura 34 – Associação de cores às notas de uma oitava.

Fonte: Autor

Com a associação das notas do sistema musical MIDI a cada uma das teclas

é possível plotar em vídeo as *bounding boxes* das notas encontradas. Para facilitar a identificação das notas, cores foram designadas à cada uma das 12 presentes em uma oitava (Figura 34).

Figura 35 – Visualização de todas as teclas com as notas correspondentes.



Fonte: Autor

4 Resultados e Discussão

Neste capítulo são apresentados os resultados da implementação da metodologia apresentada.

Todas as técnicas de processamento de imagens e análise de imagens foram implementadas utilizando a biblioteca OpenCV (OPENCV, 2019) em Python em um computador com processador Intel Core i7-7700HQ, 16GB de RAM, GTX 1050ti e sistema operacional Windows.

Os dados utilizados foram vídeos adquiridos através de uma câmera Nikon D3100. Foram filmados, sem a utilização de pedestais estabilizadores, 30 segundos de cada cena, com resolução de 1920x1080 pixels, taxa de amostragem média de 24 quadros por segundo, à uma distância de 120cm e em um ângulo frontal em relação ao teclado do piano eletrônico, enquadrando ele durante toda a filmagem.

Os testes foram realizados em vídeos gravados de pianos eletrônicos de 61 e 76 teclas. Para acelerar o processamento dos pixels, os frames adquiridos sofrem um corte de resolução pela metade em largura e altura, ficando com resolução de 960x540 mas ainda mantendo em evidência a divisão das teclas brancas.

A Figura 36 é a plotagem da detecção das teclas utilizando a metodologia apresentada em um piano eletrônico da marca Yahama, Modelo PSR-E203, que possui 61 teclas e que foi utilizado extensivamente na descrição da metodologia deste trabalho. Uma característica a ser ressaltada desse modelo de piano eletrônico são os detalhes em branco que poderiam afetar a localização da ROI. A utilização da comparação de histograma na seção 3.2.2 foi crucial para evitar que falsas detecções da ROI fossem feitas. Durante todo o processamento dos frames as teclas pretas foram detectadas e rotuladas corretamente. As teclas brancas só foram identificadas erroneamente em poucos frames que tinham alterações na angulação da filmagem, resultando na criação errada da disposição das teclas criada na etapa de Identificação (seção 3.4), crucial para a associação de valores às teclas.

O segundo teste da metodologia foi realizado em um teclado da marca YAHAMA modelo PSR-310 de 61 teclas. A filmagem possuía alta estabilidade, não tendo alterações de ângulo e resultando numa segmentação do teclado (ROI) e identificação das teclas pretas e brancas de acordo com o seu valor na escala musical durante todos os frames processados. A ilustração do processamento de um desses frames pode ser observada na Figura 37.

O último teste foi em um modelo SP76 da KurzWeil, que é um piano moderno de 76 teclas. Ele possui dimensões maiores e necessita que a filmagem tenha uma abertura

Figura 36 – Detecção de teclas em um YAMAHA PSR E203.

Figura 37 – Detecção de teclas em um YAMAHA PSR 310.

maior para cobrir todas as teclas. Por conta disso, as divisões entre as teclas brancas não foram evidenciadas corretamente em todos os frames processados. Como não houve a utilização de barra estabilizadora, alguns frames sofreram alteração de angulação. A Figura 38 ilustra a plotagem da detecção das teclas no SP76 em um certo momento da utilização da metodologia. Na figura, por conta das divisões entre as teclas brancas não serem bem evidenciadas, a extração de componentes conexos das teclas brancas não delimitaram corretamente os objetos pertencentes à essa classe, detectando duas teclas brancas como uma, comprometendo a etapa de Identificação, rotulando falsamente as teclas da metade à direita do Dó Médio (C4) encontrado. No entando, durante todo o processamento dos frames, as teclas pretas foram identificadas corretamente.

Figura 38 – Detecção de teclas em um KURZWEIL SP76.

5 Conclusão

Este trabalho apresentou uma metodologia para a segmentação de teclado de pianos eletrônicos e a identificação das teclas presentes nele através da análise em vídeo. O método proposto tem a intenção de rotular as diferentes teclas presentes em pianos eletrônicos automaticamente através da utilização de técnicas de processamento de imagens, servindo como base para qualquer aplicação de Visão Computacional que utilize o instrumento como objeto de interesse.

Para o teste da metodologia não foi encontrado um banco de dados com vídeos que pudessem ser utilizados, sendo necessária a filmagem de uma cena que tem o objeto de interesse em um ângulo frontal. Foi inicialmente estudada a possibilidade de usar uma WebCam, devido a facilidade de controle em tempo real através do OpenCV, sua conexão direta com o computador através de USB e estabilidade de filmagem. Porém, por ter resolução limitada, as imagens geradas não evidenciavam as divisões entre as teclas brancas, característica que será crucial na Segmentação (seção 3.3).

A filmagem em pianos eletrônicos de 61 teclas, que são menores, mais comuns no mercado e indicados para pessoas que estão aprendendo o instrumento, sempre evidencia a divisão entre as teclas brancas e tem alterações de angulação irrelevantes e, por isso, a metodologia teve melhores resultados, identificando e rotulando as teclas pretas e brancas de acordo com a escala musical corretamente. Em testes do modelo que possui 76 teclas, a metodologia mostrou uma sensibilidade na detecção e rotulamento das teclas brancas quando a divisão entre elas não é bem evidenciada. No entanto, as teclas pretas sempre foram detectadas e tiveram poucas identificações falsas.

Não foram encontrados trabalhos relacionados disponiveis, não sendo possível a comparação da metodologia proposta com outras existentes.

Com esses considerações feitas, propõe-se como trabalhos futuros:

- Criar uma métrica de validação usando a disposição padrão do teclado de pianos de 61 e 76 teclas e compará-los com a representação encontrada.
- Detecção da ROI independente do ângulo em que a região do teclado se encontra no frame e o ajuste de ângulo adequado.
- Estabelecer um outro método de identificação das teclas brancas através da extração de novas características dos frames adquirido. Como a disposição das teclas do teclado seguem um padrão e sempre as teclas pretas são encontradas corretamente pela metodologia, uma possibilidade é a utilização de identificação de padrões de

centróides das teclas pretas encontradas, através da clusterização dos dados com técnicas de Mineração de Dados, para inferir a posição das brancas.

- Teste da metodologia em mais modelos de teclados de 61 e 76 teclas.
- Detecção e registro do pressionamento das teclas.

Referências

ADAMS, F. M.; OSGOOD, C. E. A cross-cultural study of the affective meanings of color. *Journal of cross-cultural psychology*, SAGE PUBLICATIONS LTD St George's House/44 Hatton Garden, London EC1N 8ER, v. 4, n. 2, p. 135–156, 1973. Citado na página 14.

BALLARD, D. H.; BROWN, C. M. *Computer Vision*. 1st. ed. [S.l.]: Prentice Hall Professional Technical Reference, 1982. ISBN 0131653164. Citado na página 18.

Canny, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, n. 6, p. 679–698, Nov 1986. ISSN 1939-3539. Citado na página 29.

DUNNE, E. G. *Pianos and Continued Fractions*. 2019. Disponível em: <https://oeis.org/DUNNE/TEMPERAMENT2.html>. Citado na página 17.

FACON, J. Processamento e análise de imagens. *Curso e Mestrado em Informatica Aplicada. Pontificia Universidade Católica do Paraná.*, p. 128, 2002. Citado 4 vezes nas páginas 18, 20, 27 e 29.

FISHER, R. et al. *71 - Hypermedia Image Processing Reference (HIPR)*. http://homepages.inf.ed.ac.uk/rbf/HIPR2/morops.htm (zuletzt aufgerufen: 18.10.2012): [s.n.], 1997. Citado na página 30.

GONZALEZ, R.; WOODS, R. *Digital Image Processing*. Pearson, 2018. ISBN 9780133356724. Disponível em: <https://books.google.com.br/books?id=0F05vgAACAAJ>. Citado 9 vezes nas páginas 13, 21, 22, 23, 26, 27, 28, 30 e 32.

JAMES, M. *Pattern Recognition*. BSP Professional, 1987. ISBN 9780632018857. Disponível em: <https://books.google.com.br/books?id=OAtzQgAACAAJ>. Citado 2 vezes nas páginas 13 e 25.

Joe Wolfe. *Note names, MIDI numbers and frequencies*. 2019. Disponível em: <https://newt.phys.unsw.edu.au/jw/notes.html>. Citado na página 41.

LOW, A. A. *Introductory Computer Vision and Image Processing*. New York, NY, USA: McGraw-Hill, Inc., 1991. ISBN 0077074033. Citado na página 30.

Make Me Analyst. *Explore your Data: Graphs and shapes of distributions*. 2019. Disponível em: <http://makemeanalyst.com/explore-your-data-graphs-and-shapes-of-distributions/>. Citado na página 24.

MARIA, L. G. F. *Processamento Digital de Imagens*. [S.l.]: INPE, 2000. Citado na página 21.

OPENCV. *Open Source Computer Vision Library*. 2019. Disponível em: <https://opencv.org/>. Citado na página 43.

PEDRINI, H.; SCHWARTZ, W. *Análise de imagens digitais: princípios, algoritmos e aplicações*. CENGAGE - UM LIVRO, 2008. ISBN 9788522105953. Disponível em: <https://books.google.com.br/books?id=13KAPgAACAAJ>. Citado 11 vezes nas páginas 13, 20, 21, 22, 23, 25, 26, 28, 29, 30 e 32.

PIANO WORKSHOP. *How to Find Middle C on a Piano*. 2013. Disponível em: <http://www.pianoworkshop.co.uk/blog/finding-middle-c-on-a-piano/>. Citado na página 17.

PIZER, S. M. et al. Adaptive histogram equalization and its variations. *Comput. Vision Graph. Image Process.*, Academic Press Professional, Inc., San Diego, CA, USA, v. 39, n. 3, p. 355–368, set. 1987. ISSN 0734-189X. Disponível em: <http://dx.doi.org/10.1016/S0734-189X(87)80186-X>. Citado na página 24.

SERRA, J. *Image Analysis and Mathematical Morphology*. Orlando, FL, USA: Academic Press, Inc., 1983. ISBN 0126372403. Citado na página 25.

SMITH, S. W. *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego, CA, USA: California Technical Publishing, 1997. ISBN 0-9660176-3-3. Citado na página 23.

SONS, I. J. W. . *Digital Image Processing, 4th Edition, William K. Pratt, 2007: Digital Image Processing,*. Bukupedia, 2007. (Digital Image Processing,). Disponível em: <https://books.google.com.br/books?id=9Z5iDwAAQBAJ>. Citado na página 31.

SUZUKI, S.; BE, K. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, v. 30, n. 1, p. 32 – 46, 1985. ISSN 0734-189X. Disponível em: <http://www.sciencedirect.com/science/article/pii/0734189X85900167>. Citado na página 31.

SZELISKI, R. *Computer Vision: Algorithms and Applications*. Springer London, 2010. (Texts in Computer Science). ISBN 9781848829350. Disponível em: <https://books.google.com.br/books?id=bXzAlkODwa8C>. Citado 4 vezes nas páginas 21, 26, 27 e 28.

UNSOUND sound. *How Many Keys Are There On A Keyboard Piano?* 2019. Disponível em: <https://sound-unsound.com/how-many-keys-are-there-on-a-keyboard-piano/>. Citado na página 16.

Yamaha Keyboard Guide. *Piano Keyboard Diagram - Layout Of Keys With Notes*. 2019. Disponível em: <https://www.yamaha-keyboard-guide.com/piano-keyboard-diagram.html>. Citado 2 vezes nas páginas 16 e 39.