Musicalização de imagens

Trabalho de Conclusão de Curso

Engenharia da Computação

Luiz Eduardo Moura de Oliveira Filho
Orientador: Prof. Bruno Fernandes

Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação

Luiz Eduardo Moura de Oliveira Filho

Musicalização de imagens

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, dezembro de 2016.

Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação
Projeto de Final de Curso

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 19 de 12 de 2016, às 14:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente **LUIZ EDUARDO MOURA DE OLIVEIRA FILHO**, orientado pelo professor **Bruno José Torres Fernandes**, sob título **Musicalização de Imagens**, a banca composta pelos professores:

Alexandre Magno Andrade Maciel
Bruno José Torres Fernandes

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

☐ Aprovada X Aprovada com Restrições* ☐ Reprovada

e foi-lhe atribuída nota: 8,5 (oito e meio)

*Obrigatório o preenchimento do campo abaixo [illegible] 30 [illegible]

ALEXANDRE MAGNO ANDRADE MACIEL

BRUNO JOSÉ TORRES FERNANDES

* Este documento deverá ser encadernado juntamente com a monografia em versão final.

Dedico esse trabalho à Deus e aos meu Pais que sempre estiveram comigo.

Resumo

Este trabalho apresenta um modelo que permite a criação de músicas através de uma imagem de entrada, resultando em uma música melódica numa determinada escala musical definida pelo usuário. O modelo consiste em determinar os pontos relevantes da imagem, definir a notas desses pontos, gerar a sequência de notas e executar a sequência de notas geradas. Para definir os pontos relevantes da imagem pode-se utilizar métodos que detecte pontos relevantes de uma imagem. A definição das notas se dá pelo menor valor da distância dos pontos relevantes até os extremos do cubo de cores RGB, onde cada extremidade representa uma nota diferente. Na geração da sequência de notas foi utilizado o conceito de máquinas de estados, onde cada estado depende apenas de si para ir a um próximo estado e esses estados são as notas geradas pelo modelo. Com esse modelo proposto, foi possível criar uma aplicação que é capaz de musicalizar uma imagem utilizando o KMeans ou o SURF como método de detecção de pontos relevantes dependendo apenas da definição do usuário. Como resultado, obtemos uma sequência de notas que são executadas tornando possível ouvir o som de uma imagem.

Abstract

This work presents a model that allows the creation of music through an input image, resulting in melodic music in a certain musical scale defined by the user. The model consists of determining the relevant points of the image, defining the notes of those points, generating the note sequence and executing the generated note sequence. To define the relevant points of the image you can use any method that detects relevant points in an image. The definition of notes is given by the smallest distance value of the relevant points to the ends of the RGB color cube, where each extremity represents a different note. Was used the concept of state machines in the generation of the sequence of notes, where each state depends only on itself to go to a next state and these states are the notes generated by the model. With this, proposed model it was possible to create an application that is capable of musicising an image using KMeans or SURF as a method to detect relevant points, depending only on the user definition. As a result we get a sequence of notes that are executed making it possible to hear the sound of an image.

Sumário

Índice de Figuras

Índice de Tabelas

Índice de Equações

Tabela de Símbolos e Siglas

SURF – *Speeded-Up Robust Features*

SIFT - *Scale-Invariant Feature Transform*

1 Introdução

As imagens podem suscitar diferentes sentimentos para as pessoas em múltiplos níveis de intensidade, gerando felicidade, medo, raiva, tristeza e etc.. Essas imagens podem ser geradas a partir de uma máquina fotográfica, câmera no aparelho celular, pintura de um quadro ou um desenho, e todas elas emitem um sentimento diferente para cada pessoa. Os tipos de imagens produzidas influenciam no sentimento emitido como, por exemplo, uma fotografia particular a qual faz-nos lembrar de algum momento importante em nossa vida e o sentimento transmitido não é, necessariamente, causado pelas características da imagem e sim por uma lembrança. Em contrapartida, um quadro pintado ou o desenho de um artista, proporciona uma emoção baseada nas características da imagem, levando em consideração a forma, contorno, conjunto de cores e etc, que em um determinado conjunto representam um sentimento.

Segundo a Fundação Dorina Nowill cerca de 6,5 milhões de pessoas no Brasil possuem algum tipo de deficiência visual, seja a perda total ou parcial da visão. Isso pode impedir que essas pessoas visualizem a beleza de uma imagem, não permitindo, assim, experimentar o sentimento transferido por ela. Porém alguns sentidos são mais aguçados naqueles que têm algum tipo de deficiência, entre eles, uma audição mais apurada. Tendo isso em vista surge então, a ideia de transformar uma imagem em uma música, para que este represente e transfira o sentimento da imagem.

Visando o exposto a cima, a junção de imagem e musicalização de imagens permite que as pessoas que possuem algum tipo de deficiência visual sejam capazes de sentir a emoção de um belo quadro e estes retratem o sentimento através de um som próprio produzido pela imagem.

O Capítulo 2 deste trabalho faz uma revisão bibliográfica acerca dos assuntos abordados como segmentação de imagens, visão computacional, conceitos musicais básicos, extração de características de imagens, os quais serviram de base para fundamentação e criação do modelo proposto no experimento, que poderão ser identificados nos capítulos seguintes. Ao fim serão mostrados as conclusões obtidas com este trabalho e os trabalhos futuros.

2 Fundamentação teórica

Este capítulo irá tratar de toda a fundamentação teórica obtida para a produção deste trabalho.

2.1 Musicalização

2.1.1 Nota musical

O som é um conjunto de ondas mecânicas que se propaga em um meio material, como o ar ou a água. Algumas das características do som mudam de acordo com o meio de propagação, como a velocidade e o comprimento de onda, entretanto a frequência permanece independente e constante durante todo o percurso.

Em uma orquestra, conseguimos identificar os sons de diferentes instrumentos e decifrar de qual deles veio o som. Mas na partitura de cada músico encontramos as mesmas notas. Isso acontece por que as notas escritas na partitura representam uma frequência fundamental que pode ser enriquecida com diversos harmônicos dependendo das características do instrumento. Geralmente, instrumentos de corda apresentam uma vasta gama de harmônicos e, consequentemente, possuem uma onda mais complexa. Os instrumentos com menor número de harmônicos são os de percussão e alguns metais, sendo a flauta doce um dos instrumentos com a sonoridade mais pura entre todos.

Quando a corda de um violão é tocada com uma certa frequência, se a frequência estiver na faixa de 20 a 20.000 Hz, o ouvido humano será capaz de vibrar à mesma proporção, captando essa informação e produzindo sensações neurais, às quais o ser humano dá o nome de som. As ondas com frequência baixa, entre 20 e 100 Hz, por exemplo, soam em nossos ouvidos de forma grave, e sons com frequência elevada (acima de 400 Hz) soam de forma aguda. Nesta situação, podemos imaginar que apenas uma onda percorre a corda e assim obtemos a frequência ouvida. Essa analogia funciona para uma situação ideal, entretanto o que encontramos na prática é uma corda vibrando de forma muito mais complexa, pois um conjunto de curvas senoidais originadas por diversos

fatores como a posição em que tocamos, a densidade da corda e o tamanho da caixa do violão, somam-se e, juntas, geram o som que ouvimos.

Apesar de diferentes, quando os instrumentos reproduzem uma mesma nota, o som é agradável. Isso acontece por que todas as componentes dessa melodia são compostas por múltiplos inteiros da frequência original. Para entender melhor esse conceito podemos analisar a Figura 1, que relaciona as notas e suas respectivas frequências: A nota dó (C) corresponde a frequência de 261,63 Hz, enquanto a nota sol (G) é aproximadamente 392 Hz, que é 3/2 a frequência do dó. Podemos então caracterizar a nota sol como sendo uma harmônica de uma das harmônicas de dó. Mesmo que o som de uma flauta seja diferente do som do violão, quando reproduzem a mesma nota, as frequências que ouvimos são sempre múltiplas inteiras umas das outras, resultando em uma interação harmônica. Deve-se observar, no entanto, que existem instrumentos transpositores, e as notas da partitura do músico devem ser adaptadas para manter a harmonia, resultando que a frequência das notas deve ser observada em uma nova escala previamente descrita.

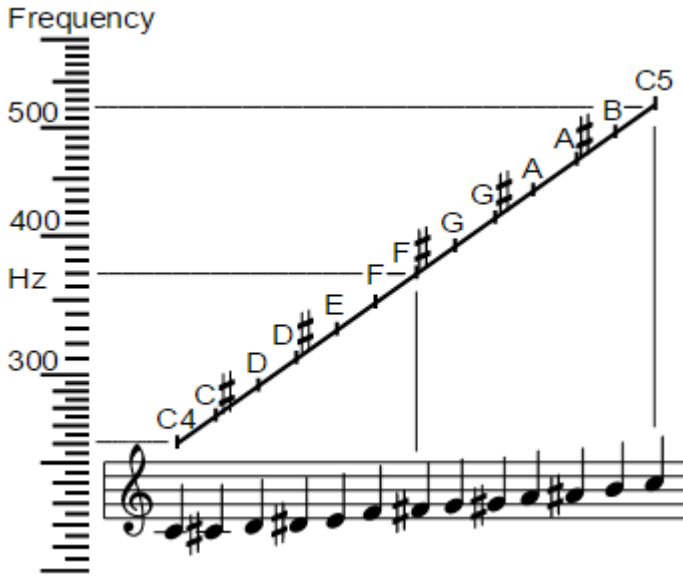


Figura 1 Frequência das notas

Outro fator importante para diferenciar os sons é a amplitude de cada harmônica, ou seja, sua intensidade. Existem instrumentos em que algumas harmônicas têm amplitudes superiores a própria frequência fundamental, tornando o som bem mais grave ou agudo. Com os avanços tecnológicos das

últimas décadas, músicos tem desenvolvido novas técnicas para amplificar harmônicos, como é o caso da guitarra, que pode contar com pedais para distorcer o som.

2.1.2 Escala musical

A partir da descoberta de artefatos musicais da antiguidade, supõe-se que a primeira escala desenvolvida tenha sido a escala de cinco sons ou pentatônica, o que é confirmado pelo estudo de sociedades antigas encontradas contemporaneamente. Observando-se, no entanto, que a palavra "pentatônica" é, na verdade, substituída no vocabulário musical, pela palavra "pentafônica", uma vez que a primeira (pentatônica), remete à ideia de cinco notas tônicas em uma mesma escala ou tonalidade sonora musical, o que não é a verdade; e a segunda (pentafônica) refere-se, mais claramente, à escala ou tonalidade formada por cinco sons ou notas diferentes. No entanto, o termo "pentatônica" ainda é muito mais utilizado popularmente do que o termo "pentafônica".

As escalas de 7 notas foram prováveis desenvolvimentos da escala pentatônica e tem-se o registro de sua utilização pelos gregos, apesar de que qualquer tentativa de resgate da sonoridade dessas escalas tratar-se-á de exercício puramente especulativo.

2.1.3 Escala Maior

Em música, escala maior é uma escala diatônica de sete notas em modo maior, um dos modos musicais utilizados atualmente na música tonal. A sequência de tons e semitons dessa escala obedece à seguinte ordem: Tom-Tom-Semitom-Tom- Tom –Semitom.

A partir da escala maior é que são formados os acordes maiores. A escala fundamental do modo maior é a escala de Dó maior, uma vez que a relação de intervalos desse modo pode ser obtida nesta escala sem a necessidade de nenhuma alteração de altura. A Figura 2 mostra as notas dessa escala e sua sequência de intervalos na sequência de intervalos:

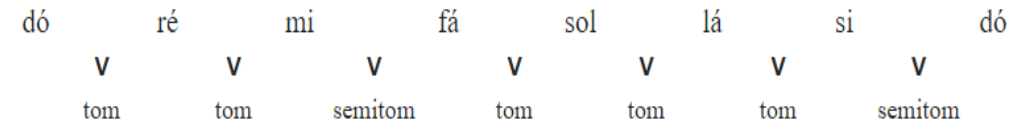


Figure 2 Escala maior

Para formar escalas maiores iniciadas por outra nota é necessário acrescentar alterações de altura a algumas notas, a fim de manter a mesma sequência de intervalos. Em uma escala de sol maior, por exemplo, o intervalo das notas serão conforme a Figura 3 mostra.

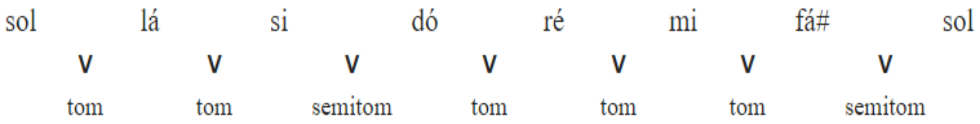


Figure 3 Escala sol maior

A nota fá não pode ser utilizada nesta sequência pois o intervalo entre mi e fá é de um semitom e entre fá e sol é de um tom. Para que a escala obedeça à ordem dos intervalos é preciso aumentar a nota fá em meio tom e torná-la um fá sustenido (fá#). Em outras escalas, para manter a relação de intervalos, é necessário reduzir a altura de algumas notas em meio tom (bemol). O ciclo das quintas define a ordem em que os sustenidos ou bemois são adicionados às escalas.

2.2 Processamento de imagens

O interesse em métodos de processamentos de imagens digitais, segundo Gonzalez [4], decorre em duas principais áreas: melhoria de informação visual para interpretação humana e o processamento de dados de cenas para percepção automática através de máquinas. Ao longo do tempo foi se aperfeiçoando os métodos de processamento em diversas áreas, como imagens médicas e aplicação industrial. Para acontecer este processamento se tem algumas etapas que podem ser seguidos.

A primeira etapa é a aquisição da imagem[4], em que consiste em adquirir uma imagem utilizando um sensor para capturar a imagem e a capacidade de digitalizar o sinal produzido pelo sensor como uma câmera fotográfica. Logo

após à obtenção da imagem uma função de pré-processamento é executada, a imagem original é melhorada a fim de aumentar as chances de sucesso do processamento, algumas técnicas de realce de contrates, remoção de ruídos ou isolamento de região podem ser utilizadas. A próxima etapa é a segmentação, uma imagem é dividida em partes que são entradas ou objetos constituintes. Em geral a segmentação é uma das tarefas mais difíceis no processamento de imagens.

A etapa seguinte é um conjunto de dados e representações que é apenas uma forma de representar a imagem em forma de dados. E por último temos o reconhecimento e interpretação dos dados obtidos no processamento. O reconhecimento é o processo de atribuir um rótulo a um objeto, baseado no seu descritor. Já a interpretação envolve a atribuição de significado ao objeto reconhecido.

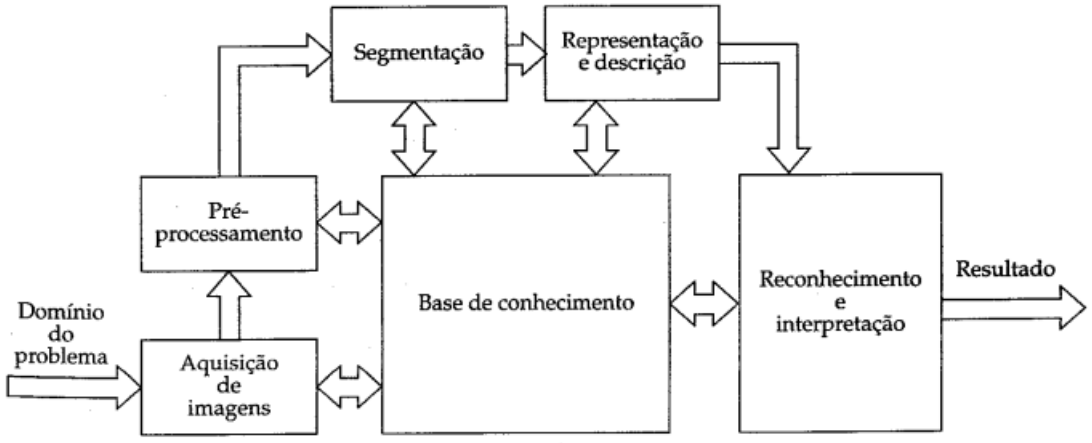


Figura 4 Fluxo do processamento de imagens

2.2.1 Segmentação

Em visão computacional, a segmentação de imagens é o processo de dividir uma imagem digital em múltiplos segmentos, objetivando simplificar ou mudar a representação de uma imagem em algo que é mais significativo e mais fácil de analisar [1] [8] A segmentação de imagens é tipicamente usada para localizar objetos e limites (linhas, curvas, etc.) em imagens. Mais precisamente, a segmentação de imagem é o processo de atribuir um rótulo a cada pixel em uma imagem de tal forma que os pixels com o mesmo rótulo compartilhem certas características.

O resultado da segmentação de imagem é um conjunto de segmentos que coletivamente cobrem toda a imagem, ou um conjunto de contornos extraídos da imagem. Cada um dos pixels em uma região é semelhante em relação a alguma característica ou propriedade calculada, como cor, intensidade ou textura. As regiões adjacentes são significativamente diferentes em relação à mesma característica. [7] e quando aplicado a uma pilha de imagens, comumente utilizada em imagens médicas, os contornos resultantes após a segmentação de imagem podem ser usados para criar reconstruções em 3D com a ajuda de algoritmos de interpolação.

2.2.2 Clustering

Clustering é uma técnica de mineração de dados para realizar agrupamentos automáticos de dados segundo seu grau de semelhança. O critério de semelhança faz parte da definição do problema e, dependendo, do algoritmo pode ser alternado. Uma das técnicas existentes de *Clustering* é o KMeans.

O *K-means* é uma técnica utilizada para particionar uma imagem em *K* grupos. [2] O algoritmo identifica um centro para um grupo, aleatoriamente ou com base em algum método heurístico e atribui cada pixel da imagem ao grupo que resulta na menor distância entre o pixel e o centro do *cluster* recalculando os centros do grupo pela média de todos os pixels, repetindo essas ações até que a convergência seja atingida, ou seja, nenhum pixel altera o grupo. No caso, a distância, normalmente, é a diferença quadrática ou absoluta entre um pixel e um centro e ela é tipicamente baseada na cor, intensidade, textura e localização do pixel, ou uma combinação ponderada desses fatores. Este algoritmo é garantido para convergir, mas pode não retornar uma solução ideal e a qualidade da mesma depende do conjunto inicial do grupo e do valor de K.

2.2.3 Extração de características

Em visão computacional e processamento digital de imagens, uma característica é uma informação que é relevante para uma determinada área de uma imagem e tem o mesmo sentido da característica de aprendizagem de uma máquina e reconhecimento de padrão, embora o processamento de imagens

tenha uma coleção de características muito sofisticada, as características podem ser estruturas específicas na imagem, como pontos, bordas ou objetos, elas também podem ser o resultado de uma operação de vizinhança geral ou detecção de recurso aplicada à imagem tudo que possa ser utilizado para caracterizar uma imagem.

O conceito de característica é muito geral e a escolha de recursos em um sistema de visão computacional pode ser altamente particular dependendo do problema específico à mão.

Quando as características são definidas em termos de operações locais de vizinhança aplicadas a uma imagem, um procedimento comumente referido como extração de características, pode-se distinguir entre abordagens de detecção de características que produzem decisões locais se existe uma característica de um determinado tipo em um dado ponto de imagem, e aqueles que produzem dados não-binários como resultado. A distinção torna-se relevante quando as características detectadas resultantes são relativamente escassas embora sejam tomadas decisões locais, a saída de uma etapa de detecção de características não precisa ser uma imagem binária o resultado é muitas vezes representado em termos conjuntos de coordenadas (conectadas ou não) dos pontos de imagem onde as características foram detectadas, às vezes com precisão de subpixel.

Quando a extração de característica é feita sem tomada de decisão local, o resultado é muitas vezes referido como uma imagem de característica. Consequentemente, uma característica da imagem pode ser vista como uma função de variáveis espaciais (ou temporais) da imagem original, mas onde os valores de pixel mantêm informações sobre as características da imagem em vez de intensidade ou cor. Isto significa que uma característica da imagem pode ser processada de forma semelhante a uma imagem normal gerada por um sensor de imagem. Estas características também são frequentemente computadas como passo integrado em algoritmos para detecção [5].

Em *machine learning*, reconhecimento de padrões e no processamento de imagens, a extração de características começa a partir de um conjunto inicial de dados medidos criando valores derivados (características) destinados a ser

informativos e não redundantes, facilitando as etapas de aprendizado e generalização subsequentes e, em alguns casos, interpretações humanas. A extração de características está relacionada à redução da dimensionalidade.

Quando os dados de entrada para um algoritmo são muito grandes para serem processados e se suspeita serem redundantes (por exemplo, a mesma medição em ambos os pés e metros, ou a repetitividade das imagens apresentadas como pixels), então ele pode ser transformado em um conjunto reduzido de características (também chamado de vetor de características). Esse processo é chamado de seleção de características. Espera-se que os recursos selecionados contenham as informações relevantes a partir dos dados de entrada, de modo que a tarefa desejada possa ser executada usando esta representação reduzida em vez dos dados iniciais completos.

A extração de características envolve a redução da quantidade de recursos necessários para descrever um grande conjunto de dados e ao realizar a análise de dados complexos, um dos principais problemas decorre do número de variáveis envolvidas. A análise com um grande número de variáveis geralmente requer uma grande quantidade de memória e poder computacional, também pode causar uma de classificação no treinamento de amostras e generalizar mal as novas amostras. Extração de características é um termo usado para métodos de construção de combinações das variáveis para contornar esses problemas, enquanto ainda descreve os dados com uma certa precisão.

No processamento de imagens, o conceito de detecção de características refere-se a métodos que visam computar abstrações de informação da imagem e tomar decisões locais em cada ponto de imagem. Se existir uma característica de um determinado tipo nesse ponto as características resultantes serão subconjuntos do domínio da imagem, muitas vezes sob a forma de pontos isolados, curvas contínuas ou regiões conectadas.

Não existe uma definição universal ou exata do que constitui um recurso e muitas vezes depende do problema ou do tipo de aplicação. Uma característica é definida como uma parte "interessante" de uma imagem, e elas são usadas como um ponto de partida para muitos algoritmos da visão computacional o qual

é uma das principais primitivas para os algoritmos de detecção. Consequentemente, a propriedade em que um detector de característica desejavelmente deve ter é a repetibilidade: se a mesma característica será ou não detectada em duas ou mais imagens diferentes da mesma cena.

A detecção de características é uma operação de processamento de imagem de baixo nível. Isto é, geralmente é executado como a primeira operação em uma imagem, e examina cada pixel para ver se há uma característica presente nesse pixel. Se isso for parte de um algoritmo maior, então o algoritmo normalmente apenas examinará a imagem na região de interesse. Como um pré-requisito interno, a imagem de entrada é geralmente suavizada por um *kernel gaussiano* em uma representação de espaço em escala e uma ou várias imagens de característica são computadas, muitas vezes expressas em termos de operações de derivadas de imagens locais.

Ocasionalmente, quando a detecção é computacionalmente cara e há restrições de tempo, um algoritmo de nível mais alto pode ser usado para guiar o estágio de detecção de característica, de modo que apenas determinadas partes da imagem são pesquisadas.

Muitos algoritmos de visão computacional usam a detecção de características como etapa inicial, assim, como resultado, um grande número de detectores foi desenvolvido variando nos tipos de características detectados, a complexidade computacional e a repetibilidade.

2.2.4 Bordas

Bordas são pontos onde existe um limite (ou uma aresta) entre duas regiões de imagem. Em geral, uma aresta pode ser de forma quase arbitrária, e pode incluir junções. Na prática, as arestas são geralmente definidas como conjuntos de pontos na imagem que têm uma grande magnitude de gradiente. Além disso, alguns algoritmos comuns encadeiam pontos elevados para dar forma a uma descrição mais completa de uma borda. Esses algoritmos normalmente colocam algumas restrições sobre as propriedades de uma borda, como a forma, a suavidade e o valor do gradiente. Localmente, as arestas têm uma estrutura unidimensional.

2.2.5 Cantos e pontos de interesses

Os termos cantos e pontos de interesse são largamente utilizados e referem-se as características semelhantes entre os pontos de uma imagem. O termo canto surgiu nos primeiros algoritmos para detecção de borda, e depois analisar as bordas para encontrar mudanças rápidas na direção (cantos). Estes algoritmos foram, então, desenvolvidos para que a detecção de borda explícita não fosse mais necessária, por exemplo procurando por altos níveis de curvatura no gradiente de imagem. Foi então notado que os chamados cantos também estavam sendo detectados em partes da imagem que não eram cantos no sentido tradicional (por exemplo, um pequeno ponto brilhante em um fundo escuro pode ser detectado). Estes pontos são frequentemente conhecidos como pontos de interesse, mas o termo "canto" é usado pela tradição.

2.2.6 SURF

Speeded up robust feature usualmente conhecido como SURF, é um detector de características utilizado para reconhecimento de objetos, registro de imagens, classificação ou reconstrução. Ele é parcialmente inspirado pelo descritor de características *Scale-Invariant Feature Transform* ou comumente chamado de SIFT e, segundo Herbert Bay [6], mais rápido do que o SIFT.

Para detectar pontos de interesses o SURF utiliza uma aproximação inteira do detector determinante de Hessian, que pode ser calculado com operações inteiras usando uma imagem integral previamente calculada. Seu descritor é baseado na soma de resposta em torno dos pontos de interesses de *wavelet* de Haar. A imagem é transformada em coordenadas usando técnica de pirâmide, para copiar a imagem original com forma de pirâmide Gaussiana ou pirâmide Laplaciana para obter uma imagem com o mesmo tamanho, mas com largura de banda reduzida. Isto consegue um efeito de desfoque especial na imagem original e garante que os pontos de interesse são invariantes.

2.3 Máquina de estados

Uma máquina de estados finitos, é um modelo matemático de computação usado para projetar programas de computador e circuitos de lógica

sequencial. É concebido como uma máquina abstrata que pode estar em um número finito de estados e está em apenas um estado por vez. O estado em que se encontra é chamado de estado atual e, quando iniciado por um evento ou condição desencadeante, pode mudar de um estado para outro esse ato é chamado de transição. Uma determinada máquina de estado é definida por uma lista de seus estados, seu estado inicial e a condição de disparo para cada transição.

O comportamento pode ser observado em muitos dispositivos na sociedade moderna que executam uma sequência predeterminada de ações dependendo de uma sequência de eventos com os quais eles são apresentados. Exemplos simples são máquinas de venda automática, que distribuem produtos quando uma combinação de moedas é depositada.

Máquinas de estado finito podem modelar um grande número de problemas, entre os quais estão automação de projeto eletrônico, projeto de protocolo de comunicação, análise de linguagem e outras aplicações de engenharia. Em biologia e pesquisa de inteligência artificial, máquinas de estado ou hierarquias de máquinas de estado têm sido usadas para descrever sistemas neurológicos. Na linguística, eles são usados para descrever partes simples das gramáticas de línguas naturais.

Considerado como um modelo abstrato de computação, a máquina de estados finitos tem menos poder computacional do que alguns outros modelos de computação, como a máquina de Turing [3]. Ou seja, existem tarefas que nenhuma máquina de estados pode fazer, mas algumas máquinas de Turing. Isso ocorre porque a memória de uma máquina de estados é limitada pelo número de estados que possui.

Um exemplo de um mecanismo muito simples que pode ser modelado por uma máquina de estado é uma catraca. [7] [9]. Inicialmente, os braços estão bloqueados bloqueando a entrada e impedindo que os clientes passem. Depositar uma moeda ou *token* que desbloqueia os braços, permitindo que um único cliente possa empurrar e passar normalmente. Depois que o cliente passa, os braços são bloqueados novamente até que outra moeda seja inserida.

Considerado como uma máquina de estado, a catraca tem dois estados: bloqueado e desbloqueado. [7]. Existe ainda duas entradas que afetam seu estado: colocar uma moeda no e empurrar o braço. No estado bloqueado, empurrar o braço não tem efeito e não importa quantas vezes o impulso de entrada é dado, ele permanece no estado bloqueado. Colocar uma moeda - isto é, dar à máquina uma entrada de moeda - desloca o estado de bloqueado para desbloqueado onde colocar moedas adicionais no estado desbloqueado não tem efeito. Ou seja, dar insumos de moeda adicionais não altera o estado, no entanto, um cliente empurrando através dos braços, desloca o estado de volta para bloqueado.

A máquina de estado do exemplo anterior pode ser representada por uma tabela de transição de estado, mostrando para cada estado o novo estado e a saída (ação) resultante de cada entrada conforme mostra a Tabela 1 e a Figura 3.

Estado atual	Entrada	Próximo estado	Saída
Bloqueado	Moeda	Desbloqueado	Desbloqueia a catraca e espera ela ser girada
	Girar	Bloqueado	Nenhuma
Desbloqueado	Moeda	Desbloqueado	Nenhuma
	Girar	Bloqueado	Quando o cliente girar a catraca irá travar

Tabela 1 Estados de uma máquina

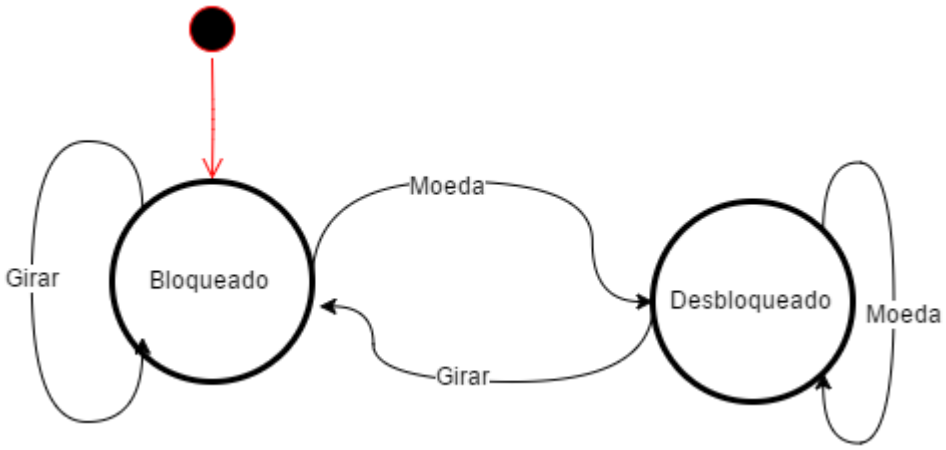


Figura 5 Máquina de estados

3 Modelo Proposto

O modelo proposto foi idealizado para que cada etapa seja independente ao método utilizado para se chegar ao resultado, reconhecendo uma entrada e uma saída que deverá ser comum a todas as formas de geração. Com isso, o modelo é formado da seguinte maneira: detecção de pontos relevantes da imagem, definição das notas dos pontos relevantes e geração da sequência de notas. A figura 1 mostra o diagrama do modelo proposto que, ao fim, resulta em uma música identificada para uma determinada imagem de entrada. Cada caixa representa uma etapa do modelo e trabalha como uma interface entre elas, não importando como é feito, tornando relevante apenas o resultado na saída, e isso permite que o modelo seja flexível à diferentes métodos de aplicação.

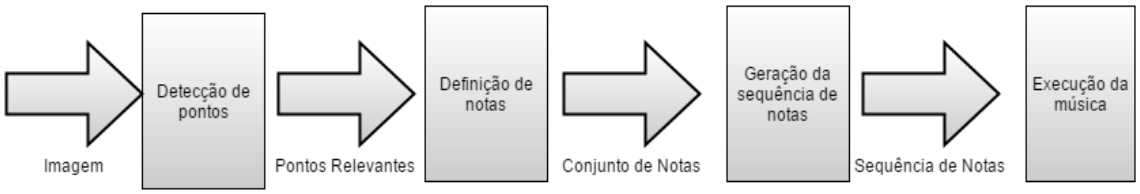


Figura 6 Modelo para geração de música de uma imagem

3.1 Detecção de pontos

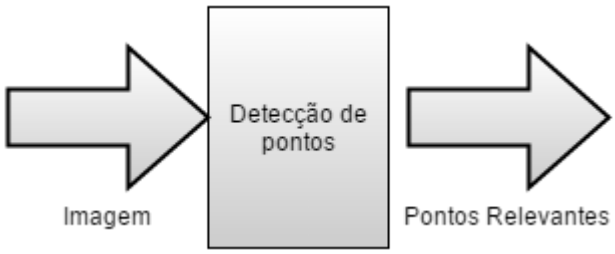


Figura 7 Modelo - Etapa de detecção de pontos

Esta é a primeira etapa do modelo para musicalização de imagens e tem a finalidade de detectar os pontos relevantes e, a partir deles, gerar as notas da música, tendo como entrada uma imagem e resultando em um conjunto de pontos de interesse identificados na imagem. Pode-se utilizar diferentes métodos para extração dos pontos relevantes, como SURF ou KMeans, é nesta etapa que o modelo identifica áreas de acordo com o sentimento passado pela imagem e

ele quem define quais os pontos serão repassados para a próxima etapa. Nesta etapa também é feita o agrupamento dos pontos relevantes da imagem, a quantidade de grupos encontrados está ligada à quantidade de possíveis notas dentro de uma determinada escala musical e cada grupo manda um representante para ser o ponto relevante do grupo resultando em um conjunto de pontos relevantes que será utilizado na etapa seguinte.

3.2 Definição de notas

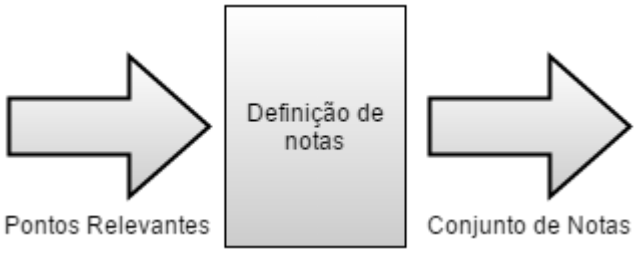


Figura 8 Modelo – Definição de notas

Esta etapa tem como entrada os pontos relevantes da imagem e o resultado é o conjunto de notas relativos aos pontos de interesse que foi dado como entrada como mostra a Figura 8. Uma das formas de definir as notas é através da distância entre o ponto de interesse e as extremidades do cubo de cores RGB. A Equação 1 informa como é feito o cálculo da distância entre os dois pixels. Onde “R” representa o valor do canal *Red* do pixel, “G” o valor do canal *Green* no pixel, “B” o valor do canal *Blue*, “x” e “y” são pixels.

$$D = \sqrt{(Rx - Ry)^2 + (Gx - Gy)^2 + (Bx - By)^2}$$

Equação 1 Cálculo da distância Euclidiana entre dois píxels

Cada extremidade é uma nota em determinada escala musical, por exemplo, o pixel que tem valor RGB igual a 0,0 e 0, respectivamente, pode corresponder a nota Dó, porém, não está restrito a esse método para a definição das notas. A escolha da nota a qual o pixel representa é dada pela menor distância entre os pixels correspondente a cada nota musical e o valor da nota no cubo RGB. O modelo permite que seja utilizada qualquer métrica ou definição que gere as notas ao fim desta etapa, importando apenas o conjunto de notas geradas.

3.3 Geração da sequência de notas

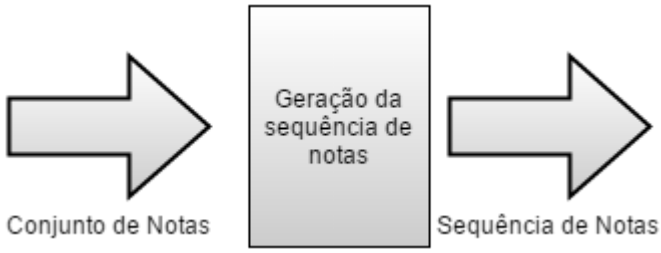


Figura 9. Model – Geração da sequência de notas

Nesta etapa criamos a música para imagem e o resultado será uma sequência melódica que poderá gerar um sentimento que identifique a imagem, porém, não é obrigatório que gere com um sentimento. Este resultado representa a imagem em forma de música.

3.4 Execução da música

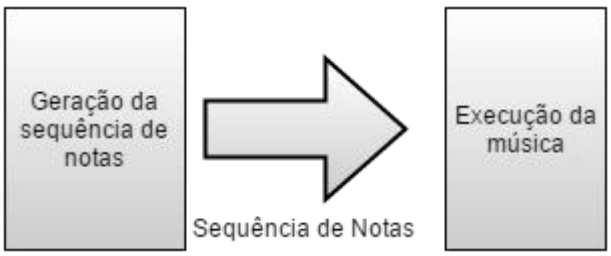


Figura 10. Model – Execução da música

Esta é a etapa final, a qual será realizada a execução da sequência de notas geradas que podem ser em formato de nota pura, ou seja, apenas uma única frequência, ou então em formato de acorde de uma nota, por exemplo, o acorde de Dó Maior.

4 Experimento

Conforme mostra a sessão 3.1 do capítulo anterior, o modelo de musicalização de imagens permite que diferentes métodos para extração e identificação de pontos relevantes da imagem sejam utilizados, e tendo isso em vista, foi implementado dois métodos referentes à essa etapa, o KMeans e o SURF. Nas sessões seguintes serão mostrados a implementação do modelo de acordo com cada um dos métodos escolhidos para identificação dos pontos mais importantes da imagem que para o intuito deste experimento foi o ponto mais focado, porém, não é necessariamente a etapa que deve ser mais explorada, depende apenas de aonde se deseja trabalhar.

4.1 Detecção de pontos e definição de notas

4.1.1 KMeans

O KMeans foi escolhido por ser um método de segmentação de imagem em que dividimos os pixels da imagem em ‘K’ grupos, com isso o número de grupos é definido como a quantidade de possíveis notas musicais diferentes dentro de uma escala, ou seja, um grupo poderá representar a nota Dó, outro grupo Ré, e assim por diante, até que todos os grupos sejam contemplados com uma nota na escala, que neste caso, são utilizados 7 grupos. Porém isso não quer dizer que todos os grupos representarão uma nota diferente. De acordo com a característica do grupo a mesma nota pode ser representada mais de uma vez e por esta razão a definição das notas musicais de cada grupo será mostrado com mais detalhes a frente neste capítulo.

O KMeans separa a imagem em grupos de pixels, e cada um tem uma determinada quantidade de elementos. Destes é escolhido um centro que é o ponto mais próximo de todo o grupo, ou seja, a menor distância dele para os demais elementos utilizando-se a distância Euclidiana entre os pixels da imagem. A quantia de elementos pertencentes a um grupo será utilizada para sabermos qual é a nota predominante da imagem e que, naturalmente, deveria

aparecer mais vezes. A Tabela 2 mostra a relação entre as imagens, a quantidade de elementos e o centro de cada grupo.

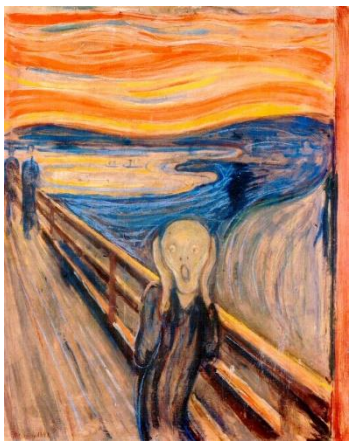
Imagem	Centro	Nº de elementos
 Dimensões (200x163)	167;93	2653
	34;52	3160
	194;87	3258
	148;102	5024
	15; 17	5855
	71;21	6172
	97;118	6478
 Dimensões (200x252)	175;108	5055
	92;199	6002
	4;58	6056
	133;205	6241
	31;27	7653
	0;171	9438

Tabela 2 Resultado da execução do KMeans

Uma imagem colorida geralmente é representada por três canais (RGB, YMC, HSV, etc) formando um vetor de três dimensões conforme a Figura 6. Isto quer dizer que é possível formar um cubo com extremidades definidas como podemos ver na imagem da Figura 7 e com isso, definimos um cubo de musicalização baseado no cubo de cores, onde cada extremidade representa uma nota musical diferente, isso faz com que uma determinada cor tenha um

som específico, ou seja, quanto mais próximo de uma extremidade, maior a chance daquele pixel ter a nota representada pela extremidade. Na Tabela 3 vemos a relação de cada extremidade e cada nota musical

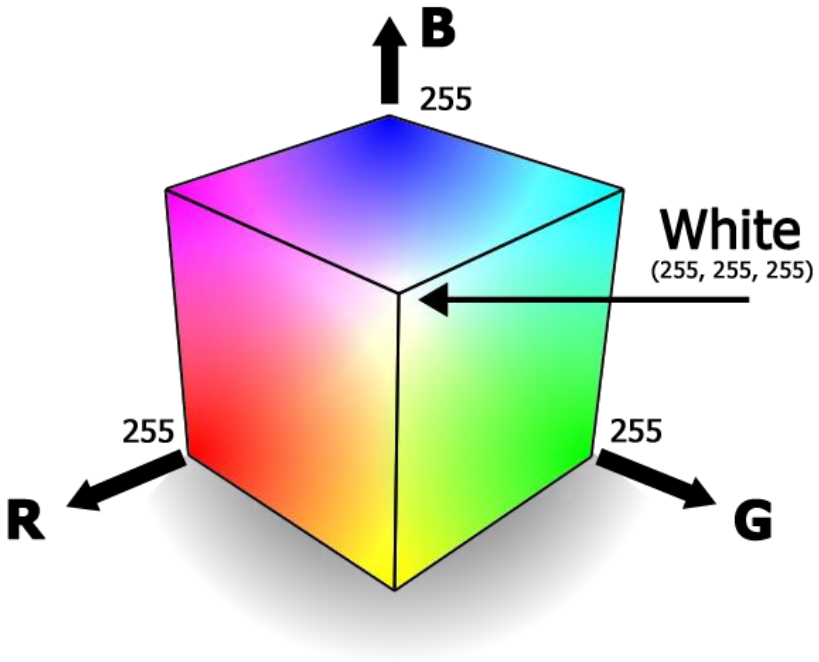


Figura 11. Vetor de cores RGB

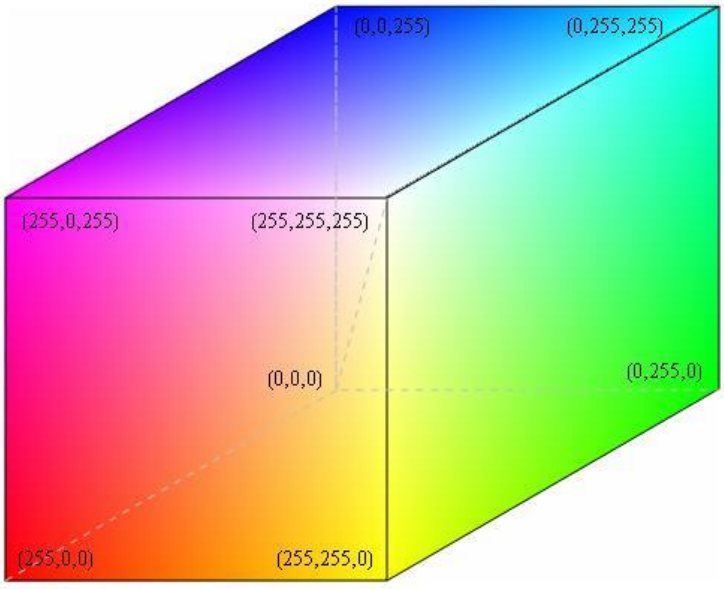


Figura 12. Cubo de cores RGB

Nota	Pixel (R, G, B)
Dó	(0, 0, 0)
Ré	(255, 0, 0)
Mi	(0, 255, 0)
Fá	(255, 255, 0)
Sol	(0, 255, 255)
Lá	(255, 0, 255)
Si	(255, 255, 255)

Tabela 3 Pixels das notas musicais

4.1.2 SURF

O SURF foi utilizado para detecção de pontos de interesse na imagem, sendo o resultado utilizado na etapa de geração de notas vista no Capítulo 3. Diferentemente do que foi implementado no KMeans, o SURF não agrupa os pontos de interesse em torno de um centro e pode retornar um grande número de elementos, acarretando na necessidade de se encontra um meio de agrupar esses pontos de interesse e assim formar as sete notas musicais utilizadas na musicalização nas imagens. A seguir podemos ver a região de interesse calculada pelo algoritmo para algumas imagens e o total de pontos encontrados.

Imagem Original	Imagem SURF	Total de pontos
 Dimensões 4736x2625	 Dimensões 4736x2625	9835
 Dimensões 3840x2160	 Dimensões 3840x2160	43445
 Dimensões 3840x2160	 Dimensões 3840x2160	8473

Tabela 4 Resultado da aplicação do SURF

Após os pontos serem encontrados é necessário dividi-los em grupos para formarmos as notas de cada grupo. Da mesma forma que utilizamos no KMeans, temos no total 7 grupos que foram divididos em 7 notas musicais na escala de Dó maior. Então, para direcionar os pontos para seu respectivo grupo, foi calculado a distância Euclidiana entre o ponto que estamos direcionando e os pontos definidos para cada nota musical de acordo com o cubo de cores definido na Tabela 4. Aquele que tiver a menor distância será o grupo encontrado para o ponto direcionado, ou seja, nesta etapa estamos definindo em qual nota musical o ponto a ser verificado pertence. Ao fim escolhemos uma coordenada para ser utilizada no restante do processo e neste caso é calculado o centro de cada

grupo e também contabilizamos o número de elementos em casa nota musical. A Tabela 5 mostra o resultado dessa etapa para uma determinada imagem.

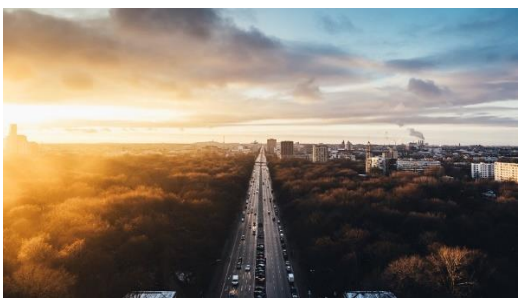
Imagem	Centro	Nota	Nº elementos
 Dimensões 4736x2625	1810;326	Si	2908
	2864; 1623	Dó	4419
	3104; 1325	Ré	1472
	2612; 1803	Fá	241
	863; 2386	Sol	634
	2233; 1351	Lá	41
	1461; 2586	Mi	120
 Dimensões 3840x2160	2230; 1579	Si	15263
	2945; 253	Dó	22406
	2349;1007	Ré	3080
	1068;1702	Fá	2525
	3434; 1778	Sol	135
	2825;2058	Lá	24
	3433;1952	Mi	12
 Dimensões 3840x2160	3535;1173	Dó	5011
	1746;2043	Si	1690
	3802;1192	Ré	1146
	3382;1910	Fá	520
	2811;2111	Sol	68
	1894;1373	Lá	28

Tabela 5 Resultado dos grupos encontrados após a aplicação do SURF

4.2 Geração da sequência de notas

Para esta etapa do modelo, foi montado uma máquina de estados onde cada estado representa uma nota encontrada. No total temos sete estados que representam as sete notas musicais. Como bem sabemos, há transições entre os estados e no universo musical também existem transições entre as notas dentro de uma escala. Visando esse conceito, todos os estados, ou seja, notas, podem ir para um outro ou permanecer no mesmo estado. Uma transição entre estados ocorre quando uma determinada condição é atingida, seja ela qual for, e no caso deste experimento utilizamos um cálculo de probabilidade de um estado sair para um outro. Para calcular essa probabilidade utilizamos a Equação 2.

$$P_{(x,y)} = \frac{n_y + D_{(x,y)}}{n_x + \sum_{i=0}^{j} n_i + \sum_{i=0}^{j} D_{(x,i)}}$$

Equação 2. Cálculo da probabilidade

Onde, x, y e i são estados, $P_{(x,y)}$ é a probabilidade de transição de um determinado estado para outro ocorrer, n_x e n_y é o total de elementos do grupo ao qual pertence a nota gerada, $D_{(x,y)}$ é a Distância Euclidiana entre os pixels representante, $\sum_{i=0}^{j} n_x$ é o somatório dos número de elementos que têm transição a partir de x e $\sum_{i=0}^{j} D_{(x,i)}$ a distância Euclidiana entre os pixels de cada representante que fazem conexão com x. A Figura 8 mostra um exemplo de uma possível máquina de estados onde a probabilidade de ir da nota para outra é mostrada na transição de uma nota para outra e o número de elementos dentro do círculo junto ao nome da nota do estado. Neste exemplo utilizamos apenas 3 estados para representar as notas Dó, Ré e Mi, os pixels representantes de cada nota musical são: (0, 0, 0), (0, 255, 0) e (0, 255, 255) respectivamente e a probabilidade de irmos de um estado para outro se encontra na transição, por exemplo, a probabilidade de sair do estado que contém a nota Dó para o estado que contém a nota Sol é de 57%.

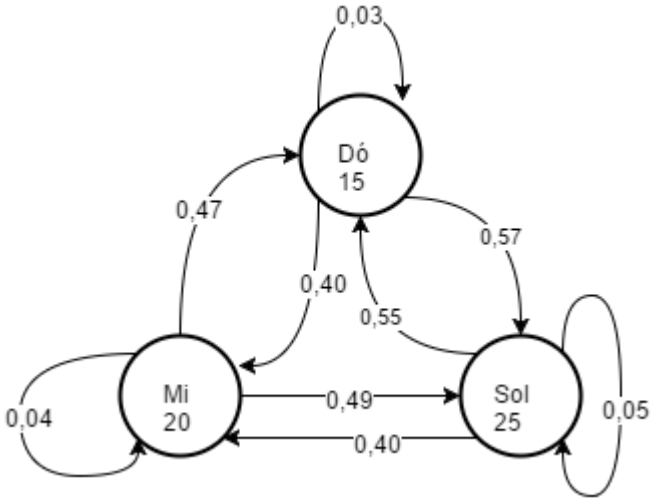


Figura 13. Exemplo de uma máquina de estados gerada.

Ao fim desta etapa, teremos criado uma máquina de estados conforme ilustrado na Figura 8 e que será utilizada para a geração das sequências de notas. A Tabela 6 mostra o resultado para três estados de uma determinada imagem.

Agora que temos a máquina gerada, é possível, então, criar a sequência de notas que serão utilizadas para formar a música da imagem e transformar, de fato, a imagem em um som musical. Para isso, foi utilizado a geração de um número randômico a fim saber qual nota irá compor a sequência, mas primeiramente, é necessário definir o alcance de cada nota. Como a probabilidade está compreendida em valores entre 0 e 1 é possível criar um intervalo fechado que é maior ou igual a 0 e menor ou igual a 100, dado que o valor aleatório também se encontra neste intervalo, assim pode-se distribuir um subintervalo para cada nota, por exemplo, utilizando a Figura 13 temos o seguinte: a probabilidade de sairmos do estado Sol para o estado Dó é de 0,57 ou seja 57%, logo, um possível intervalo pode ser encontrado nos valores entre 1 e 57 garantindo que se tem 57% de chance de o número aleatório cair nesse subintervalo.

Imagem	Estado (Nota)	Transições	Probabilidade
	Si	Dó	0,51
		Ré	0,07
		Fá	0,06
		Sol	0,01
		Lá	0,01
		Mi	0,01
	Dó	Si	0,35
		Ré	0,07
		Fá	0,06
		Sol	0,01
		Lá	0,01
		Mi	0,01
	Ré	Si	0,35
		Dó	0,51
		Fá	0,06
		Sol	0
		Lá	0
		Mi	0

Tabela 6 Resultado da máquina de estado encontrada

O mesmo acontece para os demais estados, para sair do estado Sol para o estado Mi temos a probabilidade de 0,4, ou 40%, e deve-se utilizar o mesmo intervalo definido anteriormente, que nesse caso foi valores entre 1 e 100, e criar um novo subintervalo totalmente distinto dos que já existem, nesse caso, os possíveis intervalos que se pode encontrar são valores entre 58 e 100, e como temos a probabilidade de 40% um novo possível subintervalo é compreendido entre os valores 58 e 98. Os outros 2 valores restantes é probabilidade de permanecer no estado atual. Os subintervalos não precisam ser necessariamente sequenciais, basta obedecer a premissa da probabilidade. Após a definição do alcance de cada estado no intervalo, verificamos o valor do número aleatório e caso ele dê um valor 40, por exemplo, a transição do estado Sol para o estado Dó irá ocorrer, então será utilizado a nota Dó para ser a próxima nota após a nota Sol. No final será gerada uma sequência que pode ser vista na Figura 14. A iteração para gerar o número de notas está diretamente ligada a quantidade de notas definida na entrada do modelo, ou seja, se definirmos que será gerada 20 notas músicas então o modelo irá gerar 20 números aleatórios que serão utilizados para formar a sequência de notas. A Tabela 7 mostra a sequência de 10 notas gerada a partir de uma imagem e esta sequência será executada tornando possível transformar uma imagem em música.

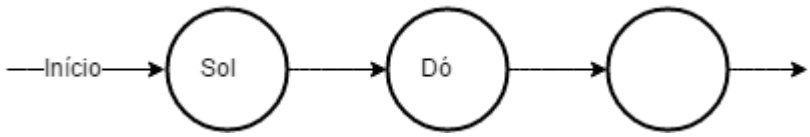


Figura 14. Sequência gerada a partir de uma máquina de estados

Imagem	Sequência
	Si
	Si
	Dó
	Fá
	Si
	Dó
	Si
	Si
	Si
	Dó

Tabela 7 Sequencia de notas encontradas

5 Conclusão

Neste experimento pode-se ver que é possível gerar uma música para uma determinada imagem de entrada, utilizando conceitos de computação e música, formando uma melodia com notas próprias e diferentes a cada imagem. Isso faz com que se abra espaço para aprofundamento nesta área como por exemplo, melhorar a forma de buscar pontos de interesses da imagem, identificar e classificar os pontos da imagem, identificar qual o sentimento que a imagem quer passar para aquele que a observa, e apesar de ter sido utilizado números aleatórios para a geração da sequência de notas, nesta etapa em particular, é possível ir mais adentro dessa parte do modelo proposto a fim de deixar a música mais parecida possível com aquilo que a imagem representa. Muitos outros conceitos podem ser aplicados para que seja possível produzir algo acessível e possível às pessoas que serão beneficiadas com esse experimento. Com isso o objetivo de transformar uma imagem em música foi alcançado.

Conforme falado anteriormente, pode ser realizado um futuro trabalho que melhore as condições de escolha das notas musicais que mais se adequem à imagem, como por exemplo, utilizar as cores da imagem detectando se imagem apresenta um aspecto mais sombrio ou alegre, e dependendo dessa análise a produção de uma música utilizando determinadas notas seja realizada com mais proximidade ao real.

Bibliografia

[1] Barghout, Lauren, Lawrence W. Lee. **Perceptual information processing system**. Paravue Inc. U.S. Patent Application 10/618,543, filed July 11, 2003.

[2] Barghout, Lauren; Sheynin, Jacob (2013). **Real-world scene perception and perceptual organization: Lessons from Computer Vision**. Journal of Vision

[3] Belzer, Jack; Holzman, Albert George; Kent, Allen (1975), ***Encyclopedia of Computer Science and Technology**, Vol. 25*. USA: CRC Press. p. 73

[4] Gonzalez, Rafael; Woods, Richard, **Processamento de Imagens Digitais**, ano 2000.

[5] Guyon and A. Elisseeff. **An introduction to variable and feature selection**. JMLR, 3:1157–1182, March 2003. T. Hastie, R. Tibshirani,

[6] Herbert Bay, Andreas Ess, Tinne Tuytellars, **Speed Up Robust Features.** ETH Zurich, Katholieke Universiteit Leuven

[7] Koshy, Thomas (2004)*, **Discrete Mathematics With Applications***. Academic Press*. p.762*

[8] Linda G. Shapiro, George C. Stockman (2001): **Computer Vision**, pp 279-325, New Jersey, Prentice-Hall.

[9] Wright, David R. (2005)*, **Finite State Machines*** Class Notes. Prof. David R. Wright website, N. Carolina State Univ. Retrieved, July 14,2012*.*