

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

LUCIANO FIORIN JUNIOR

**Seleção dinâmica de classificadores para o
reconhecimento de gêneros musicais**

Maringá
2016

LUCIANO FIORIN JUNIOR

**Seleção dinâmica de classificadores para o
reconhecimento de gêneros musicais**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Yandre Maldonado e Gomes da Costa

Maringá
2016

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá – PR., Brasil)

F521s Fiorin Junior, Luciano
Seleção dinâmica de classificadores para o reconhecimento de gêneros musicais / Luciano Fiorin Junior. -- Maringá, 2016.
95 f. : il. col., figs., tabs.

Orientador: Prof. Dr. Yandre Maldonado e Gomes da Costa.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Tecnologia, Departamento de Informática, Programa de Pós-Graduação em Ciência da Computação, 2016.

1. Reconhecimento de padrões. 2. Seleção dinâmica de classificadores. 3. Combinação de classificadores. 4. Reconhecimento de gêneros musicais - Padrões (Informática) - Sistema de reconhecimento. 5. Recuperação de informação por conteúdo. I. Gomes, Yandre Maldonado e Gomes da, orient. II. Universidade Estadual de Maringá. Centro de Tecnologia, Departamento de Informática, Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDD 21.ed.006.32

ECSL

FOLHA DE APROVAÇÃO

LUCIANO FIORIN JUNIOR

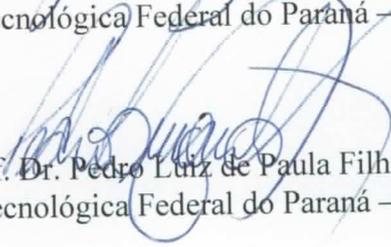
Seleção dinâmica de classificadores para o reconhecimento de gêneros musicais

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA


Prof. Dr. Yandre Maldonado e Gomes da Costa
Universidade Estadual de Maringá – DIN/UEM


Prof. Dr. Diego Bertolini Gonçalves
Universidade Tecnológica Federal do Paraná – UTFPR-CM


Prof. Dr. Pedro Luiz de Paula Filho
Universidade Tecnológica Federal do Paraná – UTFPR-MD

Aprovada em: 19 de fevereiro de 2016.

Local da defesa: Sala 118, Bloco C56, *campus* da Universidade Estadual de Maringá.

Agradecimentos

Agradeço inicialmente ao Prof. Dr. Yandre Maldonado e Gomes da Costa, meu orientador, por ter me acolhido e guiado. Pela sua paciência e orientação. Por todo o conhecimento que me foi passado e pela oportunidade proporcionada.

À Profa. Dra. Linnyer Beatrys Ruiz pela motivação no início dessa caminhada.

À Inês, por todo o auxílio, esforço e incentivo.

Aos meus companheiros de mestrado Pipi, Rodo, Ariel, Tio B e todos aqueles que se fizeram presentes nos meus dias, nos churrascos da turma e nas tardes de café.

Também aos meus companheiros Dudu, Rolinha e Presto, pelos anos de convivência e pelas madrugadas em claro.

À Rafa, por todas as conversas significativas e companherismo.

À minha família, meus pais e meu irmão, por serem referência no meu crescimento. Por todos os ensinamentos, todo o suporte e confiança.

Por fim, agradeço a agência fomentadora Capes que possibilitou o desenvolvimento deste trabalho.

Resumo

Classificação é uma das tarefas presentes na área de reconhecimento de padrões, e pode ser entendida como uma rotina na qual uma classe é atribuída a uma instância analisada. A observação dos atributos de um determinado elemento e a representação desses valores por denota um conjunto de características, popularmente representado na forma de vetor e posteriormente submetido a um classificador. A partir de um conjunto de classificadores, é entendido que a qualidade da decisão tomada pela combinação de opiniões do conjunto supera, na maioria das vezes, a decisão individual de qualquer um dos seus componentes. A seleção de classificadores é definida como um processo para manipulação de um conjunto inicial de classificadores, a fim de obter um grupo reduzido, mas que melhor contribua com o objetivo. No que diz respeito ao reconhecimento de gêneros musicais, a seleção de classificadores visa a melhora nas taxas de classificação das amostras. O presente trabalho aborda a aplicação de estratégias para seleção dinâmica de classificadores na tarefa de reconhecimento de gêneros musicais. Essas estratégias visam a composição de subconjuntos capazes de classificar corretamente padrões desconhecidos e combinação das respectivas estimativas para tomada de decisão. Também foram avaliadas abordagens adaptadas do método KNORA, para elaborar uma vizinhança dinamicamente que auxilie no processo de classificação de amostras semelhantes, além da combinação de métodos que comprovem a relevância da complementaridade de opiniões entre os classificadores envolvidos, de forma a obter uma melhor decisão.

Palavras-chaves: reconhecimento de padrões; seleção dinâmica de classificadores; combinação de classificadores; reconhecimento de gêneros musicais; recuperação de informação por conteúdo.

Abstract

Classification can be defined as a routine in which a class is assigned to a pattern described by a set of attribute. This routine is one of the tasks present in the pattern recognition field. The observation process of the attributes from a specific element and the representation of its values with a data structure, often a vector, denotes their vector of features, thereafter applied to a classifier. Common knowledge in a set of classifiers, at most cases, reaches a better decision than a single one. The classifiers selection is defined as a process for handling an initial set of classifiers to obtain a better small subset, which can achieve better results, closer to the goal. In music genre recognition, the main goal by using classifiers selection is to improve the final recognition rates. This work discusses a bunch of strategies for dynamic classifiers selection in music genre recognition tasks to obtain classifiers subsets who are able to correctly classify unknown patterns using combining strategies for individual estimatives. KNORA method variations, are also investigated were generated dynamically to support a classification cycle. In these cases, the neighborhood members, based on similar samples from a well known set. Finally, investigate the complementary methods with the combination of individual outputs and verify the improvement in recognition rates.

Keywords: pattern recognition; dynamic classifiers selection; classifiers combination; music genre recognition.

Lista de ilustrações

Figura 1	– Limite superior para diferentes n mínimos na Latin Music Database	16
Figura 2	– Espectrogramas de três diferentes gêneros: (a) espectrograma de música de forró; (b) espectrograma de música gaúcha; e (c) espectrograma de música de bolero.	28
Figura 3	– Limiarização de pixel com LBP	30
Figura 4	– Uniformidade de um padrão em LBP	31
Figura 5	– Etapas para reconhecimento de padrões com combinação de classificadores	35
Figura 6	– As três diferentes razões para combinar classificadores.	37
Figura 7	– Três diferentes esquemas para seleção e combinação de classificadores: (a) Seleção estática de conjunto; (b) Seleção dinâmica de classificador; (c) Seleção dinâmica de conjunto. A linha sólida indica processo estático, aplicado uma única vez para todos os padrões de teste, e as linhas tracejadas indicam processo dinâmico repetido uma vez para cada padrão de teste.	41
Figura 8	– KNORA ELIMINATE utiliza somente os classificadores que classificam corretamente todos os k padrões mais próximos. No espaço de características, o círculo escuro corresponde ao padrão de teste e os 5 vizinhos mais próximos os círculos preenchidos. No espaço de classificadores as regiões preenchidas correspondem ao conjunto de classificadores utilizados.	45
Figura 9	– KNORA UNION utiliza os classificadores que classificam corretamente algum dos k padrões mais próximos. No espaço de características, o círculo escuro corresponde ao padrão de teste e os 5 vizinhos mais próximos os círculos preenchidos. No espaço de classificadores as regiões preenchidas correspondem ao conjunto de classificadores utilizados.	46
Figura 10	– Etapas do sistema de classificação	49
Figura 11	– Segmentação do sinal e geração do espectrograma	49

Figura 12 – Representação das etapas do sistema de classificação.	51
Figura 13 – Segmentação do sinal	51
Figura 14 – Espectrogramas gerados a partir do segmento de 60 segundos de sinais de áudio da Latin Music Database (LMD). Na figura 14a um espectrograma de axé e na figura 14b de bolero	53
Figura 15 – Representação de espectrograma e zoneamento com escala Linear	54
Figura 16 – Representação de espectrograma e zoneamento com escala Mel .	55
Figura 17 – Processo de combinação de classificadores na etapa de classificação	57
Figura 18 – Esquema geral de classificação para um único classificador . . .	59
Figura 19 – Vetor de resultados dos erros cometidos pela aplicação do $LBP_{8,2}$ na base LMD	61
Figura 20 – Vetor de resultados dos erros cometidos pela aplicação do LPQ_7 na base LMD	61
Figura 21 – Combinação dos resultados apresentados na figura 19 e na figura 20 em que somente uma das aplicações comete erro	61
Figura 22 – A combinação dos resultados apresentados na 19 e na figura 20 em um cenário no qual as amostras são classificadas incorreta- mente pelas duas aplicações apresentadas	61
Figura 23 – Taxas de reconhecimento variando k vizinhos para variações do método KNORA aplicadas sobre o Local Binary Pattern (LBP) $_{8,2}$ para regra da soma	75
Figura 24 – Taxas de reconhecimento variando k vizinhos para variações do método KNORA aplicadas sobre o $LBP_{8,2}$ para regra do produto	76

Lista de tabelas

Tabela 1 – Síntese dos trabalhos em classificação de gêneros musicais mais relevantes à esta proposta	25
Tabela 2 – Matriz de confusão para um sistema de classificação binário, com 100 amostras pertencentes à classe A e 100 amostras pertencentes à classe B. Somente 45 amostras da classe A foram corretamente classificadas, enquanto que para a classe B foram 85	63
Tabela 3 – Matriz de confusão para um sistema de classificação binário . . .	63
Tabela 4 – Matriz de confusão multiclasse para <i>fold</i> 1 da LMD, utilizando descritor LBP e regra da soma	64
Tabela 5 – Resultados obtidos a partir da utilização do LBP _{8,2}	70
Tabela 6 – Resultados obtidos a partir da utilização do Local Phase Quantization (LPQ)	70
Tabela 7 – Resultados obtidos a partir da utilização do Statistical Spectrum Descriptor (SSD)	71
Tabela 8 – Resultados obtidos a partir do LBP _{8,2} com seleção de classificadores Overall Local Accuracy (OLA)	73
Tabela 9 – Resultados obtidos a partir do LPQ com seleção de classificadores OLA	73
Tabela 10 – Resultados obtidos a partir do LBP _{8,2} com seleção de classificadores Local Class Accuracy (LCA)	73
Tabela 11 – Resultados obtidos a partir do LPQ com seleção de classificadores LCA	73
Tabela 12 – Limite superior nas bases LMD e GTZAN para cada descritor de textura	74
Tabela 13 – Resultados obtidos a partir do LBP com KNORA UNION . . .	76
Tabela 14 – Resultados obtidos a partir do LPQ com KNORA UNION . . .	77
Tabela 15 – Resultados obtidos a partir do LBP com KNORA UNION-W .	77
Tabela 16 – Resultados obtidos a partir do LPQ com KNORA UNION-W .	77
Tabela 17 – Resultados obtidos a partir do LBP com KNORA ELIMINATE	78

Tabela 18 – Resultados obtidos a partir do LPQ com KNORA ELIMINATE	78
Tabela 19 – Resultados obtidos a partir do LBP com KNORA ELIMINATE-W	78
Tabela 20 – Resultados obtidos a partir do LPQ com KNORA ELIMINATE-W	79
Tabela 21 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA UNION	79
Tabela 22 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA UNION	80
Tabela 23 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA UNION W	80
Tabela 24 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA UNION W	80
Tabela 25 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE	81
Tabela 26 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE	81
Tabela 27 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE-W	81
Tabela 28 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE-W	82
Tabela 29 – Resultados obtidos a partir da soma das porcentagens estimadas para cada classe em um sistema LBP com os respectivos valores do LPQ	82
Tabela 30 – Resultados obtidos a partir da estimativa de maior valor entre as saídas dos sistemas com LBP e LPQ	83
Tabela 31 – Resultados obtidos a partir da soma de estimativas das saídas dos sistemas com LBP e SSD	84
Tabela 32 – Resultados obtidos a partir da soma de estimativas das saídas dos sistemas com LPQ e SSD	84
Tabela 33 – Resultados obtidos a partir da soma de estimativas das saídas dos sistemas com LBP, LPQ e SSD	84
Tabela 34 – Limite superior para as base utilizadas nos experimentos para cada descritor de textura	85

Tabela 35 – Resultados obtidos com a aplicação do KNORA sobre a combinação de classificadores dos métodos LBP e LPQ na base LMD	86
Tabela 36 – Melhores resultados apresentados na experimentação e seus correspondentes na literatura	88

Dedico este trabalho à minha família.

Sumário

1	INTRODUÇÃO	14
1.1	Motivação	14
1.2	Desafios	16
1.3	Objetivos	17
1.4	Contribuições	17
1.5	Estrutura do Trabalho	18
2	REVISÃO BIBLIOGRÁFICA	19
3	FUNDAMENTAÇÃO TEÓRICA	26
3.1	Extração de características	26
3.1.1	Características para descrever conteúdo de textura	27
3.1.1.1	Representação Estatística	28
3.1.1.2	Representação Espectral	28
3.1.1.3	Representação Estrutural	29
3.1.2	Descritores de conteúdo acústico	34
3.2	Sistema com Múltiplos Classificadores	35
3.2.1	Combinação de Classificadores	35
3.2.2	Seleção de Classificadores	40
3.3	Considerações	46
4	METODOLOGIA	48
4.1	Visão Geral	48
4.2	Segmentação do Sinal	50
4.3	Extração de características	52
4.3.1	No domínio visual	52
4.3.1.1	Geração do Espectrograma	52
4.3.1.2	Zoneamento	53
4.3.1.3	Descritores de textura	55
4.3.2	Características acústicas	56

4.4	Combinação de Classificadores	56
4.5	Seleção de Classificadores	57
4.6	Classificação	58
4.6.1	Complementaridade	60
4.6.2	Avaliação de resultados	62
4.7	Considerações	65
5	EXPERIMENTOS	66
5.1	Bases de músicas	66
5.1.1	Gtzan	67
5.1.2	LMD	67
5.1.3	Base de músicas africana	68
5.2	Classificação com único descritor	69
5.2.1	LBP	69
5.2.2	LPQ	70
5.2.3	SSD	71
5.3	Classificação com seleção de classificadores	72
5.3.1	Métodos clássicos	72
5.3.2	KNORA	74
5.3.3	KNORA com seleção de tamanho de vizinhança	78
5.4	Classificação com combinação de descritor	80
5.4.1	LBP e LPQ com soma de estimativas finais	81
5.4.2	LBP e LPQ com comparação individual de classificador	82
5.4.3	Combinação de LBP e LPQ com SSD	83
5.4.4	Limite Superior	85
5.4.5	KNORA aplicado na combinação de LBP e LPQ	85
5.5	Conclusão e avaliação de resultados	86
6	CONSIDERAÇÕES FINAIS	89
	Referências	91

1 Introdução

A classificação é uma das tarefas compreendidas na área de reconhecimento de padrões, na qual é atribuída uma classe a para uma instância analisada, comumente representada na forma de um vetor de características. Popularizada a partir do trabalho de Tzanetakis e Cook (2002) como uma abordagem de reconhecimento de padrões em sinais, a classificação automática de gêneros musicais permanece como uma tarefa desafiadora e desperta interesse por parte da comunidade de pesquisa no contexto de recuperação de informações de música. Entende-se por gênero o rótulo estabelecido para descrever ou categorizar títulos musicais, frequentemente empregada por usuários em tarefas de busca. Entretanto, fatores como subjetividade de percepção humana dificultam o uso de apenas etiquetas geradas manualmente nessas tarefas.

Por algum tempo o processo de organização de músicas se mostrou dependente de interação manual, no formato de meta tags, como dados ID3, normalmente vinculados ao formato de áudio MPEG-1 Audio Layer (MP3). Embora a vinculação do ID3 possua informações relevantes, no que diz respeito à indexação, seleção e classificação, frequentemente os dados apresentados nessas meta tags se mostram incompletos ou errôneos. Essa dificuldade é um dos principais motivos pelo qual se mostra interessante a classificação de gêneros baseada em conteúdo, na qual informações adicionais são recuperadas diretamente do sinal da música, na forma de características acústicas ou, mais recentemente, no domínio visual (COSTA, 2013).

1.1 Motivação

Embora existam várias propostas descritas na literatura relacionadas a automatização dessas tarefas (STURM, 2012), é possível observar, nos resultados apresentados nesses estudos, uma margem para exploração de abordagens em diferentes etapas do processo visando melhoria nas taxas de reconhecimento. Essa margem é observada por meio do cálculo de um limite superior para os classifica-

dores envolvidos, o qual descreve um potencial para o sistema muito superior às taxas alcançadas individualmente ou com a combinação simples e direta entre os classificadores envolvidos.

O limite superior é uma medida auxiliar para avaliar o potencial do sistema de classificadores. Essa medida é apresentada na forma de um cenário ideal, no qual é possível verificar, em cada conjunto de classificadores, a existência de ao menos um classificador com capacidade de classificar corretamente uma determinada amostra. Caso seja identificado esse classificador, a amostra em questão é tomada como possível de ser classificada. Essa estimativa pode ser interpretada de maneira mais realista, como potencial alcançável pelo sistema se existir uma abordagem suficientemente boa para localizar ao menos um classificador capaz de realizar corretamente a classificação, caso exista, em todos os possíveis cenários.

Estendendo a restrição do limite superior para variações nas quais uma amostra só é tomada como possível de classificação caso um mínimo de n classificadores a consiga classificar corretamente. A figura 1 ilustra o limite superior para diferentes valores de n estipulados como quantidade mínima de classificadores corretos, a partir de um conjunto inicial de classificadores. Essa diferença entre os valores obtidos para o limite superior com requisitos mais rigorosos demonstra que é necessário selecionar um subconjunto com exatamente um classificador para obter a melhor taxa de desempenho possível. A dificuldade de elaboração uma estratégia que garanta a utilização do classificador que melhor classifica uma amostra analisada em qualquer cenário possível motiva a exploração e otimização das técnicas de seleção de subconjunto de classificadores (KO *et al.*, 2008).

Como parte do processo de classificação, o uso de estratégias na criação dos conjuntos de classificadores permite abstrair diversas características de um elemento, ou representar diferentes visões de uma mesma característica. Além de que, a possibilidade de selecionar, dentro de um conjunto inicial, a combinação desses classificadores se mostra como uma linha de desenvolvimento promissora, no que diz respeito à melhorias de desempenho no sistema para valores de classificação próximos do que é tomado como ideal (KOERICH; POITEVIN, 2005).

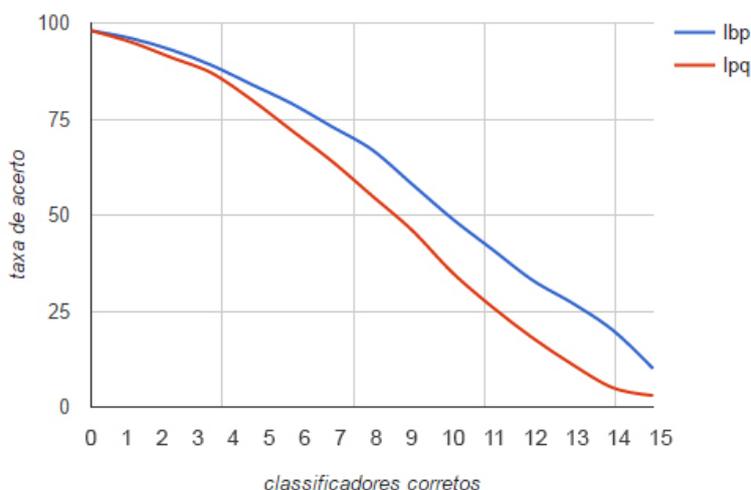


Figura 1 – Limite superior para diferentes n mínimos na Latin Music Database

1.2 Desafios

A principal dificuldade para obter taxas de reconhecimento próximas ao potencial observado no cálculo do limite superior é selecionar, dentre um conjunto inicial de classificadores, o subconjunto ideal que melhor avalie os atributos extraídos da amostra analisada. A descrição de atributos é um processo relevante no que se trata da construção de um sistema de reconhecimento de padrões. Um descritor que não estabelece bons critérios para extração das características em uma determinada amostra dificulta a manipulação do conjunto inicial de classificadores na tarefa de atribuição da classe e afeta as taxas de reconhecimento (DUDA *et al.*, 2012).

Na extração de características, a escolha de métodos que proporcionem a construção de classificadores que possam contribuir para obtenção de uma decisão final correta surge como um dos maiores problemas para o sistema de classificação. Embora um classificador ideal seja entendido como um componente capaz de interpretar qualquer cenário no qual foi aplicado, no decorrer do desenvolvimento de esquemas em busca de soluções para problemas de reconhecimento de padrões foi observado que o conjunto de padrões classificados incorretamente para diferentes abordagens não necessariamente se sobrepõe, contrariando assim o comportamento uniforme esperado pela interpretação. A diversidade na construção e interpretação entre os classificadores permite uma melhoria nos resultados apresentados. Um

classificador pode contribuir na classificação de um determinado padrão de maneira a complementar um outro classificador, caracterizando assim a noção de complementaridade. É dito que um par de classificadores é complementar, se cometem erros diferentes para um mesmo conjunto de padrões (DIETTERICH, 2000).

1.3 Objetivos

O objetivo principal desse trabalho é avaliar estratégias de seleção dinâmica e combinação de classificadores, construídos a partir de características extraídas do sinal de áudio, para rotulação de gêneros musicais. Com base no cálculo de potencial dos sistemas de classificação, entende-se que ainda exista uma melhora nas taxas de reconhecimento a ser alcançada, inerente aos erros cometidos na atribuição de classes para o conjunto de amostras. Desse modo, e viável a exploração de técnicas, que quando aplicadas nas diferentes etapas do processo de classificação, possam contribuir para obtenção de melhor desempenho, no contexto de reconhecimento de padrões em gêneros musicais.

Para atingir os objetivos pretendidos, toma-se como tarefas específicas:

- A utilização de estratégias de extração de características para construção de conjuntos de classificadores complementares utilizados no sistema de classificação.
- Avaliação de desempenho de abordagens em bases de músicas estruturalmente distintas, no que diz respeito a composição das classes de gênero musical.
- Investigação de complementaridade dos classificadores de métodos para um mesmo conjunto de amostragem, a partir de técnicas de combinação e seleção de classificadores

1.4 Contribuições

A principal contribuição esperada com o desenvolvimento deste trabalho é a de identificar a viabilidade de estratégias para a seleção de subconjuntos de classificadores ideais no processo de classificação de gêneros musicais. Dentre as

abordagens avaliadas, incluem-se métodos para extração de características obtidas em espectrogramas, cujo principal atributo visual é a textura, utilizando técnicas de processamento de imagens e posteriormente associadas a geração de classificadores.

Entende-se que um conjunto de classificadores ideal seja aquele que contenha os classificadores mais aptos para classificar corretamente a amostra de uma base de teste. Este conceito visa a seleção de subconjuntos de classificadores, de modo que as taxas de reconhecimento alcançadas pelas estratégias envolvidas estejam mais próximas do potencial estimado para o sistema.

1.5 Estrutura do Trabalho

Este trabalho está organizado como se segue. No capítulo 2 é feita uma revisão bibliográfica que descreve os principais trabalhos na literatura no que diz respeito à abordagens para reconhecimento de padrões. Ainda são apresentados trabalhos que mais impactaram no desenvolvimento desta proposta e uma breve descrição de resultados para efeito de comparação. Para o capítulo 3 fica reservada uma contextualização acerca dos principais conceitos referentes a sistemas de classificação e estratégias aplicadas nas etapas do processo. Já no capítulo 4 são detalhadas as rotinas envolvidas no desenvolvimento da experimentação presente neste trabalho, tal como os estratégias e bases de dados utilizadas. O capítulo 5 apresenta os resultados obtidos utilizando a metodologia descrita anteriormente. Finalmente, o capítulo 6 mostra a discussão sobre os experimentos e conclui o trabalho.

2 Revisão Bibliográfica

O crescimento e popularização de mídias digitais, assim como a facilidade de acesso do público a uma enorme quantidade de dados disponibilizados, tornou necessária a exploração de tarefas automatizadas envolvendo recuperação de informações. Com propósito de facilitar a manipulação de dados de mídia, as soluções computacionais para essas tarefas possibilitaram abstrair a percepção humana em relação a determinadas atividades de avaliação, tal como a classificação de músicas. Neste capítulo serão detalhados os trabalhos mais relevantes encontrados na literatura, em sequência cronológica, relacionados ao desenvolvimento da aplicação de conceitos e metodologias na classificação automática de gêneros musicais. O entendimento de classificação automática de gêneros musicais como uma tarefa pertinente ao reconhecimento de padrões ganhou notoriedade a partir do trabalho de Tzanetakis e Cook (2002). Embora previamente apresentado como uma tentativa de classificar músicas levando em consideração gêneros musicais, o trabalho de Deshpande *et al.* (2001) fazia uso de tarefas automatizadas no domínio visual, definido no processo de extração de características em espectrogramas a partir de Mel-Frequency Cepstral Coefficients (MFCC). Esse trabalho não obteve repercussão expressiva na comunidade acadêmica. O experimento contava com uma pequena amostra de 157 títulos de apenas três diferentes gêneros, selecionados a partir de domínio público disponível na internet sem rigor nos critérios de seleção.

Na proposta apresentada por Tzanetakis e Cook (2002) as características a serem analisadas eram extraídas diretamente do sinal da música a fim de representar o espectro sonoro (*timbral texture*), padrão rítmico (*beat-related*) e altura da nota (*pitch-related*), descrevendo assim uma abordagem baseada em conteúdo. Com base na percepção de que músicas de um mesmo gênero compartilham certas similaridades, como a presença de instrumentos, padrões de ritmo e distribuição de variações de frequência de vibrações, foram propostos conjuntos dessas características para a representação de ritmo, harmonia e textura de timbre, com intuito de identificar os gêneros de cada amostra.

Para a experimentação foi elaborada uma base de dados composta por

1000 faixas de música entre 10 diferentes gêneros. A base denominada GTZAN foi disponibilizada em segmentos sequenciais de 30 segundos de sinal e equilibrada em quantidade com 100 músicas para cada um dos gêneros. O conteúdo para representação de timbre foi obtido com métodos para discriminação e reconhecimento de fala em música, calculados a partir de Short Term Fourier Transform (STFT) para cada frame presente no sinal. O vetor de características para discriminação de timbral consistia de média e variação para um conjunto de aspectos relacionados à textura sobre uma determinada janela observada na amostra. A representação de estrutura rítmica da música se fez baseada em Discrete Wavelet Transform (DWT), uma variação da técnica Wavelet Transform (WT) que provê uma visão uniforme de resolução de tempo para todas as frequências presentes na amostra, porém otimizada para ser computada com apresentação de um sinal mais compacto em relação as unidades de tempo e frequência. O vetor final de características na proposta apresentava um total de 30 características, na qual 6 eram pertinentes a conteúdo rítmico, 19 para textura de timbre e 5 para conteúdo de vibrações. No experimento foram utilizados classificadores Gaussian Mixture Models (GMM) e k Nearest Neighbors (KNN), além da aplicação de um sistema simplificado para zoneamento baseado em frames na atribuição de classes.

A melhor taxa de acerto obtida na experimentação de Tzanetakis e Cook (2002) foi de 61%, valor comparável às taxas obtidas no experimento com tarefas de reconhecimento, realizadas por humanos, apresentado por Gjerdigen e Perrott (2008), o qual era focado na obtenção de dados que seriam utilizados como referência sobre a capacidade humana de reconhecimento gêneros musicais para uma série de amostras. A taxa final de acerto nesse trabalho foi de 70% para uma base de músicas distribuídas entre 10 gêneros musicais populares da década de 90. Além de introduzir de maneira expressiva a classificação de gêneros musicais como uma tarefa de reconhecimento de padrões, o trabalho de Tzanetakis e Cook (2002) apresentaram a primeira base de dados criada especificamente para tarefas de reconhecimento de gêneros musicais e tornou-se referência para desenvolvimentos posteriores no mesmo domínio.

Na sequência, Li *et al.* (2003) propuseram um novo formato de extração de conteúdo, Daubechies Wavelet Coefficient Histogram (DWCH), que captura informações locais e globais do sinal da música simultaneamente conforme os

cálculos de Histogramas de Coeficientes fornecidos por coeficientes de ondulação. Inicialmente realizado sobre a base GTZAN e depois em uma base arbitrária, contendo 756 músicas divididas em 5 diferentes gêneros, o experimento empregou diversos algoritmos de aprendizagem de máquina no processo de classificação, incluindo Support Vector Machine (SVM) e Linear Discriminant Analysis (LDA), para comparação com os estudos anteriores presentes na literatura. Os autores verificaram que o desempenho das tarefas nas quais era utilizado o classificador SVM obtinha resultados melhores do que as tarefas que utilizando os classificadores já utilizados em trabalhos anteriores. O melhor acerto obtido no experimento foi de 72%, para o mesmo conjunto de características disponibilizado por Tzanetakis e Cook (2002).

Embora não esteja diretamente relacionado à área de reconhecimento de padrões, Surowiecki (2005) argumenta que decisões tomadas por um grupo de elementos frequentemente apresenta melhores resultados, se comparadas à realizada por um indivíduo qualquer isoladamente. Nessa abordagem, as demonstrações envolvem grupos de pessoas estimando medidas para ocorrências do cotidiano: peso de uma vaca por observação; localização de um submarino. O autor descreve a diversidade dos elementos que compõe o conjunto como um dos principais fatores para o sucesso da proposição, já que cada indivíduo desenvolve individualmente o raciocínio culminando numa decisão descentralizada. Uma vez produzidas as opiniões individuais, se faz necessário um critério de composição para obter uma decisão consensual. Essa proposição é comparável a geração de um conjunto inicial de classificadores com a tomada de decisão independente e a necessidade de estratégias para a combinação dos resultados.

Analogamente, os experimentos apresentados no trabalho de Koerich e Poitevin (2005) demonstram que a combinação de classificadores distintos atinge resultados superiores e mais acurados que os obtidos a partir da aplicação de processo que utilize um único classificador. O processo para geração de grupos de classificadores depende da forma como são extraídas características, tal como delimitação de regiões utilizadas da amostra ou diversificação de descritores. Nessa proposta são extraídas características a partir de três segmentos da música e as saídas individuais dos classificadores são combinadas para obtenção de uma decisão. A combinação se faz por meio de regras de fusão, introduzidas no trabalho de

Kittler *et al.* (1998), como regra do máximo, soma e produto. O uso de segmentos passou a ser utilizado posteriormente, tendo em vista uma melhor amostragem obtida do sinal, a possibilidade de variações no tempo e tolerância a possíveis erros.

O trabalho de Li *et al.* (2003) introduz uma abordagem diferenciada para extração de conteúdo, na qual a informação é simultaneamente obtida a partir de toda a extensão da amostra e de regiões menores, contendo informação local: o DWCH. Esse método atua como um filtro que decompõe o sinal em etapas, na forma de histograma e inclui atributos para análise de timbre, popularmente utilizado em tarefas de reconhecimento de voz. Diante do fato de que a maioria dos trabalhos desconsiderava a possibilidade de relação definida entre as classes envolvidas no sistema, Li e Ogihara (2005) abordaram a incorporação de hierarquia no processo de classificação, de modo que essa estrutura de relações permitia aplicações mais tolerantes a erros individuais, pois cada classificador teria seu foco voltado à situações de classificação menos conflitantes. Para essa abordagem, foi elaborada uma matriz de confusão para identificar o grau de confusão entre determinadas classes e possibilitar a inferência de relacionamento entre os gêneros.

Tendo em vista o cuidado com uma avaliação adequada por parte dos classificadores de um sistema, que divergem do reconhecimento de gênero e tendenciosamente classificam artistas, Pampalk *et al.* (2005) introduziu o conceito de *artist filter*, no qual um conjunto de treinamento não poderia conter títulos de um mesmo artista contido no conjunto de teste. Os resultados obtidos nas experimentações demonstraram grande redução das taxas de acertos, quando comparados a experimentos que não utilizavam o conceito. A partir desse trabalho o *artist filter* passou a ser utilizado como um valioso instrumento para uma avaliação mais realista nas tarefas de reconhecimento de gêneros.

A carência de músicas disponibilizadas publicamente surgia como um complicador, no que dizia respeito ao desenvolvimento de trabalhos de recuperação de conteúdo. A existência de mídias classificadas equivocadamente nas bases, conteúdo pouco expressivo ou de curta duração e gêneros repetitivos eram tidos como limitadores. Com base nessas dificuldades, Silla *et al.* (2008) apresentaram a Latin Music Database, uma base composta por 10 gêneros provenientes de países latino americanos. O desenvolvimento da base visava o desafio da classificação de gêneros oriundos de regiões culturalmente semelhantes e teve como critério para atribuição

de gênero para os título o modo como cada música é dançada.

Nos experimentos precursores utilizando os gêneros latinos, o resultado com múltiplos classificadores era de 65% de taxa de acerto. Posteriormente, Silla *et al.* (2011) descrevem abordagens mais elaboradas que envolvem a combinação de diferentes conjuntos de características extraídas do sinal: Statistical Spectrum Descriptor, Rhythm Histogram (RH) e IOIHC. Dentre os resultados obtidos, a melhor taxa atingida foi de 89% na base LMD, sem uso de *artist filter*.

Seyerlehner *et al.* (2010) apresentaram o *block-level features*, um conjunto de características capaz de capturar informações temporais em maior quantidade, se comparada a outros tipos de características popularmente utilizadas. Utilizando do classificador SVM, os resultados dos experimentos, considerando 10 gêneros, com a GTZAN apontaram taxas de reconhecimento próximas a 88%. Na sequência, o resultado obtido por experimentos na LMD alcançou uma taxa de acerto de 79,86%, fazendo uso do conceito de *artist filter*.

Costa *et al.* (2012) fazem uso de uma abordagem para reconhecimento de gêneros musicais no domínio visual a partir das características de textura extraídas em espectrogramas, imagens que descrevem densidade espectral das frequências do sinal ao longo do tempo e empregam estratégias de zoneamento para as imagens de entrada, além da combinação de diferentes descritores de textura para geração de classificadores, combinados por meio de regras de fusão. Na avaliação dos resultados, é percebido para o sistema de classificação a preservação de características em determinadas regiões da superfície da imagem nas estratégias de zoneamento e o desempenho individual dos descritores de textura na discriminação de gêneros. Os resultados obtidos na base LMD utilizando um descritor de textura Local Binary Pattern e zoneamento em escala Mel atingem em torno de 82,33%, o melhor resultado até então para a base utilizando *artist filter*. Ainda no domínio de reconhecimento de gêneros musicais, Costa *et al.* (2013a) introduz uma proposta inicial para a seleção dinâmica de classificadores a partir de características extraídas do LBP. O K-Nearest-ORAcles (KNORA) foi utilizado para a etapa de seleção do subconjunto de classificadores e os resultados obtidos na LMD se mostram muito próximos aos alcançados em experimentações anteriores, alcançando 83%.

Embora a disponibilidade de conteúdo no formato digital tenha contribuído para o desenvolvimento de pesquisas em recuperação de informação de música,

existe uma concentração de experimentos utilizando bases de dados tradicionais. Na proposta de Lidy *et al.* (2010) é levantada a questão sobre a aplicabilidade dessas abordagens em cenários diversos, em particular músicas étnicas com características bem diferentes dos padrões já reconhecidos. Como parte do experimento desse trabalho é apresentada uma base de músicas africanas, que apresenta quatro tipos de classificações possíveis para o mesmo conjunto de músicas: país, função, instrumento e etnia. E os melhores resultados alcançados são: 82,21%; 48,25%; 69,06%; e 88,57% respectivamente.

Já no contexto de seleção dinâmica de classificadores, Vriesmann *et al.* (2015) apresenta um comparativo de experimentações empregando uma abordagem adaptada do método KNORA, além de sua combinação com KNN. Embora não tenha apresentado uma melhora significativa nas taxas de reconhecimento, a aplicação dessa adaptação demonstrou desempenho superior ao método original, de modo a fortalecer a necessidade de explorar mais a fundo variantes desse e de outros métodos, visando melhores resultados.

Mais recentemente, o trabalho de Nanni *et al.* (2016) demonstra a eficácia de uma abordagem baseada em *bag-of-features*, uma técnica para classificação de imagens que ignora informação espacial e ocorrência de elementos no conteúdo. Para o reconhecimento de gêneros essa proposta aborda a combinação de características obtidas no domínio visual e acústico. No domínio visual as características são extraídas de espectrogramas divididos com a utilização de escala Mel, a partir de descritores de textura e projeções de *bag-of-features*. A decisão final é tomada utilizando a combinação de classificadores SVM. No domínio acústico são extraídas características diretamente do sinal, como o timbre, e a decisão é tomada pela fusão de classificadores SVM e *Random Subspace of AdaBoost*. Nesse trabalho ainda foi investigada a possibilidade de combinação de classificadores baseado em diferentes descritores. O melhor resultado obtido para a base LMD, foi de 86,1%, com a combinação de classificadores baseados em LBP, multi-scale LBP histogram Fourier descriptor (LBP-HF), e multi-scale rotation invariant co-occurrence of adjacent LBP (RICLBP).

A tabela 1 apresenta uma síntese dos principais trabalhos presentes na literatura relacionados à esta proposta.

Tabela 1 – Síntese dos trabalhos em classificação de gêneros musicais mais relevantes à esta proposta

Autores	Características	Classificador	Base Gêneros	Melhor Acerto
	Conteúdo rítmico			
Tzanetakis e Cook (2002)	frequências de vibração	k-NN e GMM	GTZAN	61,00%
Li <i>et al.</i> (2003)	DWCH	SVM	GTZAN	72,00%
Li e Ogihara (2005)	DWCH	SVM	GTZAN	72,70%
Lidy e Rauber (2005)	SSD e RH	SVM	GTZAN	74,90%
Silla <i>et al.</i> (2008)	Textura de timbre conteúdo rítmico frequências de vibração	J48, 3NN, MLP, NB e SVM	LMD	84,70%
			GTZAN	85,49%
Seyerlehner <i>et al.</i> (2010)	Block-level	SVM	LMD	79,86%
			(Base africana)	
			Função	48,25%
			Instrumentos	69,06%
Lidy <i>et al.</i> (2010)	Marsyas SSD, RH, RP IOIHC	SVM	País	82,21%
			Etnia	88,57%
			LMD	88,06%
Costa <i>et al.</i> (2012)	GLCM, LBP e MARSYAS	SVM	LMD	82,33%
Costa <i>et al.</i> (2013b)	LPQ e Filtros de Gabor	SVM	LMD	80,78%
Costa <i>et al.</i> (2013a)	LBP e KNORA	SVM	LMD	83%
Vriesmann <i>et al.</i> (2015)	Atributos visuais com CFS	k-NN	LMD	70,22%
Nanni <i>et al.</i> (2016)	LBP, LBP-HF, LPQ, RICLBP entre outros	SVM	LMD	86,1%

3 Fundamentação Teórica

Classificação é um dos problemas abordados no contexto de reconhecimento de padrões, e pode ser entendido como o processo que atribui uma determinada classe C_i à um conjunto de características x , extraídas a partir de uma amostra a ser classificada. Para atribuição dessas categorias um sistemas de classificação se baseia em etapas bem definidas: pré-processamento dos dados a serem analisados; extração de características; e a classificação das amostras com atribuição de classes (DUDA *et al.*, 2012).

A abordagem clássica de avaliação por hipóteses determina que uma amostra é aceita ou rejeitada conforme um critério estatístico pré-estabelecido. Um sistema de classificação para reconhecimento de padrões, em geral, utiliza processos baseados em probabilidades, os quais buscam pela hipótese mais provável dentro de um conjunto inicial de hipóteses. A extração das características consiste em identificar atributos de uma determinada amostra de entrada e produzir uma saída que descreva as suas características relevantes para a classificação, representada como uma simplificação dos dados originais. Esse processo que caracteriza o reconhecimento de padrões implica na redução de informação de entrada e representação dos atributos identificados por meio de uma sequência limitada de valores, na forma de um vetor de características, para análise e atribuição de uma categoria para uma determinada amostra.

3.1 Extração de características

A etapa de extração de características apresenta uma parte crucial no sistema de classificação, a fim de potencializar a identificação das classes. Um descritor que não apresente bons critérios para obter as características de uma amostra dificulta a atuação do classificador no reconhecimento de padrões e impacta diretamente nas taxas de classificação. A tarefa de extrair características é diretamente ligada ao contexto na qual a amostra está inserida: as características relevantes; as particularidades ou similaridades das amostras; presença de ruído; ou conhecimento

prévio. Uma vez que as características são extraídas é possível aplicar técnicas de aprendizagem de máquina para realizar a classificação, independente do foco da aplicação a que serão destinadas.

3.1.1 Características para descrever conteúdo de textura

Textura pode ser entendida como o aspecto de uma superfície, geralmente relacionado a distribuição de características visuais em uma determinada representação. No caso de imagens, pode ser entendida como a disposição dos pixels e as propriedades dos elementos contidos, como cores e dimensões. A textura pode, entre outros aspectos, ser avaliada de acordo com a sua força: uma textura é dita fraca quando há reduzidas interações espaciais entre as primitivas e homogeneidade na sua distribuição; uma textura é dita forte quando apresenta uma maior regularidade de interação entre as primitivas. É entendido como interação a distribuição de tipos de primitivas, no qual a textura é descrita pela frequência dessas interações em uma determinada vizinhança (WANG; HE, 1990).

A textura é considerada um importante atributo visual para o reconhecimento de padrões a partir de imagens de espectrogramas. Espectrogramas são imagens bastante ricas em informação que descrevem a densidade espectral das frequências do sinal, eixo vertical, ao longo do tempo, eixo horizontal. Nas Figura 2a e 2b são apresentados dois espectrogramas extraídos de músicas classificadas em gêneros bastante similares, forró e gaúcha respectivamente, no que diz respeito ao conteúdo rítmico e de timbre. Já na Figura 2c pode-se ver o espectrograma extraído de uma música típica de um gênero bastante diferente dos dois anteriores, o bolero. Nessa imagem, percebe-se a presença mais intensa de linhas no sentido horizontal, o que é característico de músicas nas quais há mais ênfase no conteúdo relacionado à harmonia.

Uma das formas de representar essas características é quantificar o conteúdo de textura descritas em uma determinada região. Segundo Gonzalez *et al.* (2004) as técnicas de representação podem ser divididas em: representação estatística, representação espectral e representação estrutural.

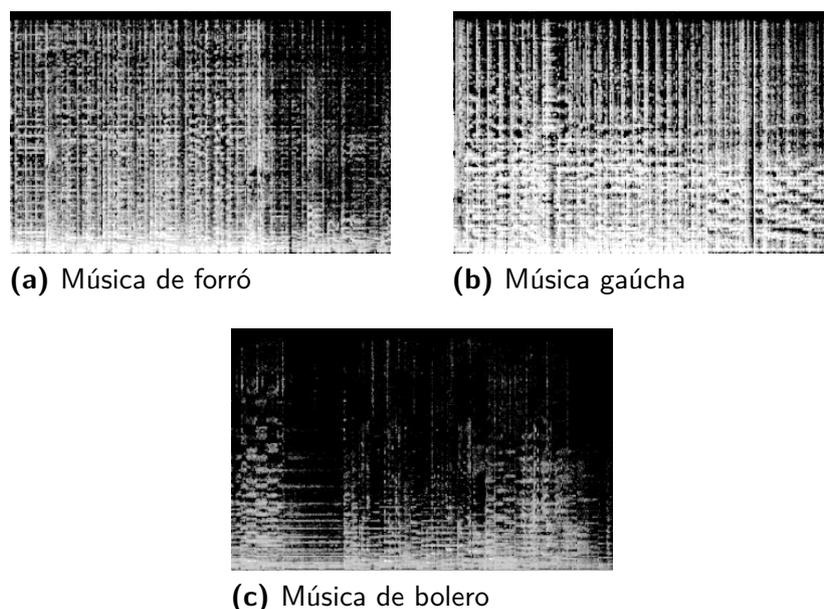


Figura 2 – Espectrogramas de três diferentes gêneros: (a) espectrograma de música de forró; (b) espectrograma de música gaúcha; e (c) espectrograma de música de bolero.

3.1.1.1 Representação Estatística

A representação estatística é uma abordagem que se baseia em propriedades estatísticas presentes no histograma de intensidades. Os dados estatísticos são obtidos a partir da contagem de ocorrências dos níveis de cinza presentes na imagem ou obtidos através da relação de pixels em diferentes níveis de cinza distribuídos em uma determinada região. O uso de matrizes de co-ocorrência é um exemplo de abordagem proposta por Haralick *et al.* (1973) que avalia as formas de relacionamento espacial entre os pixels de diferentes intensidades de cor através de medidas estatísticas. O uso de técnicas nesse tipo de abordagem permite associar à imagem, da qual foram extraídas as características de textura, atributos como lisura, rugosidade e granularidade.

3.1.1.2 Representação Espectral

Baseada na função de densidade do espectro, a representação espectral permite a visão da imagem no domínio da frequência. Medidas espectrais de

textura são baseadas no espectro de Fourier. Textura espectral é um atributo útil para descrever a direcionalidades de padrões periódicos, facilmente distinguíveis pela concentração de regiões de grande energia no espectro, e para discriminar padrões de textura periódicos ou não periódicos. Conseqüentemente uma abordagem ideal para quantificar as diferenças de periodicidade dos padrões observados (GONZALEZ *et al.*, 2004). A Discrete Fourier Transform (DFT) é uma das técnicas de representação espectral que permite decompor uma imagem em suas componentes elementares seno e cosseno. Embora essa técnica se mostre limitada pela representação do conteúdo somente no domínio de frequência é amplamente utilizada em processamento de imagens digitais. Filtros de Gabor, por sua vez, é uma das abordagens utilizadas com êxito por Costa *et al.* (2013b), que permite extrair informações no domínio de frequência e tempo.

3.1.1.3 Representação Estrutural

As abordagens de representação estrutural descrevem características de uma amostra baseadas no princípio de que texturas são compostas de primitivas distribuídas com regularidade e repetitividade em determinada região espacial, as quais podem ser definidas como uma unidade de medidas ou um conjunto de unidades descritas por uma lista de atributos.

A fim de demonstrar o potencial de utilizar um espectro de textura nos métodos de análise de conteúdo, o trabalho de Wang e He (1990) introduziu um modelo de representação estrutural empregando um processo mais elaborado na obtenção das unidades de textura para cada um dos pixels avaliados. Essas unidades eram obtidas com a análise de vizinhos dispostos em 8 diferentes direções em relação ao pixel analisado. O espectro de textura, por sua vez é utilizado para representar a ocorrência dessas unidades de textura computadas sobre uma determinada região. A seguir são apresentadas algumas técnicas para descrição de textura com representação estrutural.

Local Binary Pattern

O Local Binary Pattern (LBP) é um operador apresentado por Ojala *et al.* (1996) e definido sobre as medidas de variações da escala de cinza na representação

da textura. Inicialmente proposto como uma medida complementar para análise de contraste nas imagens e originalmente trabalhava em um bloco de imagem 3×3 . A utilização de uma medida fixa agia como um limitador para descrever detalhes de informações semelhantes de informações em variadas escalas, posteriormente aprimorado para que fosse possível a aplicação em blocos de vizinhança variáveis, a partir do alinhamento circular em relação ao pixel central e interpolação de regiões intermediárias. Os pixels nesse bloco eram limiarizados pelo valor do pixel central em uma representação binária e multiplicados por pesos correspondentes a porção da vizinhança e finalmente somados para obter uma rotulação do pixel central.

Aplicado, o LBP possibilita a representação de aspectos da textura de uma amostra e estabelece um vetor de saída com 59 valores para vizinhanças com 8 vizinhos avaliados e 243 para vizinhanças com 16 vizinhos. Os valores representam características dos padrões e são obtidos com base na diferença de intensidade entre cada um dos pixels da imagem de entrada e seus vizinhos. Das principais motivações para a utilização do método, destaca-se a robustez para análise de mudanças monotônicas em escala de cinza, como variações de iluminação, e a simplicidade computacional. A figura 3 demonstra o processo de cálculo do valor central com base na limiarização de vizinhança.

Nesse método a vizinhança da figura 3(a) é limiarizada pelo valor do pixel central. Os valores dos pixels na vizinhança são multiplicados pelos pesos dados aos pixels correspondentes na figura 3(b). Finalmente os valores da vizinhança são somados para obter o valor da unidade de textura correspondente ao espectro de textura, como pode ser visto na figura 3(c).

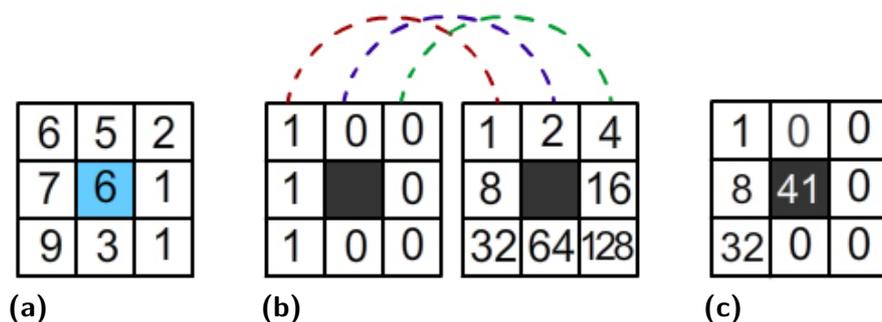


Figura 3 – Limiarização de pixel com LBP

O operador $LBP_{P,R}$ pode ser entendido como a soma das limiarizações aplicadas aos pesos correspondente e formalmente definido por

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p. \quad (3.1)$$

A equação 3.1 representa um operador LBP, para um pixel de coordenadas x_c e y_c , aplicado em uma vizinhança de tamanho P com raio R em relação ao pixel central. O operador é obtido pela soma da função de limiarização $s(z)$ para um vizinho de valor z , descrita na equação 3.2, na qual g_c representa o pixel central, o pixel analisado da vizinhança g_p .

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \quad (3.2)$$

Assim, para um determinado pixel da vizinhança, a diferença entre seu valor é o valor do pixel central, multiplicado pelo peso da posição correspondente, determina parte da somatória para obtenção do operador.

Para uniformidade na avaliação dos padrões é utilizada uma medida relacionada ao número de transição dos bits que compõe os padrões sequenciais circulares, de 0 para 1 e vice versa. Um padrão binário é dito uniforme se o número de transições é no máximo 2. Padrões que apresentem um número maior são ditos não uniformes. Durante a aplicação do LBP, cada um dos padrões uniformes recebe um rótulo único, representado no histograma, enquanto que todos os padrões não uniformes são associados ao mesmo rótulo (PIETIKÄINEN *et al.*, 2011). A representação binária da vizinhança e as uniformidade dos padrões pode ser observada na figura 4.

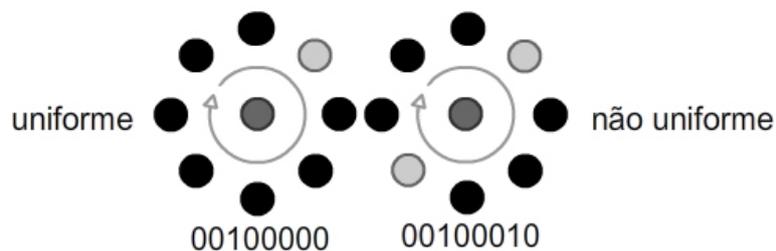


Figura 4 – Uniformidade de um padrão em LBP

Como os padrões representados a partir do operador LBP são obtidos de uma amostra circular a partir de um pixel central, a rotação de uma imagem de entrada implica que os pixels de uma amostra de vizinhança serão analisados em uma orientação diferente (PIETIKÄINEN *et al.*, 2000).

Local Phase Quantization

Informações de textura podem ter sua aplicabilidade limitada em algumas aplicações por degradações no conteúdo da imagem de entrada. Uma das causas desse tipo de degradação é devido ao movimento, e o ruído está diretamente relacionado a forma de aquisição do conteúdo. Em geral, o uso de algoritmos para remover essa degradação são computacionalmente caros, e conforme o grau de ruído, ineficientes. Tendo em vista essa dificuldade, Ojansivu *et al.* (2008) introduziram um método para análise de textura robusto e insensível ao borramento.

Baseado na fase de quantização da DFT computada em pequenas regiões retangulares da imagem denominadas janelas, o LPQ é um operador insensível ao borramento de simetria centralizada, que inclui movimentação, desfoque e turbulência atmosférica. Seu funcionamento aplicado à identificação de textura consiste em computar o operador localmente em cada pixel e apresentar os resultados obtidos na forma de um histograma. Na prática, essa invariância à borramento não pode ser completamente alcançada devido às dimensões limitadas das imagens analisadas. Assim, a convolução da imagem inicial se estende além das bordas e a informação da região é perdida. Esse efeito se torna mais aparente quando o borramento é grande o suficiente, se comparado ao tamanho da imagem (OJANSIVU *et al.*, 2008).

O operador faz uso da informação de fase local extraída utilizando uma DFT de duas dimensões, mais precisamente um STFT calculado sobre uma região retangular N_x de dimensões $M \times M$ para cada posição x dos pixels da image $f(x)$ definida pela equação 3.3

$$F(u, x) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u^T y} \quad (3.3)$$

com o vetor base da 2D DFT na frequência u . Por conveniência, a STFT é

expressa através do vetor de notação da equação 3.4

$$w_u^T f_x. \quad (3.4)$$

sendo que f_x é um vetor contendo as M^2 amostras de imagem de cada vizinhança N_x . O LPQ considera apenas quatro coeficientes complexos correspondentes as frequências 2D: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$ e $u_4 = [a, -a]^T$, sendo a uma frequência escalar que satisfaz à DFT da função de expansão pontual da imagem, *Point Spread Function (PSF)* de $f(x)$. Sendo

$$F_x^c = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)], e \quad (3.5)$$

$$F(x) = [Re\{F_x^c\}, Im\{F_x^c\}]^T \quad (3.6)$$

tendo em vista que *Re* e *Im* retornam, respectivamente as partes reais e imaginária de um número complexo, a matriz de transformação $8-M^2$ pode ser resumida em

$$F(x) = W f_x \quad (3.7)$$

De modo que a função de imagem $f(x)$ seja assumida como resultado de um processo de primeira ordem de Markov, Ojansivu *et al.* (2008) estabelecem que a informação é melhor preservada na quantização quando as amostras avaliadas são independentes. Para garantir essa independência, assumindo uma distribuição Gaussiana da informação, é utilizada uma transformação de clareamento para cada uma das posições da imagem. O resultado obtido é aplicado em um quantizador, descrito como

$$q_j = \begin{cases} 1, & \text{se } g_j \geq 0 \\ 0, & \text{para outros casos} \end{cases} \quad (3.8)$$

sendo g_j cada um dos componentes resultantes da transformação. Os coeficientes quantizados são representados na forma de um inteiro em 0 e 255 utilizando a conversão de binário para decimal

$$b = \sum_{j=1}^8 q_j 2^{j-1} \quad (3.9)$$

finalmente armazenado em um vetor de 256 posições para uso como vetor de características no processo de classificação e correspondente ao histograma do operador.

3.1.2 Descritores de conteúdo acústico

Características de baixo nível obtidas a partir do conteúdo acústico do sinal de áudio têm sido empregadas desde o início dos trabalhos relevantes na área de classificação automática de gêneros musicais. Tzanetakis e Cook (2002) apresentaram como principais tipos de características acústicas: o timbre, harmonia e ritmo. O timbre é tido como uma característica associada a estratégias para diferenciar sons que apresentem mesma intensidade e frequência, geralmente computadas no domínio do tempo e englobando diferentes fontes e instrumentos simultaneamente. A harmonia reflete características presentes no eixo vertical da representação do áudio, diretamente ligada à melodia, que reflete o conteúdo no eixo horizontal. Ritmo está relacionado a regularidade de eventos, como a existência de hierarquia das batidas em função do tempo e a periodicidade de suas ocorrências.

Statistical Spectrum Descriptor e Rhythm Histogram

Baseado na importância das transformações psico-acústicas para manipulação de informação do conteúdo, o trabalho de Lidy e Rauber (2005) introduziram dois métodos com foco nas problemáticas dos algoritmos existentes: Statistical Spectrum Descriptor (SSD) e o Rhythm Histogram (RH). O SSD é um método para extração de atributos rítmicos dos padrões encontrados no sinal de áudio. Inicialmente o sinal de entrada é submetido a transformação utilizando uma Fast Fourier Transform (FFT) para obter um espectrograma. Em um segundo passo, é aplicada uma divisão de 24 zonas no espectro com a escala Bark para representar as características rítmicas separadas por bandas de frequência.

De acordo com a ocorrência de batidas, ou outras variações rítmicas de energia, medidas estatísticas são associadas à uma determinada banda para descrever o conteúdo de áudio. A partir das medidas contabilizadas para cada uma das 24 bandas, são avaliados critérios como média, mediana, variância, assimetria, valores máximos e valores mínimos. O resultado desse processo de avaliação é produto final

do SSD é um vetor contendo 168 características. O RH se difere do SSD na maneira como são armazenadas as informações de ondulação das frequências, agrupando as 24 bandas de modo a formar um histograma de energia rítmica por modulação de frequências. O vetor de características resultante desse histograma apresenta 60 características acústicas.

3.2 Sistema com Múltiplos Classificadores

As publicações da literatura corrente tem apresentado mudanças na concepção original de etapas no sistema de classificação, seja na forma como são extraídas as características quanto no particionamento da etapa final com o uso de técnicas que envolvam multiclassificadores. Os experimentos apresentados nos trabalhos de Koerich e Poitevin (2005) demonstram que a combinação de classificadores distintos atingem resultados superiores e mais acurados que os obtidos de características utilizando somente um classificador. A combinação se faz, na maioria das vezes através de uma regra de fusão e o processo envolvendo multiclassificadores está diretamente relacionado às formas de extração de características, como a utilização de diferentes regiões de uma superfície analisada ou no emprego de diferentes descritores para um mesmo espaço. O processo de classificação utilizando multiclassificadores é ilustrada na Figura 5. Nas próximas seções são descritos alguns aspectos relacionados à seleção e combinação de classificadores no domínio de reconhecimento de gêneros musicais.

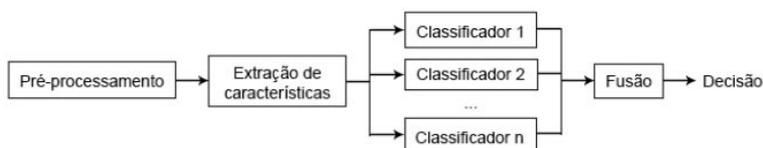


Figura 5 – Etapas para reconhecimento de padrões com combinação de classificadores

3.2.1 Combinação de Classificadores

Os algoritmos empregados na etapa de classificação, como KNN e SVM foram projetados inicialmente visando a resolução de problemas com base na construção de

um único classificador. Diante do desenvolvimento de diferentes esquemas em busca de soluções para os problemas de classificação foi observado entre as abordagens que o conjunto de padrões classificados incorretamente não necessariamente se sobrepunham. Isto sugere que diferentes construções de classificadores oferecem informações complementares dos padrões a serem classificados e que a combinação dessas abordagens poderia permitir uma melhora no desempenho geral da solução (BREIMAN, 1996).

A combinação de classificadores parte da ideia de que uma decisão não se baseia em opiniões individuais. Ao invés disso, diferentes estratégias são utilizadas para gerar um conjunto de classificadores, na qual a diversidade entre eles contribui para existência de diferentes interpretações do cenário e obtenção de melhores resultados (JAIN *et al.*, 2000). O conjunto de classificadores, ou parte dele é utilizado para produzir uma decisão consensual.

A diversidade entre os classificadores permite uma melhoria dos resultados, assim determinados classificadores podem contribuir para classificar corretamente uma amostra que um classificador individual classificou incorretamente. Segundo Dietterich (2000), dois classificadores são complementares se cometem erros diferentes para um mesmo conjunto de padrões. E ainda aponta três razões para a construção de agrupamentos de classificadores: estatística, computacional e representacional.

A primeira razão é estatística e define um algoritmo de aprendizagem como uma busca em um espaço H de hipóteses para identificar a melhor hipótese. O problema estatístico surge quando a quantidade de dados de treinamento é pequena se comparada ao espaço de hipóteses. Sem dados suficientes, o algoritmo de aprendizagem pode encontrar várias hipóteses em H com a mesma precisão nos dados de treinamento. Construindo um agrupamento destes classificadores, é possível estabelecer uma média sobre os resultados e minimizar o risco de decisões equivocadas. A Figura 6(a) ilustra a situação. A curva externa representa o espaço de hipóteses H . A curva interna representa o conjunto de hipóteses com boa precisão. O ponto f representa a hipótese real, aproximada pela média das hipóteses selecionadas.

A razão computacional é a que expõe que algoritmos de aprendizagem realizando busca local podem se prender a mínimos locais. Mesmo nos casos onde há dados suficientes no conjuntos de treinamento pode ser difícil, em termos compu-

tacionais, para o algoritmo encontrar a melhor hipótese. O agrupamento construído para busca local a partir de diferentes pontos de partida pode proporcionar uma melhor aproximação à verdadeira. A Figura 6(b) ilustra a situação.

E finalmente, a terceira razão é a representacional e aponta que em diversas situações de aprendizagem o espaço de hipótese não pode representar uma solução ideal, assim a partir de somas ponderadas de determinadas hipóteses permite expandir o espaço de soluções representáveis. Novamente a situação é ilustrada na Figura 6(c).

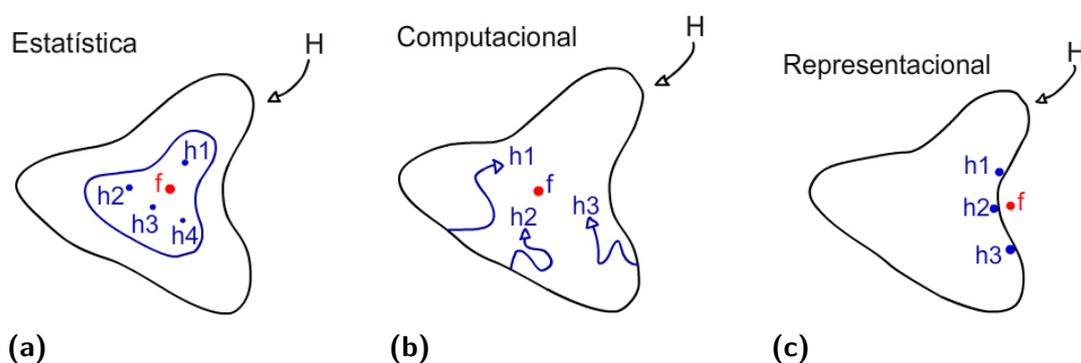


Figura 6 – As três diferentes razões para combinar classificadores.
Fonte: (DIETTERICH, 2000)

A utilização de diferentes classificadores pode ter as saídas representadas na forma de probabilidades estimadas associadas a cada classe envolvida no problema. Diante disso se faz necessário o uso de estratégias para combinação dos resultados a fim de obter uma decisão. A seguir será detalhada a regra do voto majoritário, a mais simples e popular forma de combinação de classificadores, e na sequência as regras descritas por Kittler *et al.* (1998): produto, soma, média, máximo e mínimo. Essas últimas podem ser utilizadas quando as saídas dos classificadores apresentam probabilidades estimadas associadas para cada uma das classes envolvidas no problema.

Regra do voto majoritário

Para a regra do voto majoritário, dado um conjunto inicial de classificadores, cada um dos resultados produzidos é associado a uma classe dentre as possíveis e

computado um voto para a classe. A classe que obtiver o maior número de votos será atribuída ao padrão analisado. Caso exista empate entre duas, ou mais classes, no número total de votos atribuídos deve existir um critério de desempate ou estratégia de rejeição. Formalmente a regra pode ser definida pela equação 3.10

$$mv(x) = \max_{k=1}^c \sum_{i=1}^n y_{i,k} \quad (3.10)$$

na qual a decisão final por voto majoritário é a classe com maior somatória entre as saídas $y_{i,k}$ para um número n de classificadores.

Regra do produto

Na regra do produto a decisão é obtida a partir do produtório entre as probabilidades associadas às saídas dos classificadores e pode ser descrita pela equação 3.11

$$pr(x) = \max_{k=1}^c \prod_{i=1}^n P(w_k | y_i(x)) \quad (3.11)$$

na qual, x representa o padrão a ser classificado, n é o número de classificadores a serem combinados e $P(w_k | y_i(x))$ a probabilidade da amostra x pertencer a uma determinada classe y_i . Essa é considerada uma regra severa de combinação dos classificadores por não possuir mecanismos eficientes para resolver situações envolvendo valores individuais de probabilidades muito reduzidos ou próximos a zero, comprometendo a decisão final em um pior caso envolvendo todos os classificadores e fazendo necessário o uso de uma estratégia de apoio para rejeitar ou aceitar uma hipótese.

Regra da soma

A regra da soma calcula o somatório entre as probabilidades associadas às saídas dos classificadores e pode ser definida pela equação 3.12

$$sr(x) = \max_{k=1}^c \sum_{i=1}^n P(w_k | y_i(x)) \quad (3.12)$$

sendo x o padrão a ser classificado, n o número de classificadores envolvidos e $P(w_k | y_i(x))$ a probabilidade da amostra x pertencer à classe determinada pelo

classificador em questão. É de conhecimento que a regra da soma apresenta maior tolerância à erros de estimativa e na visão geral da literatura consta com os melhores resultados.

Regra da média

Calculada através da média entre as probabilidades associadas a saídas dos classificadores, a regra da média apresenta resultados semelhantes aos da regra da soma, devido a similaridade de resiliência a erros individuais. Essa regra é definida pela equação 3.12

$$mr(x) = \frac{1}{n} \max_{k=1}^c \sum_{i=1}^n P(w_k | y_i(x)) \quad (3.13)$$

dado x o padrão a ser classificado, n o número de classificadores envolvidos e $P(w_k | y_i(x))$ a probabilidade da amostra x pertencer à classe determinada pelo classificador em questão.

Regra do máximo

Para cada um dos classificadores é selecionado o maior valor dentre as saídas do mesmo, de modo a formar um novo conjunto contendo os maiores valores de todos os classificadores envolvidos inicialmente. A regra do máximo seleciona o maior valor contido nesse último conjunto e é atribuído como resultado da amostra a classe associada a esse valor. Essa regra é dita de baixa severidade, pois uma classe que obtenha alta probabilidade no respectivo classificador já tem possibilidades de ser escolhida como decisão final. Formalmente essa regra pode ser definida como

$$max(x) = \max_{k=1}^c \max_{i=1}^n P(w_k | y_i(x)) \quad (3.14)$$

sendo x o padrão a ser classificado, n o número de classificadores envolvidos e $P(w_k | y_i(x))$ a probabilidade da amostra x pertencer à classe determinada pelo classificador em questão.

Regra do mínimo

A regra do mínimo se comporta de forma semelhante a regra do máximo, porém dentre o novo conjunto formado pelos maiores valores de probabilidades de cada um dos classificadores é selecionado o menor valor. Essa regra é considerada severa, pois um classificador que não tenha uma probabilidade conclusiva e esteja indeciso entre diversas classes pode acabar selecionado e impactar no erro de classificação da amostra. Essa regra é descrita por

$$\min(x) = \max_{k=1}^c \min_{i=1}^n P(w_k|y_i(x)) \quad (3.15)$$

sendo x o padrão a ser classificado, n o número de classificadores envolvidos e $P(w_k|y_i(x))$ a probabilidade da amostra x pertencer à classe determinada pelo classificador em questão.

3.2.2 Seleção de Classificadores

Seleção de classificadores é uma estratégia em que, a partir de um conjunto inicial de classificadores, é selecionado um classificador, ou um subconjunto de classificadores, para associar um padrão de teste a uma das classes envolvidas no problema. As técnicas de seleção podem ser divididas quanto ao momento em que são selecionados os classificadores como estática ou dinâmica (KO *et al.*, 2008). No primeiro caso, as regiões de competência são definidas durante a fase de treinamento, enquanto que no segundo, elas são definidas durante o processo de classificação, levando em conta características da amostra a ser classificada. A Figura 7 apresenta diferentes esquemas de seleção de classificadores. Na Figura 7(a) é ilustrado o processo de seleção estática de agrupamento de classificadores. Nela o conjunto de classificadores definido é utilizado para classificar qualquer padrão apresentado ao sistema. Na Figura 7(b), é demonstrado o esquema de seleção dinâmica de classificador, no qual um único classificador é selecionado dinamicamente a cada classificação, levando em conta características do padrão a ser classificado. A Figura 7(c) apresenta o esquema de seleção dinâmica com agrupamento de classificadores. Nele é escolhida uma combinação de classificadores para cada padrão a ser classificado, levando em conta características do padrão corrente. As saídas dos classificadores são combinadas a partir de regras de fusão.

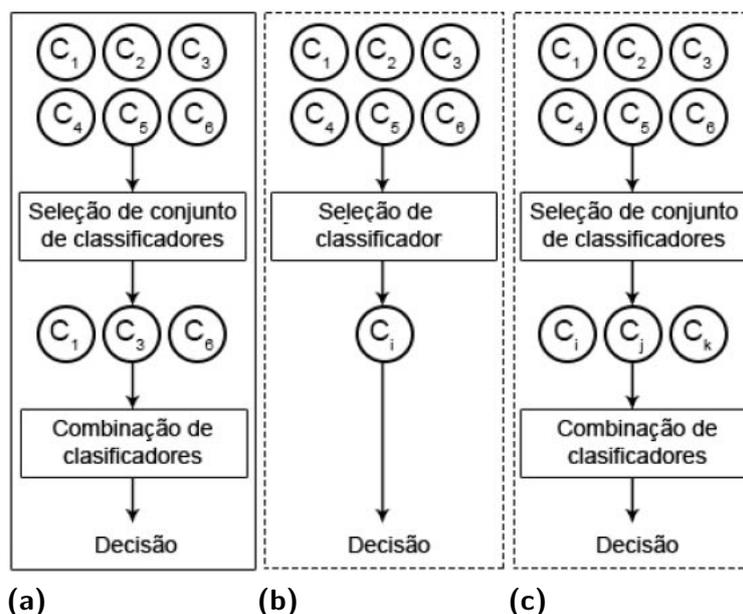


Figura 7 – Três diferentes esquemas para seleção e combinação de classificadores: (a) Seleção estática de conjunto; (b) Seleção dinâmica de classificador; (c) Seleção dinâmica de conjunto. A linha sólida indica processo estático, aplicado uma única vez para todos os padrões de teste, e as linhas tracejadas indicam processo dinâmico repetido uma vez para cada padrão de teste.

Fonte: (KO *et al.*, 2008)

Existem diferentes métodos apresentados na literatura para seleção dinâmica de classificadores a partir de um conjunto inicial, seja através da seleção de um único classificador, quanto na seleção de um subconjunto. A seguir serão descritos alguns detalhes sobre diferentes métodos de seleção de classificadores com potencial para uso no desenvolvimento deste trabalho.

Overall Local Accuracy

O princípio da abordagem Overall Local Accuracy (OLA) é estimar o desempenho de eficiência individual dos classificadores em regiões específicas do espaço de características vizinha a uma determinada amostra e então fazer uso do conceito de classificador local mais eficiente para obter a decisão (WOODS *et al.*, 1996). O esquema de seleção utilizado pelo OLA é ilustrado na Figura 7(b) e as taxas de

eficiência locais são estimadas na forma de porcentagens de amostras de teste classificadas corretamente na região. O pseudo-código 1 descreve o funcionamento do OLA.

Algoritmo 1: Método OLA (MARTINS, 2014)

Entrada: conjunto de validação, rótulos associados por cada classificador C_k
para todo conjunto de validação, tamanho h da vizinhança

Saída : classificador C_k que maximiza a taxa de vizinhos corretamente classificados

- 1 ordene os elementos do conjunto de validação considerando suas distâncias Euclidianas em relação a amostra analisada, em ordem crescente
 - 2 para cada classificador C_k faça
 - 3 conte os h_k vizinhos mais próximos corretamente classificados por C_k
 - 4 calcule a taxa de vizinhos corretamente classificados por C_k
 - 5 fim para
 - 6 selecione C_k que maximiza a taxa de vizinhos corretamente classificados calculada nos passos 2-5
-

Local Class Accuracy

Este método é semelhante ao descrito no Local Class Accuracy (LCA), também ilustrado na Figura 7(b), porém se diferencia no fato de que a estimativa de eficiência é obtida considerando as classes de saída. O classificador atribui uma classe c_i para cada instância de teste, assim é possível determinar a porcentagem de amostras de uma determinada classe c_i atribuídas por este classificador que foram corretamente classificadas. O pseudo-código 2 descreve o funcionamento do LCA.

Algoritmo 2: Método LCA (MARTINS, 2014)

Entrada: conjunto de validação, rótulos associados por cada classificador C_k
 para todo o conjunto de validação, tamanho h da vizinhança,
 rótulo da classe

Saida : classificador C_k que maximiza a taxa de vizinhos corretamente
 classificados

- 1 ordene os elementos do conjunto de validação considerando suas distâncias
 Euclidianas em relação a amostra analisada, em ordem crescente
 - 2 para cada classificador C_k faça
 - 3 selecione os h_i vizinhos mais próximos no conjunto de validação
 - 4 conte os $h_i k$ vizinhos mais próximos corretamente classificados por C_k
 - 5 calcule a taxa de vizinhos corretamente classificados por C_k
 - 6 fim para
 - 7 selecione C_k que maximiza a taxa de vizinhos corretamente classificados
 calculada nos passos 2-6
-

K-Nearest-ORAcles

Introduzido a partir do trabalho de Ko *et al.* (2008), o método K-Nearest-ORAcles (KNORA) apresenta conceitos similares aos demais, OLA e LCA, na utilização de uma vizinhança de padrões para cada instância de teste, porém diferencia-se por incorporar estratégias de identificação do melhor subconjunto de classificadores com potencial para classificar corretamente uma dada amostra, com base na similaridade de amostras presentes em um conjunto de validação. Esse esquema de seleção de conjunto é ilustrado na Figura 7(c). Para cada instância de teste, o KNORA encontra seus k vizinhos mais próximos presentes no conjunto de validação e identifica quais classificadores identificam corretamente as classes desses vizinhos. Depois de identificados é formado um subconjunto de classificadores que serão empregados na classificação do padrão de teste. Para aplicação do KNORA os autores propõem quatro diferentes variações do algoritmo: KNORA-ELIMINATE, KNORA-UNION, KNORA-ELIMINATE-W e KNORA-ELIMINATE-W. O pseudo-código 3 ilustra o funcionamento do KNORA-ELIMINATE.

Algoritmo 3: Método KNORA-ELIMINATE (MARTINS, 2014)

Entrada: conjunto de validação, rótulos associados por cada classificador C_k
 para todo o conjunto de validação, tamanho h da vizinhança

Saída : classe predita

- 1 ordene os elementos do conjunto de validação considerando suas distâncias Euclidianas em relação a amostra analisada, em ordem crescente
 - 2 para cada classificador C_k faça
 - 3 contar os h_k vizinhos corretamente classificados por C_k
 - 4 fim para
 - 5 enquanto $h > 0$ e não houver um subconjunto de classificadores C^* que classifica corretamente todos os h elementos faça
 - 6 decremente h
 - 7 fim enquanto
 - 8 se $h > 0$
 - 9 selecione C^*
 - 10 senão
 - 11 enquanto pelo menos um classificador não classificar um vizinho corretamente faça
 - 12 incremente h
 - 13 fim enquanto
 - 14 selecione C^*
 - 15 fim se
 - 16 use a regra de fusão 'voto majoritário' para identificar a classe eleita dentre aquelas apontadas pelos classificadores C^*
-

KNORA-ELIMINATE

Dado k vizinhos de um padrão de teste x_j , no qual $1 \leq j \leq k$, e supondo que um conjunto de classificadores $C(j)$, $1 \leq j \leq k$ classifica corretamente os k mais próximos vizinhos, então cada classificador $c_i \in C(j)$ submete um voto associado a amostra x . No caso de nenhum classificador conseguir classificar corretamente todos os k vizinhos mais próximos da amostra é selecionado o classificador que

tenha recebido uma maior quantidade de votos no subconjunto K ou cada um dos classificadores com o maior número de votos, em caso de empate. A Figura 8 ilustra esta estratégia.

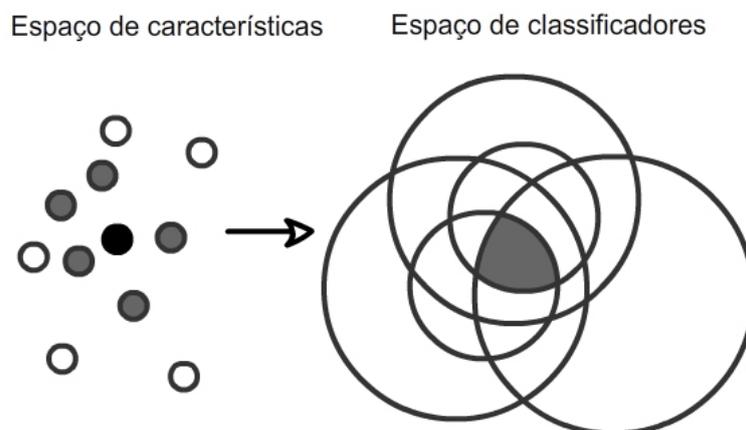


Figura 8 – KNORA ELIMINATE utiliza somente os classificadores que classificam corretamente todos os k padrões mais próximos. No espaço de características, o círculo escuro corresponde ao padrão de teste e os 5 vizinhos mais próximos os círculos preenchidos. No espaço de classificadores as regiões preenchidas correspondem ao conjunto de classificadores utilizados.

Fonte: (KO *et al.*, 2008)

KNORA-UNION

Dado k vizinhos x_j , $1 \leq j \leq k$ de um padrão de teste x , e supondo que o um vizinho j seja classificado corretamente pelo conjunto de classificadores $C(j)$, $1 \leq j \leq k$, então cada classificador $c_i \in C(j)$ submete um voto associado à amostra x . Como todos os k vizinhos mais próximos são considerados, um classificador pode estar associado a mais de um voto se classificar corretamente mais do que um vizinho. Quanto mais vizinho um determinado classificador puder classificar corretamente, mais votos este terá na classificação do padrão. A Figura 9 ilustra esta estratégia.

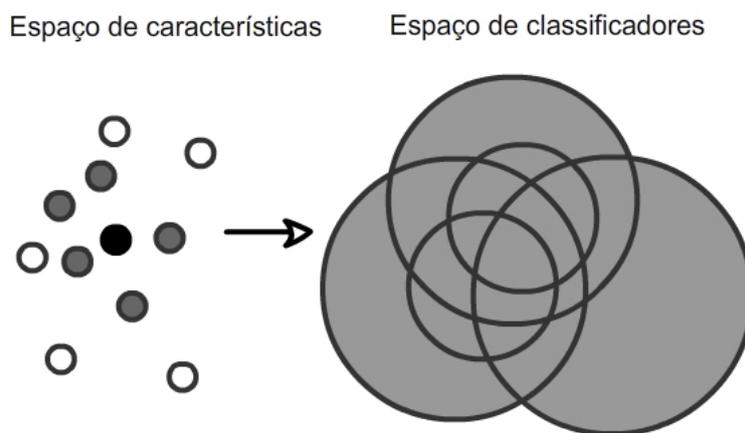


Figura 9 – KNORA UNION utiliza os classificadores que classificam corretamente algum dos k padrões mais próximos. No espaço de características, o círculo escuro corresponde ao padrão de teste e os 5 vizinhos mais próximos os círculos preenchidos. No espaço de classificadores as regiões preenchidas correspondem ao conjunto de classificadores utilizados.

Fonte: (KO *et al.*, 2008)

KNORA-ELIMINATE-W

Este esquema é similar ao KNORA-ELIMINATE, porém cada voto é submetido a um peso baseado na distância euclidiana entre o vizinho x_j e o padrão de teste x .

KNORA-UNION-W

Este esquema é similar ao KNORA-UNION, porém cada voto é submetido a um peso baseado na distância euclidiana entre o vizinho x_j e o padrão de teste x .

3.3 Considerações

Neste capítulo foram apresentados os principais conceitos tratando-se de representação de características e desenvolvimento do sistema de classificação, no que se refere ao reconhecimento de padrões relacionado ao foco do presente estudo, o qual, visa avaliar abordagens reconhecidas sobre conjuntos de classificadores estabelecidos a partir de vetores de características e aprimorar as taxas de eficiência

no domínio de reconhecimento de gêneros musicais. Inicialmente, na seção 3.1, foi tratada a divisão estabelecida na literatura para as abordagens de representação de textura: estatística, estrutural e espectral.

Posteriormente, a seção 3.2 aborda as interpretações já aceitas na literatura sobre etapas do sistema de classificação envolvendo multiclassificadores e descreve pela subseção 3.2.1 os aspectos relacionados à combinação desses classificadores, gerados a partir do conjunto inicial de vetores de características. Ainda, na subseção 3.2.2 foram descritas as particularidades de algumas estratégias que visam a seleção de um subconjunto ideal classificadores dentro do conjunto inicial, de modo que seja possível realizar a melhor classificação entre as classes existentes no sistema classificador para a base de amostras. Detalhes sobre o uso de cada uma dessas técnicas, assim variantes das mesmas, serão apresentadas no capítulo 4.

4 Metodologia

Dentre as dificuldades encontradas para o desenvolvimento de abordagens que visam melhorar as taxas de classificação, no domínio de reconhecimento de padrões, pode-se citar a eficiência dos processos de geração e seleção dos classificadores utilizados no sistema de classificação. Segundo Dietterich (2000) um dos maiores problemas, tratando-se de aprendizado de máquina, está na escolha de métodos que proporcionem a construção de bons classificadores, os quais possam contribuir para obtenção de uma decisão final aceitável para um determinado domínio de aplicação.

Este capítulo inicia com uma visão extendida do modelo de classificação introduzido na figura 5, para um sistema estruturado com foco em conjuntos de classificadores, seguido de um detalhamento das etapas do processo, tal como a avaliação das configurações utilizadas para a extração de características e construção dos classificadores. No que diz respeito a manipulação de dados e seleção de saídas dos classificadores serão descritos, com base nos conceitos apresentados no capítulo 3, as técnicas e estratégias utilizadas na experimentação e as respectivas decisões de projeto.

4.1 Visão Geral

Na tarefa de reconhecimento de padrões, um sistema de classificação trata de atribuir à um determinado objeto uma classe, baseada na análise de suas características. A aprendizagem supervisionada é, em geral, mais utilizada. Nela as classes possíveis para atribuição são previamente definidas, e essencialmente, os sistemas de reconhecimento de padrões são estruturados em quatro etapas: pré-processamento e segmentação; extração de características; classificação; e pós-processamento (SHAH; JETHAVA, 2013). A figura 10 apresenta as etapas do sistema de classificação.

No pré-processamento, operações de filtragem são aplicadas ao volume de dados inicial, de modo a otimizar a qualidade do conteúdo. Na segmentação

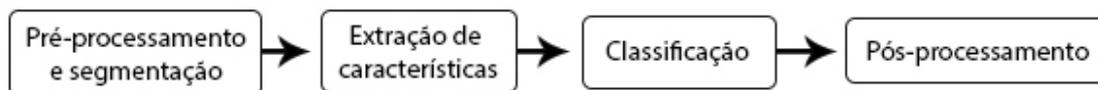


Figura 10 – Etapas do sistema de classificação

são estabelecidos critérios para o particionamento dos dados, tornando assim o volume de informação a ser trabalhada mais adequado às necessidades do cenário de aplicação. Como processo adicional à segmentação, este trabalho contempla ainda a geração de imagens de espectrogramas a partir dos segmentos obtidos, de modo que a extração de características para descritores de textura seja realizada sobre essas imagens, enquanto que a extração por descritores acústicos é realizada diretamente nos segmentos do sinal. A etapa é ilustrada na figura 11.

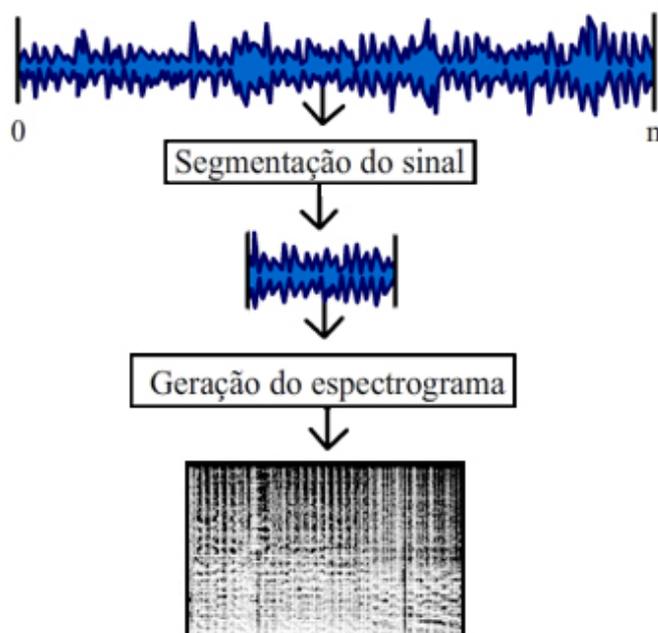


Figura 11 – Segmentação do sinal e geração do espectrograma

Na etapa de extração de características é produzida uma representação do conteúdo observado na amostra para facilitar o trabalho de classificação. Esse conteúdo é submetido a um classificador, cuja tarefa é produzir a melhor classificação possível, dado um conjunto de atributos extraídos. Embora uma mesma

representação possa ser aplicada em diferentes cenários, sua eficácia está diretamente ligada ao tipo de classificador utilizado, de modo que não é possível obter um modelo universal para representar características.

A aplicação das duas tarefas age de forma independente no sistema, embora a eficácia de uma representação esteja diretamente ligada ao tipo de classificador que será aplicado, de modo que não é possível obter um modelo universal de representação de características.

O sistema de classificação reconhece como entrada uma determinada amostra na forma de um vetor de características, o qual é submetido a um classificador, responsável pela atribuição de uma classe que julga mais adequada dentre as possíveis. O principal motivo para a abordagem de classificadores que utilizem de características em forma de vetor é a possibilidade de abstração do processo em relação à aplicação, na qual um mesmo esquema de classificação pode ser empregado para reconhecimento de padrões em cenários distintos. A qualidade do processo de classificação está diretamente relacionada a habilidade do sistema em discriminar cada exemplo entre as classes disponíveis, de modo que exemplares de diferentes classes apresentem particularidades distintas de informação compreensíveis ao classificador (DOUGHERTY, 2012).

Pós-processamento é a etapa final do sistema de classificação e diz respeito às ações tomadas com base no resultado obtido pela classificação das amostras. No caso deste trabalho, podemos associar à etapa intermediária de reclassificação empregada nos experimentos adicionais, na qual as amostras classificadas no primeiro nível de um sistema de classificação multinível são selecionadas com base em um critério de rejeição e reinsertadas no segundo nível do sistema. A figura 12 ilustra as etapas do sistema de classificação.

4.2 Segmentação do Sinal

A segmentação é uma tarefa presente na etapa de pré-processamento responsável pela redução do volume de dados a serem analisados por meio da criação de segmentos obtidos a partir da amostra inicial. Com base no trabalho apresentado por Costa *et al.* (2004), a experimentação presente neste trabalho faz uso de segmentação do sinal, embora se diferencie na quantidade de segmentos, e siga

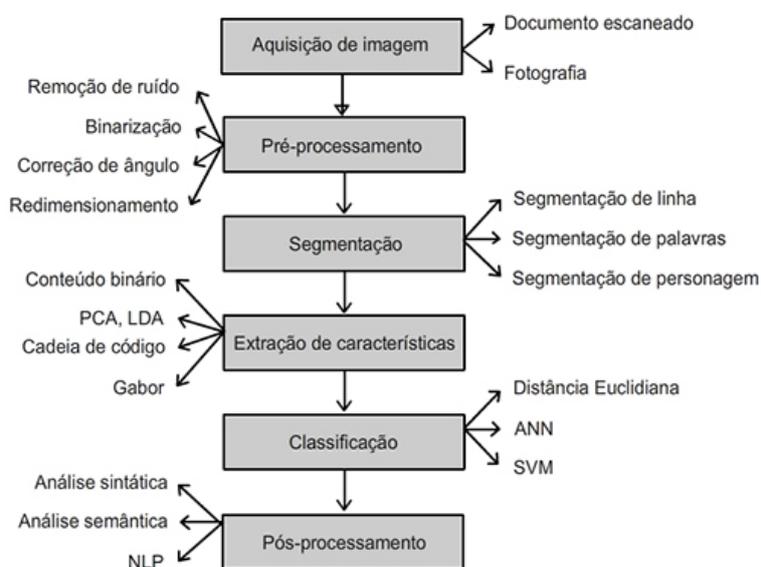


Figura 12 – Representação das etapas do sistema de classificação.
Adaptado: (SHAH; JETHAVA, 2013)

a abordagem apresentada por George e Shamir (2014), na qual são utilizados segmentos únicos de maior duração, ao invés de três segmentos de regiões distintas do sinal.

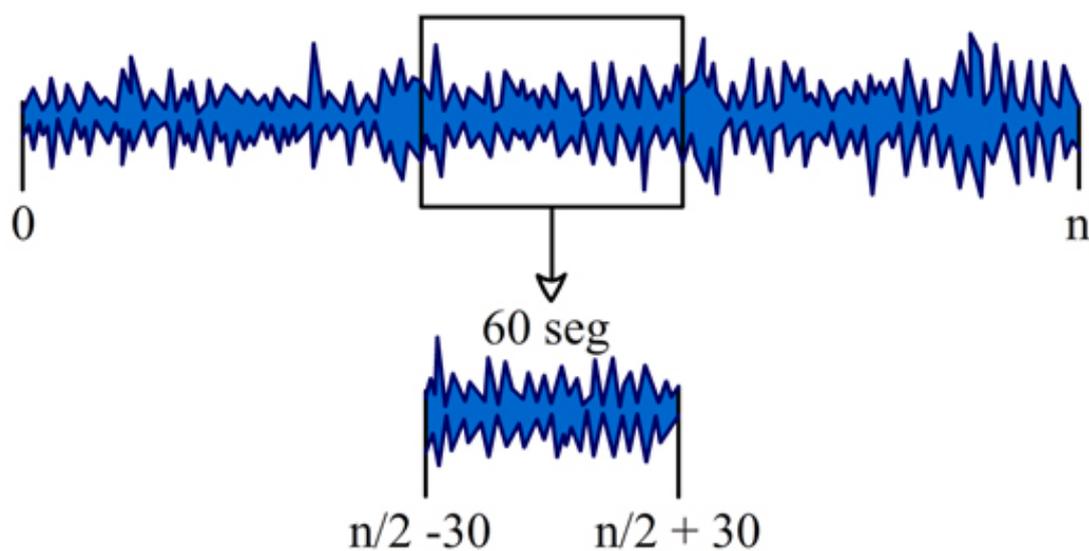


Figura 13 – Segmentação do sinal

Para extração do segmento, foram selecionados trechos de 60 segundos compreendidos entre os segundos $((n/2)-30)+1$ e $n/2+30$, sendo n a duração da música em segundos, conforme demonstrado na figura 13. No caso de amostras com tempo inferior aos 60 segundos, foi utilizado todo o sinal.

4.3 Extração de características

A etapa de extração de características consiste de estratégias para representação de características identificadas na amostra que sejam suficientes para alimentar o processo de classificação e atribuição de um resultado final. Para o reconhecimento de gêneros musicais se mostra possível a extração dessas características diretamente do sinal: características acústicas; ou através de suas representações visuais na forma de espectrogramas: características no domínio visual.

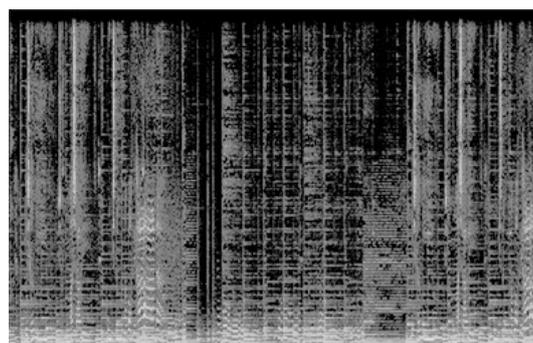
4.3.1 No domínio visual

O conceito de análise de imagens através de informação de conteúdo extraída somente do aspecto visual foi introduzida no trabalho de Bonet e Viola (1997) com propósito de possibilitar a manipulação de dados e busca em grandes bancos de imagens a partir de um conjunto de características extraídas de uma determinada amostra. O reconhecimento de padrões baseado em textura segue o mesmo princípio de comparação de semelhança, no qual a partir de espectrograma é obtido um vetor de características que representa a amostra.

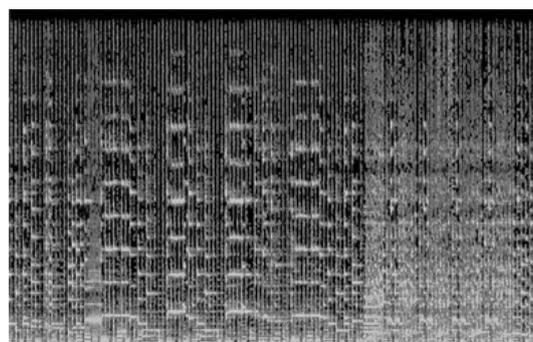
4.3.1.1 Geração do Espectrograma

Espectrograma é uma representação na forma de imagem de um sinal de áudio, na qual o eixo horizontal representa o tempo. O eixo vertical corresponde à frequência e intensidade do brilho de cada pixel da imagem, de modo a representar a amplitude do sinal. Dessa maneira fica representada a variação do sinal no decorrer do tempo.

Na realização dos experimentos presentes neste trabalho, para geração dos espectrogramas foi utilizado o software Sound eXchange (SoX) 14.4.2, disponível em *sox.sourceforge.net* e parametrizado para a geração das imagens já em escala



(a)



(b)

Figura 14 – Espectrogramas gerados a partir do segmento de 60 segundos de sinais de áudio da LMD. Na figura 14a um espectrograma de axé e na figura 14b de bolero

de cinza, para melhor se adequarem aos processos de extração de características. A figura 14 demonstra um espectrograma gerado a partir de amostras da base de músicas LMD.

4.3.1.2 Zoneamento

Tendo em vista a falta de uniformidade ao longo do eixo vertical na textura apresentada pelas imagens de espectrograma extraídas das músicas, foi adotada uma estratégia de segmentação da imagem denominada zoneamento. A fim de preservar informações locais em regiões específicas da imagem o zoneamento tem se mostrado uma estratégia eficiente na extração de características. Além da preservação de informação local, essa estratégia permite a criação de um conjunto com diferentes classificadores, onde cada zona estabelecida pode ser associada a um classificador

específico. A utilização de um conjunto de classificadores no processo de extração de características abre possibilidade para a aplicação de técnicas para a fusão dos resultados, conforme detalhado na seção 3.2.1.

Escala Linear

Para o zoneamento utilizando escala linear é definida uma quantidade de zonas e a imagem é dividida igualmente com base nessa quantidade, no qual cada divisão corresponde a uma faixa de frequências. A figura 15 demonstra as divisões de um espectrograma pela escala linear para uma amostra da LMD. Baseado no experimento apresentado por Costa *et al.* (2013b) a escala Mel apresenta melhores resultados, assim nesta experimentação a escala linear é apresentada brevemente para efeito comparativo em um dos experimentos.

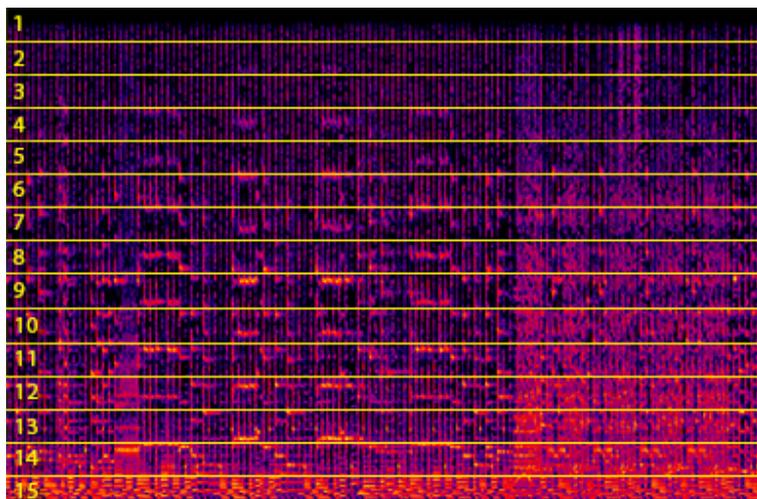


Figura 15 – Representação de espectrograma e zoneamento com escala Linear

Escala Mel

Segundo Umesh *et al.* (1999), a escala Mel é uma estratégia de zoneamento fundamentada na psicoacústica, na qual a distribuição de zonas está condicionada a relação das frequências reais com as frequências perceptíveis por humanos. Para essa escala são estabelecidas 15 zonas que correspondem às bandas de frequência. Embora nos experimentos deste trabalho, nas variações da base africana, foram

empregadas somente 13 zonas pois acima de uma determinada frequência não existe conteúdo relevante nas amostras utilizadas, assim foram descartadas as duas zonas de maior frequência, de modo a não prejudicar o sistema de classificação. A figura 16 ilustra as divisões estabelecidas pela escala Mel para uma amostra da LMD.

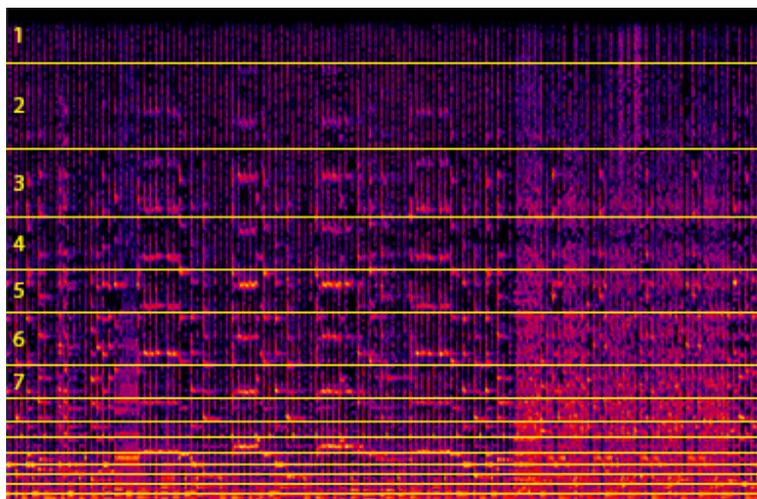


Figura 16 – Representação de espectrograma e zoneamento com escala Mel

4.3.1.3 Descritores de textura

O uso de descritores de textura se justifica pelo fato de que esse é o principal atributo perceptível na imagem de espectrograma. Para obtenção do vetor de características no domínio visual, é aplicado um descritor capaz de extrair informação perceptiva e estatística contida na imagem. Foram utilizados nestes experimentos para extração das características o LBP e o LPQ.

Para desenvolvimento da proposta a variação do LBP utilizada foi $LBP_{8,2}$, de modo a seguir a linha de desenvolvimento apresentada por Costa *et al.* (2012), que apresenta um vetor de características resultante para a amostra analisada com 59 valores. O $LBP_{8,2}$ é a variação mais encontrada em publicações na literatura e apresenta melhores resultados em relação às outras parametrizações (NANNI *et al.*, 2016). Já para o LPQ, na experimentação, foram utilizadas janelas com dimensões quadradas de 7 unidades de altura e largura.

4.3.2 Características acústicas

No que se trata de características acústicas, o Statistical Spectrum Descriptor (SSD) foi selecionado com base no experimento apresentado no trabalho de Silla *et al.* (2011) por apresentar bom desempenho em tarefas de classificação de músicas, no qual o descritor é aplicado sobre uma das bases presentes neste trabalho: LMD; embora o número de amostras seja diferente, por decorrência da não atribuição ao conjunto de treinamento do conceito de *artist filter*. A utilização desse conceito impõe uma maior complexidade no sistema de classificação, pois reduz a interferência de características associadas ao intérprete na avaliação das amostras e implica em taxas de reconhecimento mais expressivas quando o conceito é utilizado. O SSD resulta em medidas estatísticas vinculadas à amostra de sinal, tal como assimetria ou variância. O conjunto de características obtido conta com 168 valores e representa as flutuações nas bandas críticas e informação timbral. Para extração das características acústicas foi utilizado o *Audio Feature Extrator por Rhythm Patterns, Rhythm Histograms and Statistical Spectrum Descriptors*, disponível em ifs.tuwien.ac.at/mir.

4.4 Combinação de Classificadores

A premissa da combinação de classificadores é que a utilização de diferentes estratégias para a criação de um conjunto de classificadores contribui para a obtenção de melhores resultados, pois a diversidade entre os classificadores existentes oferece diferentes interpretações do cenário. Dessa forma, o conjunto de classificadores, ou parte, dele são utilizados para produzir um resultado, conforme o operador utilizado (DIETTERICH, 2000).

Esse processo se faz por meio de regras de fusão, detalhadas na seção 3.2.1, nas quais o conjunto de saídas dos classificadores é submetido a um processo criterioso que seleciona um subconjunto de classificadores entre o conjunto inicial para contribuir à tomada de decisão. Processo inexistente no caso de sistemas de classificação que contém um único classificador no conjunto inicial de classificadores. A figura 17 ilustra a combinação das saídas de um conjunto de classificadores para obtenção de uma decisão final.

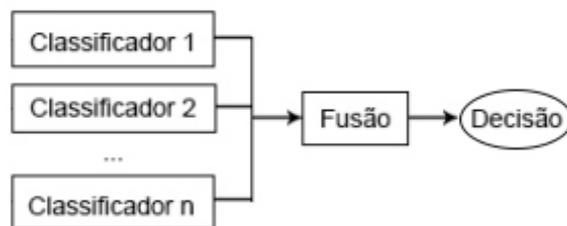


Figura 17 – Processo de combinação de classificadores na etapa de classificação

Com base nos resultados apresentados em Costa *et al.* (2013b) e nas taxas de eficiência promissoras durante a experimentação, foram priorizadas as regras de soma e produto para avaliação de melhores resultados.

4.5 Seleção de Classificadores

Dentre as complicações que envolvem sistemas com uso de múltiplos classificadores uma das mais relevantes é o mecanismo de seleção do subconjunto adequado de classificadores entre os existentes no conjunto inicial, de modo que o grupo selecionado contribui para a otimização das taxas de reconhecimento. Essa tarefa pode ser realizada a partir de seleção estática de classificadores, seleção dinâmica de classificador ou seleção dinâmica de subconjunto de classificadores.

A análise de uma série de padrões está relacionada à diferentes dificuldades de classificação pela interpretação das características da amostra. Isso implica que é correto assumir que para diferentes cenários exista uma seleção mais adequada de classificadores, de modo que um mecanismo de seleção dinâmica de classificador, baseado nas características observadas da amostra, apresente resultados mais promissores que uma abordagem equivalente, mas que utilize seleção estática (GIACINTO; ROLI, 1999). Uma estratégia de seleção dinâmica explora o uso de classificadores distintos para avaliação de um determinado padrão, de forma que para diferentes atributos ou regiões do padrão é feita a seleção de um classificador e associado à amostra.

O uso de uma abordagem de seleção dinâmica de classificador está condicionado diretamente a confiabilidade na estimativa individual do classificador

selecionado para tomada de decisão em relação aos demais. A seleção dinâmica de um subconjunto de classificadores se apresenta como uma opção mais interessante para o sistema de classificação, na qual o risco de estimativa é distribuído sobre um grupo de classificadores no lugar de um classificador selecionado individualmente para um determinado padrão de teste.

Com base nos resultados apresentados no trabalho de Ko *et al.* (2008), que demonstra um desempenho superior nas taxas de reconhecimento para estratégias de seleção dinâmica de conjunto de classificadores, em relação a seleção individual, a experimentação presente neste trabalho faz uso dessa abordagem. O principal método para seleção dos classificadores nos experimentos é o KNORA, e a aplicação é feita na forma de variações do método, no que diz respeito a critérios da seleção. Para efeito comparativo, são apresentados ainda resultados realizados com a utilização de métodos já conceituados na literatura, como OLA e LCA

4.6 Classificação

Para realização das tarefas de classificação o classificador utilizado foi o SVM, um modelo de algoritmos de aprendizagem apresentado por Vapnik e Kotz (1982) e amplamente utilizado em tarefas de reconhecimento de padrões e análise de dados, com desempenho satisfatório descritos na literatura recente (COSTA *et al.*, 2013a).

O SVM foi originalmente desenvolvido para classificação binária, na qual existem somente duas classes para decisão, no entanto a maioria dos cenários envolvendo reconhecimento são problemas multiclasse. De modo a solucionar problemas envolvendo um maior número de classes possíveis, foram extendidos os critérios de avaliação do SVM original, assim o mecanismo se baseia na construção de modelos distintos combinados para descrição de subconjuntos de características, associados aos grupo de amostras. Cadaum dos modelos construído é um classificador binário associado a uma das classes envolvidas no sistema. A construção dos modelos é dividida em dois processos: treinamento e teste. O processo de treinamento consiste na utilização de um conjunto de vetores para a construção de modelos de cada uma das classes e o processo de teste na triabuição de uma classe para a amostra, com base nas características extraídas (VAPNIK; KOTZ, 1982).

Para o processo de classificação do conjunto de amostras da base, foi adotado o conceito de validação cruzada, uma metodologia de confrontamento de conjunto de teste alternados contra os demais conjuntos utilizados para o treinamento, introduzida por Efron e Efron (1982). Fatores como a escassez de dados e a variância de resultado decorrente de má construção do conjunto de teste estão entre os principais motivos para utilização da validação cruzada. Um dilema no sistema de classificação é a alocação de dados relevantes entre o conjunto de treinamento e teste. Em uma situação onde é priorizado reservar uma maior quantidade de dados para construção de um bom modelo, os dados restantes podem se mostrar insuficientes para um conjunto de teste representativo.

Na validação cruzada, o conjunto de dados é dividido em conjunto denominados *folds*, no qual um dos conjuntos é utilizado para teste e os demais para a construção do modelo. O processo é repetido de forma que todos os *folds* tenham sido utilizados como conjunto de teste e treino. A utilização desse conceito permite obter uma melhor estimativa de probabilidades menos sensível a variabilidade dos conjuntos. As estimativas são obtidas com a aplicação das regras de fusão apresentadas por Kittler *et al.* (1998) sobre as saídas do conjunto de classificadores, descritas anteriormente na seção 3.2.1. O esquema geral de classificação para um classificador é ilustrado na figura 18 e as estimativas obtidas na classificação posteriormente utilizada na combinação com os demais classificadores envolvidos no sistema.

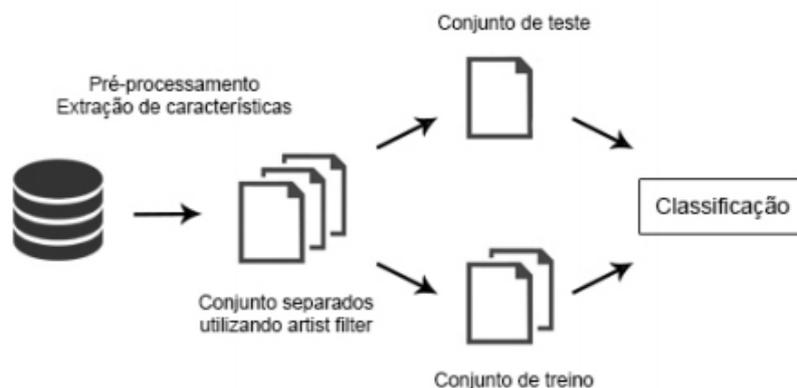


Figura 18 – Esquema geral de classificação para um único classificador

4.6.1 Complementaridade

A utilização de diferentes classificadores aplicados sobre as características extraídas de uma mesma região oferece a possibilidade de que, uma determinada amostra no sistema de classificação seja interpretada de maneiras distintas, de acordo com o comportamento de cada um dos classificadores envolvidos. A representação das características de uma amostra quando submetida à classificadores diferentes produz resultados diferentes, com base no desempenho individual do classificadores. Essa variedade de interpretações é a base do conceito de complementaridade dos resultados, no qual dois classificadores são ditos complementares se apresentam erros diferentes para um mesmo conjunto de teste. Entende-se por erro no segmento de reconhecimento de padrões a atribuição de uma classe equivocada para uma amostra em análise.

Uma das complicações recorrentes observada no desempenho individual de um classificador é a variância nos resultados obtidos para diferentes cenários. Esse problema se deve à dificuldade de abstração para o desenvolvimento de um classificador universal, de modo que a combinação de classificadores com diferentes especializações apresenta melhores taxas de reconhecimento. Embora não seja possível eliminar a complicação decorrente da variância de desempenho individual a utilização de um agrupamento permite alcançar uma maior variedade de cenários com a compensação de erros a partir do conceito de complementaridade, como apresentado em Breiman (1996), trabalho no qual é experimentada a combinação de classificadores com alto nível de variação nas taxas de eficiência.

É possível estender o entendimento de complementaridade das estimativas obtidas na etapa de pós processamento na obtenção de uma decisão final, a partir da incorporação de critérios de combinação de resultados. Com base no experimento utilizando a base LMD para um conjunto de 900 músicas analisadas, a figura 19 ilustra o vetor de resultados para os erros cometidos na aplicação do LBP_{8,2}. A figura 20 ilustra os erros cometidos pela aplicação do LPQ₇. Na figura 21 é ilustrado o resultado disjunto entre os dois métodos, em um cenário no qual somente um dos métodos combinados tenha cometido o erro.

Já na figura 22 são apresentados os erros em um cenário no qual o erro é cometidos por ambos os métodos utilizados e pode ser entendido como uma medida



Figura 19 – Vetor de resultados dos erros cometidos pela aplicação do $LBP_{8,2}$ na base LMD



Figura 20 – Vetor de resultados dos erros cometidos pela aplicação do LPQ_7 na base LMD

auxiliar semelhante ao limite superior para a compreensão do potencial de classificação do conjunto. O conjunto de amostras que são classificadas incorretamente pelas duas aplicações é ilustrado na figura 22 e pode ser entendido como uma medida auxiliar semelhante ao limite superior para compreensão do potencial de classificação do conjunto.



Figura 21 – Combinação dos resultados apresentados na figura 19 e na figura 20 em que somente uma das aplicações comete erro



Figura 22 – A combinação dos resultados apresentados na 19 e na figura 20 em um cenário no qual as amostras são classificadas incorretamente pelas duas aplicações apresentadas

Tomado como uma medida especulativa, o limite superior se baseia no conceito de oráculo, definido pela situação em que, se pelo menos um classificador de um conjunto consegue classificar corretamente uma amostra, considera-se que a

amostra em questão é dita possível de ser classificada em um cenário de seleção ideal de classificadores. Essa medida se mostra útil para representar a melhor taxa de acerto alcançável por um subconjunto de classificadores, obtida pela razão entre o número de amostras em que pelo menos um dos classificadores envolvidos seja capaz de classificar corretamente o total de amostras, conforme definida na equação 4.1.

$$limitesuperior = \frac{c}{n} \quad (4.1)$$

na qual, c corresponde ao número de amostras que teve pelo menos um dos classificadores envolvidos capaz de classificar corretamente a mesma, enquanto que n se refere ao total de amostras analisadas.

4.6.2 Avaliação de resultados

Considerando o problema de reconhecimento apresentado neste trabalho como um cenário multiclasse, foi adotada para avaliação dos resultados obtidos a matriz de confusão, uma tabela que permite o cálculo de taxas de desempenho e a identificação de conflitos entre as classes envolvidas, a fim de reconhecer focos de confusão.

Matriz de confusão é uma tabela auxiliar que demonstra os percentuais obtidos na classificação para cada uma das classes envolvidas, frequentemente utilizada para descrever os pontos críticos do desempenho de um sistema. Essa tabela representa o total de amostras analisadas e o número de equívocos e acertos na classificação. A tabela 2 ilustra uma matriz de confusão para um sistema de classificação binária, na qual n representa o total de amostras classificadas, a soma dos valores de uma linha representa a quantidade de amostras pertencente a uma determinada classe e cada coluna representa a quantidade de amostras atribuídas a uma determinada classe.

A taxa de reconhecimento de classe descrita na matriz é dada como um percentual pela equação 4.2, na qual tx_i , denota a taxa para uma determinada classe i , c_i corresponde ao número de amostras corretamente classificadas para a classe em

n = 200	classe A	classe B
classe A	45	55
classe B	15	85

Tabela 2 – Matriz de confusão para um sistema de classificação binário, com 100 amostras pertencentes à classe A e 100 amostras pertencentes à classe B. Somente 45 amostras da classe A foram corretamente classificadas, enquanto que para a classe B foram 85

n = 200	Positiva	Negativa
Positiva	TN	FP
Negativa	FN	TP

Tabela 3 – Matriz de confusão para um sistema de classificação binário

questão no processo de classificação e t_i o total de amostras que verdadeiramente pertence à classe.

$$tx_i = 100 \frac{c_i}{t_i} \quad (4.2)$$

Para obtenção da taxa de reconhecimento geral da matriz, é aplicado um peso a cada uma das taxas individuais das classes envolvidas no sistema e calculada a média com base no total de classes, t_c , conforme descrito na equação 4.3

$$txg = 100 \frac{\sum \left(tx_i * \left(\frac{n_i}{n} \right) \right)}{t_c} \quad (4.3)$$

Com base na construção de uma matriz de confusão, são obtidas quatro terminologias essenciais: Positivos Reais, descreve uma situação na qual uma amostra pertence a classe dita positiva e o resultado final é também atribuído à classe positiva; Negativos Reais, a amostra pertence à classe dita negativa e o resultado final é atribuído à classe negativa; Falsos Positivos, a amostra pertence à classe negativa, porém é atribuída à positiva; e Falsos Negativos, a amostra pertence à classe positiva, mas é atribuída à negativa; As terminologias observadas são ilustradas na Tabela 3.

Em sistemas multiclases, a matriz de confusão surge como um artifício para identificação de conflitos e pontos falhos do sistema. O valor presente na diagonal principal reflete a taxa de acerto da classe em questão, enquanto que os

Tabela 4 – Matriz de confusão multiclasse para *fold* 1 da LMD, utilizando descritor LBP e regra da soma

	Axé	Bac.	Bol.	For.	Gaú.	Mer.	Pag.	Sal.	Sert.	Tan.
Axé	86,67	3,33	0,00	0,00	0,00	0,00	3,33	0,00	6,67	0,00
Bachata	0,00	93,33	3,33	0,00	0,00	0,00	0,00	3,33	0,00	0,00
Bolero	0,00	3,33	86,67	3,33	0,00	0,00	0,00	3,33	3,33	0,00
Forró	0,00	0,00	6,67	83,33	3,33	0,00	0,00	3,33	3,33	0,00
Gaúcha	3,33	0,00	0,00	6,67	73,33	3,33	0,00	0,00	13,33	0,00
Merengue	0,00	3,33	0,00	0,00	0,00	96,67	0,00	0,00	0,00	0,00
Pagode	10,00	0,00	3,33	0,00	0,00	3,33	66,67	6,67	6,67	3,33
Salsa	0,00	0,00	0,00	0,00	0,00	0,00	3,33	96,67	0,00	0,00
Sertaneja	16,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	83,33	0,00
Tango	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00

valores distribuídos na linha descrevem os desvios de classificação para cada uma das outras classes envolvidas. A tabela 4 apresenta uma matriz de confusão para um *fold* na base LMD, na qual é possível observar gêneros propensos a confusão com outros de características semelhantes, como é o caso do axé e do sertanejo.

No caso de um sistema que utilize o conceito de *cross validation*, o resultado final é calculado pela média entre todas as repetições do processo. Fica definida como repetição cada uma das variações na tomada de um dos *folders* como conjunto de teste e os demais como conjunto de treinamento.

A partir das terminologias obtidas com a construção da matriz é possível calcular taxas relacionadas aos resultados alcançados, de modo a avaliar o desempenho de classificação do sistema. Para os experimentos deste trabalho ainda foi considerado o *F Score*, uma medida de balanceamento entre as taxas de ocorrência de positivos verdadeiros e precisão do sistema. Precisão pode ser entendida como uma medida para avaliar a capacidade de acerto de um classificador, na qual um valor baixo dessa medida pode também indicar um número grande de falsos positivos.

4.7 Considerações

A metodologia apresentada propõe, essencialmente o experimento e aplicação de estratégias nas diferentes etapas do processo de classificação, visando a otimização das taxas de reconhecimento de gêneros musicais. Primeiramente a utilização de mecanismos de extração de características, para descrever os atributos observados na base de amostras, tal como os provenientes dos aspectos de textura presentes nas imagens de espectrogramas. Ainda, um mecanismo de análise acústica para descrever características presentes diretamente no sinal de áudio.

No que diz respeito ao processo de classificação, foram apresentados métodos que tratam da manipulação dos resultados obtidos a partir da composição de um conjunto de classificadores, por meio da combinação de saídas e seleção dinâmica de subconjuntos no decorrer da classificação. Faz parte do trabalho a aplicação de métodos conceituados na literatura (WOODS *et al.*, 1996), como parâmetro de comparação, em contraste com a proposta de adaptações no mecanismo de seleção de vizinhança no KNORA.

5 Experimentos

A experimentação descrita no presente capítulo trata dos resultados obtidos para um sistema de classificação com base nas metodologias mencionadas no capítulo 4. Inicialmente, na seção 5.1, são apresentadas as bases de dados utilizadas e suas características estruturais, bem como a composição de amostras distribuídas entre as classes.

Ja a seção 5.2 demonstra os resultados para sistemas baseados na geração de classificadores a partir de descritores individuais: Local Binary Pattern (LBP) e Local Phase Quantization (LPQ) para o domínio visual; Statistical Spectrum Descriptor (SSD) para características acústicas. Na seção 5.3 estão contidos os experimentos utilizando seleção de classificadores. De modo a estabelecer um parâmetro de comparação de resultados, na subseção 5.3.1 são apresentados os valores obtidos a partir da aplicação de métodos bem aceitos na literatura, no que diz respeito a seleção de classificadores.

A seguir, podem ser encontrados os experimentos que utilizam técnicas adaptadas do K-Nearest-ORAcles (KNORA) para seleção de classificadores no pós-processamento e atribuição de classes. Resultados com variações dessas técnicas com seleção dinâmica de vizinhança na subseção 5.3.2 e limitação da vizinhança por filtro de distância na subseção 5.3.3.

Na seção 5.4 são contemplados experimentos de combinação de descritores. Finalmente a seção 5.5 encerra o capítulo.

5.1 Bases de músicas

A metodologia proposta para a tarefa de reconhecimento de padrões no sistema de classificação foi aplicado em 6 diferentes distribuições de amostras, cada qual correspondente a uma base de dados utilizada. Na subseção 5.1.1 é discutido a GTZAN, uma base conceituada na literatura, no que diz respeito ao reconhecimento de gêneros musicais (TZANETAKIS; COOK, 2002). A seguir, em 5.1.2, detalhada a LMD, uma base latino-americana de músicas com rotulação utilizando critério

baseado em como a música é dançada. Já na subseção 5.1.3, é apresentada uma base de músicas africanas a ser utilizada com 4 tipos de classificações diferentes.

5.1.1 Gtzan

Apresentada nas experimentações de Tzanetakis e Cook (2002), a GTZAN é uma base de músicas bem reconhecida no que diz respeito à reconhecimento de gêneros musicais. A base consiste de 1000 faixas de áudio, de 30 segundos de duração cada, distribuídas entre 10 diferentes gêneros populares da época: Blues; Classical; Country; Disco; Hip Hop; Jazz; Metal; Pop; Reggae; e Rock. Inicialmente estabelecida para propósitos de experimentação pessoal, a base não teve critérios pré-estabelecidos de condicionamento ou padrões para captura para os sinais de áudio. Dessa forma as faixas são apresentadas numa enorme variedade de condições na composição das classes.

A GTZAN é uma base reconhecida e bem estudada na literatura, de forma que é referência e serve de parâmetro para os estudos subsequentes envolvendo classificação de gêneros. Suas músicas foram divididas para a realização dos experimentos em 10 *folds*, cada um contendo 100 músicas distribuídas igualmente entre os gêneros. No caso das abordagens do KNORA, que necessitam de um conjunto de validação, o total de *folds* serviu à esse propósito, de modo que, como a base não respeita o conceito de *artist filter* na sua composição, não há preocupação de similaridade pertinente à intérprete na escolha de vizinhança.

5.1.2 LMD

A LMD é uma base de músicas latino-americanas apresentada a partir do trabalho de Silla *et al.* (2008) utilizada para tarefas relacionadas à recuperação automática de informações musicais. Originalmente construída para experimentos de classificação de gêneros musicais. O critério estabelecido para rotulação das amostras se baseia na percepção humana de como cada uma das faixas é dançada. A principal motivação para o desenvolvimento da LMD foi a carência de bases de dados disponíveis que apresentassem um número representativo de exemplos e critérios consistentes de atribuição de classes (MCKAY *et al.*, 2006).

Essa base apresenta 3227 títulos distribuídos entre 10 diferentes gêneros musicais: Axé; Bachata; Bolero; Forró; Gaúcha; Merengue; Pagode; Salsa; Sertaneja; e Tango. Uma das preocupações durante o processo de desenvolvimento, foi que além da informação de gênero ao qual pertence cada amostra, informações adicionais de autoria e técnicas fossem agregadas, de modo que fosse possível a utilização da base para outras tarefas de reconhecimento baseadas em recuperação de informação. Para este trabalho é relevante destacar entre as informações adicionais, o uso dos dados referentes ao intérprete da faixa, de modo que possibilite a aplicação do *artist filter* na separação de amostras em *folds*.

Para que fosse respeitada a divisão considerando os dados referentes a artistas, no experimento foram selecionadas 900 músicas divididas igualmente entre 3 *folds* e balanceadas igualmente entre as classes. Dos demais títulos, 2311 foram utilizados para composição de um conjunto de auxiliar denominado conjunto de validação, utilizado para aplicação do método KNORA. As amostras remanescentes foram descartadas, por não atenderem ao critério estabelecido ao se considerar a forma de segmentação do sinal adotada.

5.1.3 Base de músicas africana

Idealizada como parte do processo descrito em Cornelis *et al.* (2005) para facilitar o uso de técnicas de recuperação de informação musical, a base africana utilizada nesta proposta é parte de um conjunto de faixas digitalizadas de uma coleção privada cedida pelo *Royal Museum of Central-Africa (RMCA)*, na Bélgica. Embora exista uma quantidade grande de informação vinculada à coleção, no que diz respeito a identidade original, geográfica e musical das faixas, nem todas as amostras tem a completude dos dados, impossibilitando seu uso em algumas tarefas de classificação.

No que diz respeito as 1022 faixas disponibilizadas, é possível descrever as amostras levando em consideração pelo menos 4 linhas de classificação: função; instrumentação; país; e etnia. Função descreve o propósito a qual é destinada uma determinada faixa. Instrumentação estabelece um filtro com base na combinação de instrumentos utilizados na execução da música. País refere-se a região de origem. E por fim, etnia está relacionado ao grupo étnico de onde a música foi coletada.

Embora as classificações não sejam necessariamente gêneros musicais, é cabível a abstração como um cenário de reconhecimento de padrões e classificação, com uso da mesma metodologia aplicada nas outras bases. No que diz respeito ao experimento, cada uma das classificações pertinentes ao conjunto de amostras da base africana foi considerada como uma base independente, totalizando 4 bases utilizadas ao final da experimentação.

5.2 Classificação com único descritor

De modo a estabelecer valores utilizados como parâmetro para aplicação posterior das técnicas de manipulação dos classificadores, foi estabelecido, para cada uma das bases selecionadas, um sistema baseado em extração de características a partir de um único descritor, de textura para o domínio visual e de características acústicas diretamente no sinal de áudio. A classificação foi realizada com base na aplicação de validação cruzada para geração dos conjuntos de treinamento e teste. No caso da LMD foi empregada a distribuição de músicas respeitando o conceito de *artist-filter*, já que eram disponibilizadas informações consistentes de autoria para essa base. Na extração de características foram utilizadas estratégias para zoneamento horizontal com escala Mel, gerando assim um conjunto de classificadores. No caso da LMD e GTZAN foram estabelecidas 15 zonas, e para a base de músicas africanas somente 13, devido uma menor faixa de frequências presentes no sinal disponibilizado nessa base. Para a obtenção da decisão final foram consideradas as regras de soma e produto, descritas anteriormente na seção 3.2.1, por apresentarem resultados mais promissores.

5.2.1 LBP

A variação do método utilizada nesse trabalho é o $LBP_{8,2}$ com vetores de 59 atributos, que atua em uma região composta por 8 vizinhos a uma distância de 2 unidades do pixel analisado. No caso dos pontos de coordenadas intermediárias na diagonal em relação ao pixel central, se fez necessária uma interpolação dos valores ao redor, de modo a obter um pixel virtualizado com valor intermediário

dos pixels interpolados para aplicação do método. Os resultados obtidos no sistema de classificação descrito é apresentado na tabela 5.

Tabela 5 – Resultados obtidos a partir da utilização do LBP_{8,2}

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	73,40	0,73	73,10	0,73
LMD	86,00	0,86	85,33	0,85
País	86,02	0,53	48,85	0,57
Função	46,51	0,31	2,04	0,01
Instrumento	71,32	0,56	8,82	0,10
Etnia	82,94	0,70	50,94	0,42

5.2.2 LPQ

Para o experimento foi utilizada uma parametrização de janela de dimensão de 7 de altura e largura no processo de varredura da imagem, definido por um vetor de 256 atributos, com base em experimentos da literatura com resultados promissores. Os resultados obtidos com a aplicação do método no sistema de classificação pode ser bservada na tabela 6

Tabela 6 – Resultados obtidos a partir da utilização do LPQ

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	66,00	0,67	10,10	0,02
LMD	83,89	0,84	18,67	0,15
País	86,74	0,37	48,70	0,26
Função	48,18	0,33	2,04	0,01
Instrumento	74,54	0,38	9,82	0,04
Etnia	81,13	0,60	50,09	0,15

A aplicação dos descritores de textura na base LMD se assemelha aos experimentos realizados e resultados obtidos por Nanni *et al.* (2016). Para a base GTZAN, a carência de resultados publicados com utilização de *artist filter* estabelece novos valores como parâmetro de comparação. No caso da base africana, o resultado obtido quando considerado país de origem das amostras supera o melhor resultado

apresentado por Lidy *et al.* (2010) para essa divisão de classes. A proximidade dos resultados alcançados com os melhores conhecidos até o momento valida os uso dos descritores, como ponto de partida, para proposição de estratégias mais elaboradas, a fim de obter melhora nas taxas de reconhecimento.

5.2.3 SSD

Para efeito comparativo, foi aplicado uma estratégia para extração de características acústicas nas duas bases mais populares desta proposta, de modo a tornar possível uma comparação direta com abordagens presentes na literatura, no que diz respeito a taxas de reconhecimento. Lidy e Rauber (2005) definem o SSD como um método capaz de monitorar a variações rítmicas de energia em determinadas faixas de frequência do sinal a partir de medidas estatísticas, para cada uma das bandas especificadas. Para a realização do experimento foi adotado um software desenvolvido pela Universidade Técnica de Viena, disponível em ifs.tuwien.ac.at/mir/, e o processo de classificação se deu da mesma forma que para os descritores de textura, exceto pela aplicação de regras de fusão para combinação de resultados, pois o método gerou um único classificador. Os resultados obtidos na aplicação do método é apresentado na tabela 7

Tabela 7 – Resultados obtidos a partir da utilização do SSD

Base	Acurácia	FScore	Kappa
GTZAN	54,32	0,47	0,43
LMD	67,44	0,67	0,64

Quando comparados aos resultados obtidos com os descritores de textura, o experimentou utilizando características extraídas pelo SSD apresentou taxas de reconhecimento insatisfatórias, ainda mais distantes dos valores ideais descritos pelo limite superior. No caso do método acústico, um dos fatores que contribuiu para as baixas taxas de reconhecimento foi a geração de um único classificador para as tomadas de decisão, com base na proposição de Surowiecki (2005), na qual um grupo de elementos apresenta melhores resultados do que cada um dos elementos individualmente.

5.3 Classificação com seleção de classificadores

Em sistemas de reconhecimento envolvendo um conjunto de classificadores, cada um desses elementos impacta diretamente na decisão final de atribuição de uma classe à amostra, com base na sua eficiência individual. Seleção de classificador trata da aplicação de estratégias delimitadoras do conjunto inicial de classificadores, de modo que seja dada mais importância àqueles que, aparentemente, contribuam melhor para a tomada de decisão. Dentre as modalidades estabelecidas para desenvolvimento dessas tarefas, foi utilizada na experimentação deste trabalho a seleção dinâmica de grupo de classificadores, na qual, é definido um subconjunto a partir do conjunto inicial para cada um das amostras analisadas, de acordo com as respectivas características observadas. O conjunto inicial foi obtido a partir da utilização da estratégia de zoneamento com escala mel. Foram estabelecidas zonas horizontais, no total de 15 zonas para os experimentos utilizando a LMD ou GTZAN e 13 zonas nos experimentos utilizando a base de músicas africanas.

5.3.1 Métodos clássicos

Dentre as publicações conceituadas na literatura, no que diz respeito a seleção de classificadores, o OLA e LCA são apresentados como abordagens de resultados consistentes para uma variedade de aplicações de reconhecimento de padrões. Para fim de comparação, essas abordagens foram aplicadas nas bases LMD e GTZAN, com utilização dos descritores de textura. Embora sejam vistos como métodos consistentes, os resultados obtidos nesses experimentos descrevem um desempenho retraído, se comparados a abordagens para tarefas semelhantes.

OLA

Para a realização desse experimento, foi utilizada uma implementação baseada no trabalho de Martins (2014), que define um pseudo-código para execução do método com critério OLA, conforme apresentado anteriormente na seção 3.2.2. Os resultados referentes ao método são apresentados nas tabelas 8 e 9.

A utilização de um critério para seleção dinâmica de classificadores permite que um sistema de classificação seja adaptável às características individuais de cada

Tabela 8 – Resultados obtidos a partir do LBP_{8,2} com seleção de classificadores OLA

Base	Acurácia	FScore
GTZAN	60,50	0,60
LMD	68,67	0,68

Tabela 9 – Resultados obtidos a partir do LPQ com seleção de classificadores OLA

Base	Acurácia	FScore
GTZAN	53,20	0,52
LMD	66,00	0,66

uma das amostras analisadas, porém o método OLA se baseia na escolha de um único classificador para a tomada de decisão. Embora seja tomado o classificador com melhor potencial de classificação, o método é afetado pelo problema de decisão individual, inerente às estratégias que utilizam somente um classificador como conjunto inicial. Por consequência desse critério de seleção, as regras de fusão não alteram as estimativas do único classificador utilizado.

LCA

A aplicação do LCA se baseia em um pseudo-código de mesma origem do método anterior, e seu funcionamento está descrito na seção 6. As tabelas 10 e 11 apresentam os resultados obtidos para o descritor LBP_{8,2} e LPQ respectivamente.

Tabela 10 – Resultados obtidos a partir do LBP_{8,2} com seleção de classificadores LCA

Base	Acurácia	FScore
GTZAN	46,70	0,46
LMD	67,67	0,67

Tabela 11 – Resultados obtidos a partir do LPQ com seleção de classificadores LCA

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	63,80	0,63	63,80	0,63
LMD	65,00	0,65	65,00	0,65

De forma semelhante ao método anterior, o resultado obtido pelo LCA sofre pela limitação de um único classificador na tomada de decisão. A fim de demonstrar que a aplicação dos métodos OLA e LCA não foi satisfatória, quando comparados aos resultados esperados como taxa ideal de acurácia, a tabela 12 apresenta os valores do limite superior para cada uma das bases, LMD e GTZAN.

Tabela 12 – Limite superior nas bases LMD e GTZAN para cada descritor de textura

	LBP_{8,2}	LPQ
GTZAN	95,40	95,80
LMD	98,89	98,11

5.3.2 KNORA

O método KNORA abrange uma estratégia para seleção dos classificadores baseada em uma vizinhança obtida dinamicamente com base em medidas de distância entre o vetor de características do elemento a ser classificado e os vetores presentes em uma base externa, denominada conjunto de validação. Variações dessa técnica, em combinação com o KNN, foram exploradas no trabalho de Vriesmann *et al.* (2012) para um domínio de aplicação que envolve reconhecimento de caligrafia. Embora não tenham sido apresentados ganhos significativos nesses experimentos, a utilização de um conjunto gerado dinamicamente para moderar o impacto individual dos classificadores com base em semelhança de características observadas em comparação a uma base já conhecida se mostra como uma lógica promissora a ser desenvolvida.

Para aplicação desse método o cenário fica dependente da existência de um conjunto de amostras de classificação previamente conhecida, tomado como conjunto de validação. Inicialmente são calculados quais dos classificadores corretamente atribuem classes para cada uma das amostras presente nesse conjunto. É definido um valor k que corresponde ao número de vizinhos selecionados dinamicamente a cada iteração do processo de classificação. A seleção de vizinhança é realizada com base no cálculo da distância Euclidiana entre o vetor de características de cada um dos classificadores associado à amostra a ser analisada e o classificador

correspondente de cada uma das amostras presente no conjunto de validação. As distâncias obtidas para cada par de classificadores entre a amostra analisada e a amostra de validação são somadas de modo a obter um valor de distância único entre as duas amostras. As k amostras mais semelhantes são tomadas como vizinhança e a combinação da quantidade dos classificadores que atribuem classes corretamente à esses vizinhos define o impacto de cada classificador na decisão final para a amostra em questão. No caso da presente proposta a definição da quantidade de vizinhos k se deu por meio da variação desse parâmetro sobre a base LMD para o descritor LBP. Os resultados obtidos para diferentes valores de k nessas condições podem ser observados na figura 23, para a regra da soma, e na figura 24, para o produto.

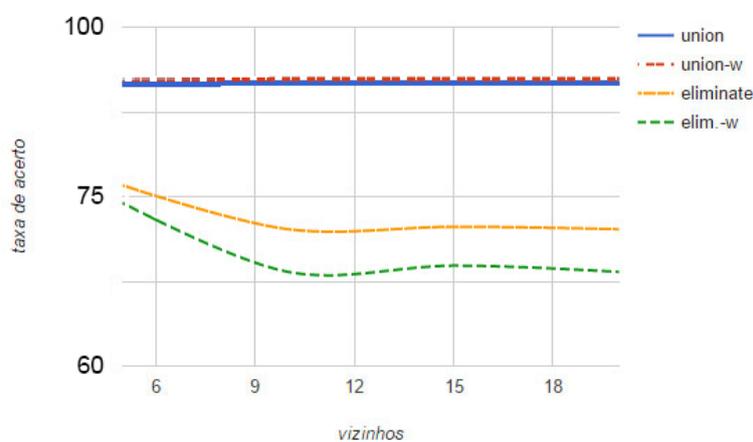


Figura 23 – Taxas de reconhecimento variando k vizinhos para variações do método KNORA aplicadas sobre o $LBP_{8,2}$ para regra da soma

Ficou definido em 10 o valor especificado de k nos demais experimentos. A utilização de vizinhanças maiores não só se mostrou mais extenso, no que se diz respeito a tempo de execução, como não contribui com a melhoria de acurácia final. No caso das variações ELIMINATE e ELIMINATE-W, uma vizinhança menor apresenta melhores resultados, a qual parte do princípio em que só é necessário um único classificador para a decisão correta, porém seus resultados ainda são inferiores às outras variações. A escolha do valor ainda considerou a possibilidade de otimização do método, com base na maior diferença encontrada entre os resultados obtidos dos métodos UNION e ELIMINATE.

Os resultados obtidos para o KNORA UNION podem ser observados nas

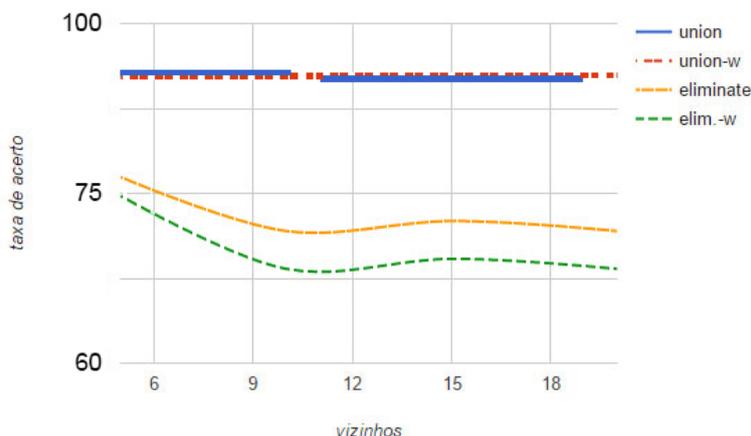


Figura 24 – Taxas de reconhecimento variando k vizinhos para variações do método KNORA aplicadas sobre o $LBP_{8,2}$ para regra do produto

tabelas 13 e 14. Para o KNORA UNION-W, nas tabelas 15 e 16. Já para para o KNORA ELIMINATE, os resultados são apresentados nas tabelas 17 e 18. E por fim a variação KNORA ELIMINATE-W nas tabelas 19 e 20, respectivamente para os descritores de textura LBP e LPQ.

Tabela 13 – Resultados obtidos a partir do LBP com KNORA UNION

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	73,70	0,73	73,20	0,73
LMD	85,44	0,85	85,33	0,85
País	86,63	0,40	34,57	0,17
Função	47,78	0,33	2,04	0,01
Instrumento	72,18	0,36	8,11	0,03
Etnia	84,98	0,67	48,71	0,13

A utilização do método KNORA na etapa de seleção de classificadores impõe um nível elevado de complexidade ao sistema de classificação, no qual se faz necessária a existência base de validação com amostras previamente classificadas. Ainda, para a construção de uma vizinhança, é realizado o cálculo da distância euclidiana entre o vetor de características da amostra analisada e o de cada uma das amostras presentes na base de validação.

Os resultados obtidos com esse método, em sua maioria, se mostram infe-

Tabela 14 – Resultados obtidos a partir do LPQ com KNORA UNION

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	72,60	0,72	72,70	0,72
LMD	84,22	0,84	85,11	0,85
País	86,94	0,38	41,28	0,23
Função	50,20	0,37	2,04	0,01
Instrumento	74,82	0,38	8,82	0,04
Etnia	81,95	0,61	47,50	0,13

Tabela 15 – Resultados obtidos a partir do LBP com KNORA UNION-W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	73,80	0,73	73,60	0,73
LMD	85,44	0,85	85,33	0,85
País	86,63	0,39	34,75	0,19
Função	47,77	0,33	2,04	0,01
Instrumento	72,18	0,36	8,11	0,03
Etnia	84,98	0,67	48,87	0,14

Tabela 16 – Resultados obtidos a partir do LPQ com KNORA UNION-W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	71,09	0,70	68,34	0,64
LMD	83,89	0,82	18,67	0,09
País	87,03	0,38	41,48	0,23
Função	50,07	0,37	2,04	0,01
Instrumento	74,96	0,38	8,82	0,04
Etnia	81,95	0,61	47,66	0,13

riores aos obtidos sem aplicação de estratégias de seleção, entretanto para a base africana, na classificação por país de origem e função, as valores obtidos superam as taxas apresentadas por Lidy *et al.* (2010), tomadas até o momento como as melhores alcançadas para a base.

Tabela 17 – Resultados obtidos a partir do LBP com KNORA ELIMINATE

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	63,10	0,62	62,80	0,62
LMD	71,67	0,72	72,11	0,72
País	88,15	0,50	3,31	0,03
Função	47,08	0,37	2,04	0,01
Instrumento	70,93	0,37	7,97	0,03
Etnia	87,40	0,73	29,87	0,09

Tabela 18 – Resultados obtidos a partir do LPQ com KNORA ELIMINATE

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	71,40	0,71	71,60	0,71
LMD	70,67	0,70	72,33	0,72
País	89,12	0,46	9,25	0,08
Função	50,28	0,39	2,04	0,01
Instrumento	74,19	0,40	8,40	0,03
Etnia	85,66	0,70	29,43	0,10

Tabela 19 – Resultados obtidos a partir do LBP com KNORA ELIMINATE-W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	59,20	0,59	59,20	0,59
LMD	68,33	0,68	68,33	0,68
País	88,06	0,49	3,31	0,03
Função	46,64	0,37	2,04	0,01
Instrumento	70,79	0,37	7,97	0,03
Etnia	86,94	0,72	29,87	0,09

5.3.3 KNORA com seleção de tamanho de vizinhança

Tendo em vista os resultados obtidos pela escolha do parâmetro k , apresentada na figura 24, é possível perceber que a vizinhança utilizada impacta significativamente nos resultados obtidos com o KNORA ELIMINATE e EMILINATE-W, embora não seja perceptível para a variação do KNORA UNION. Essa ocorrência demonstra que a definição do tamanho de vizinhança pode influenciar negativa-

Tabela 20 – Resultados obtidos a partir do LPQ com KNORA ELIMINATE-W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	65,03	0,61	59,00	0,57
GTZAN	71,40	0,71	71,60	0,71
LMD	83,89	0,84	18,67	0,15
País	88,75	0,44	9,25	0,08
Função	49,97	0,39	2,04	0,01
Instrumento	74,32	0,40	8,40	0,03
Etnia	85,66	0,70	29,43	0,10

mente nas taxas finais de reconhecimento, possivelmente pela carência de uma quantidade k vizinhos realmente semelhantes à amostra.

Com base no impacto observado pela definição de um tamanho fixo de vizinhança, esse experimento tomou por base a filtragem dos vizinhos significativos nas vizinhanças dinamicamente obtidas pelo KNORA. Essa filtragem se deu pelo cálculo de uma média das distâncias do vetor de características entre cada um dos vizinhos e a amostra, utilizada como limite de distância aceitável. Um determinado vizinho, cujo vetor de características correspondente não satisfaça a condição de distância máxima aceita, é descartado da vizinhança selecionada.

Os resultados obtidos com a aplicação do KNORA UNION são apresentados nas tabelas 21 e 22. Para o KNORA UNION-W, nas tabelas 23 e 24. Do KNORA ELIMINATE é apresentado nas tabelas 25 e 26. Já para o KNORA ELIMINATE-W, o resultado é apresentado nas tabelas 27 e 28.

Tabela 21 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA UNION

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	73,70	0,73	73,80	0,73
LMD	85,11	0,85	84,89	0,85

Para a construção de vizinhanças com tamanho definido pelas características da amostra analisada, se fez necessário estabelecer um critério de para seleção de vizinhos com potencial de contribuir para a classificação. O critério estabelecido foi

Tabela 22 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA UNION

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	73,70	0,73	73,80	0,73
LMD	83,56	0,84	18,67	0,15

Tabela 23 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA UNION W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	74,30	0,74	74,20	0,74
LMD	85,22	0,85	84,89	0,85

Tabela 24 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA UNION W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	74,30	0,74	74,20	0,74
LMD	83,56	0,84	18,67	0,15

uma média simples entre as distâncias de um total de n vizinhos mais próximos, tomada como o limiar de distância aceitável para que um vizinho fosse selecionado.

O resultados nesse experimento se mostraram inferiores aos obtidos pela utilização somente de descritores de textura, porém quando comparadas a variação do KNORA com vizinhança de tamanho variável e a correspondente, de tamanho fixo, é possível observar uma significativa melhora nos percentuais, como é o caso dos valores apresentados nas tabelas 27 e 19. Com base nessa comparação, a escolha de um critério fraco para seleção dos vizinhos pode ter impactado no potencial de classificação da estratégia, comprometendo assim as taxas de reconhecimento.

5.4 Classificação com combinação de descritor

Com base no que foi observado acerca da complementaridade de estratégias, no que diz respeito a diferentes erros cometidos para um mesmo conjunto de amos-

Tabela 25 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	66,20	0,66	66,20	0,66
LMD	77,00	0,77	77,67	0,78

Tabela 26 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	66,20	0,66	66,20	0,66
LMD	74,22	0,71	0,76	0,02

Tabela 27 – Resultados obtidos a partir do LBP com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE-W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	65,10	0,65	65,20	0,65
LMD	75,44	0,75	75,89	0,76

tras detalhado na seção 4.6.1, foram realizados experimentos combinando as saídas dos métodos de melhores resultados apresentados anteriormente na subseção 5.2. Na subseção 5.4.1 é apresentado o experimento somando as estimativas dos classificadores do primeiro método com as estimativas do classificador correspondente no segundo e posteriormente aplicada a regra de fusão. Já na subseção 5.4.2 é realizada uma combinação de saída por meio da omissão das estimativas do método que apresentar o menor valor para a classe mais provável, assim acredita-se que é tomada a decisão somente do melhor classificador de cada par. A extração de características se fez com base na estratégia de zoneamento com escala Mel, com 15 zonas para as bases LMD e GTZAN e 13 zonas para a base de músicas africanas.

5.4.1 LBP e LPQ com soma de estimativas finais

Para este experimento foram tomados os vetores contendo as estimativas de cada uma das classes envolvidas na classificação para duas abordagens, a primeiro

Tabela 28 – Resultados obtidos a partir do LPQ com seleção dinâmica de tamanho de vizinhança no KNORA ELIMINATE-W

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	65,10	0,65	65,20	0,65
LMD	74,22	0,76	10,00	0,02

utilizando LBP e o segundo LPQ, em cada uma das regras de fusão abordadas. Para cada par de classificadores correspondentes entre os métodos, os vetores de estimativas foram somados, de modo que fosse obtido um novo vetor. A decisão final foi a obtida a partir do novo vetor com aplicação das regras de fusão e a tabela 29 demonstra os resultados obtidos no experimento.

Tabela 29 – Resultados obtidos a partir da soma das porcentagens estimadas para cada classe em um sistema LBP com os respectivos valores do LPQ

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	71,90	0,72	72,10	0,72
LMD	87,78	0,88	85,33	0,85
País	86,02	0,35	48,85	0,27
Função	46,51	0,31	2,04	0,01
Instrumento	71,32	0,36	8,82	0,04
Etnia	82,94	0,64	50,94	0,15

O resultado observado a partir da combinação dos vetores de estimativas dos classificadores do LBP e LPQ, em sua maioria, não superou o obtido por cada um dos métodos individualmente, porém o valor obtido pela regra da soma apresentado na tabela 29, superou as melhores taxas da literatura, apresentada por Nanni *et al.* (2016) para a base LMD.

5.4.2 LBP e LPQ com comparação individual de classificador

Baseado na idéia de que alguns classificadores de uma das abordagens poderia contribuir negativamente na decisão final em uma determinada amostra quando somadas as estimativas, foi tomado no experimento seguinte somente o

vetor de saída do classificador que contivesse o maior valor entre as estimativas, para cada par formado pelo classificador do primeiro método e o correspondente no segundo. Esse classificador é tomado como o de maior potencial na atribuição de classes e contribui na decisão final com aplicação das regras de fusão, enquanto que as estimativas do outro classificador do par são descartadas. Os resultados obtidos com a aplicação dessa proposição pode ser observado na tabela 30

Tabela 30 – Resultados obtidos a partir da estimativa de maior valor entre as saídas dos sistemas com LBP e LPQ

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	73,40	0,73	73,10	0,73
LMD	87,11	0,87	18,67	0,15
País	86,02	0,35	48,85	0,27
Função	46,51	0,31	2,04	0,01
Instrumento	71,32	0,36	8,82	0,04
Etnia	82,94	0,64	50,94	0,15

De forma semelhante a abordagem com soma do vetor de estimativas, o resultado obtido pelo experimento apresentou valores próximos ou iguais aos obtidos individualmente pelos métodos e uma melhora mais discreta na taxa de reconhecimento para a base LMD. A semelhança de resultados entre as duas abordagens, levando em consideração que na segunda metade das estimativas envolvidas foram descartadas leva a crer na existência de classificadores que não contribuem para a decisão final, podendo assim serem descartados no decorrer das etapas de classificação com a proposição de um critério de seleção.

5.4.3 Combinação de LBP e LPQ com SSD

A fim de avaliar a possibilidade de que um classificador tomado como ruim individualmente poderia contribuir para uma melhoras nas taxas de reconhecimento quando combinado, foram tomadas as combinações dos descritores de textura com SSD para as bases LMD e GTZAN, por apresentarem uma distribuição equilibrada entre as músicas de cada gênero.

Tabela 31 – Resultados obtidos a partir da soma de estimativas das saídas dos sistemas com LBP e SSD

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	64,77	0,65	64,50	0,64
LMD	75,89	0,76	75,00	0,75

Tabela 32 – Resultados obtidos a partir da soma de estimativas das saídas dos sistemas com LPQ e SSD

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	47,64	0,48	7,21	0,01
LMD	60,56	0,61	64,33	0,65

Tabela 33 – Resultados obtidos a partir da soma de estimativas das saídas dos sistemas com LBP, LPQ e SSD

Base	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
GTZAN	70,37	0,71	69,70	0,70
LMD	82,11	0,82	79,89	0,80

Para extração de características nos descritores de textura foi aplicada uma estratégia de zonaneamento com escala Mel, gerando assim 15 classificadores para o sistema, enquanto que para SSD foi gerado somente um classificador. A fim de contornar essa diferença estrutural na etapa de combinação de classificadores, o vetor de estimativas do único classificador gerado pelo SSD foi pareado com cada um dos classificadores gerados pelos outros método utilizados. Desse modo, cada método envolvido no experimento teria as mesmas possibilidade de impactar na decisão final, independente da quantidade de classificadores gerados.

A tabela 31 apresenta os resultados obtidos a partir da combinação de classificadores do LBP e SSD. A tabela 32 apresenta os resultados da combinação de LPQ e SSD. Por fim, a tabela 33 apresenta o resultado obtido com a combinação dos classificadores pertencentes aos três métodos.

O desempenho individual do sistema de classificação utilizando o descritor de textura SSD se mostrou distante dos valores ideais descritos a partir do cálculo

do limite superior, porém com base na especulação de que classificadores de métodos distintos quando combinados podem resultar em melhores taxas de reconhecimento se fez plausível a verificação do impacto na tomada de decisão. O resultado obtido pela combinação em pares, ou trio com os descritores de textura obteve resultados menos expressivos do que os obtidos pela combinação somente do LBP e LPQ, decorrente do baixo desempenho individual do descritor acústico.

5.4.4 Limite Superior

Tomado como referência, o limite superior é uma estimativa de potencial a ser alcançado pelo conjunto de classificadores em um cenário ideal. Entende-se que uma amostra é passível de classificação se ao menos um dos classificadores envolvidos no sistema consegue atribuir corretamente na tomada de decisão a classe a qual verdadeiramente pertence a amostra analisada. O potencial calculado pelo limite superior, para cada uma das bases, é apresentado na tabela 34.

Tabela 34 – Limite superior para as base utilizadas nos experimentos para cada descritor de textura

	LBP_{8,2}	LPQ
GTZAN	95,40	95,80
LMD	98,89	98,11
País	99,41	99,61
Função	96,46	97,46
Instrumento	95,58	97,72
Etnia	99,68	99,68

5.4.5 KNORA aplicado na combinação de LBP e LPQ

Pelo resultado descrito na tabela 29 ter se mostrado promissor, quando comparado as melhores taxas já publicadas para a base LMD, foi proposto ainda um experimento envolvendo a aplicação do KNORA sobre os classificadores combinados dos métodos LBP e LPQ. A seleção de vizinhança para o KNORA se deu pelo soma da distância do vetor de características da base de validação e o vetor de cada um dos dois métodos e a abordagem utilizada para a combinação dos classificadores

Tabela 35 – Resultados obtidos com a aplicação do KNORA sobre a combinação de classificadores dos métodos LBP e LPQ na base LMD

Método	Regra da Soma		Regra do Produto	
	Acurácia	FScore	Acurácia	FScore
UNION	87,00	0,87	87,33	0,87
UNION-W	87,89	0,88	87,89	0,88
ELIMINATE	76,00	0,76	76,56	0,76
ELIMINATE-W	85,11	0,85	20,89	0,19

foi a soma das estimativas de cada par de classificadores. A tabela 35 apresenta os resultados obtidos no experimento.

Embora o resultado obtido para a base LMD, descrito na tabela 35 pelo KNORA UNION-W, tenha se mostrado superior ao obtido anteriormente com a combinação de classificadores dos descritores de textura, esse método incorpora uma maior complexidade ao sistema. A aplicação do KNORA depende da existência de uma estrutura auxiliar para seleção de vizinhança, a base de validação, e do cálculo de distância euclidiana entre os elementos dessa estrutura e a amostra analisada.

5.5 Conclusão e avaliação de resultados

Foram apresentados no presente capítulo inicialmente os resultados obtidos com dois descritores de textura e um de características acústicas. Por apresentar desempenho bem inferior aos demais, demonstrado na tabela 7, o descritor de características acústicas foi descartado para experimentações posteriores. A partir dos conjuntos de classificadores associados à extração de características, foram aplicados dois métodos tomados como referência na literatura, no que diz respeito à incorporação de estratégias para seleção de subconjuntos de classificadores, subseção 5.3.1.

No teste de análise de desempenho do método KNORA, subseção 5.3.2, foram empregados os descritores de textura, que apresentaram bons resultados individualmente, e combinados com cada uma das quatro variações: UNION, UNION-W, ELIMINATE e ELIMINATE-W. Ainda, na subseção 5.3.3, foi apresentada uma adaptação do método com seleção dinâmica de tamanho da vizinhança utilizada.

Na comparação entre o desempenho das regras de fusão, a regra do produto na maioria da vezes apresentou estimativa muito inferior à regra da soma, como índices FScore muito ruins, próximos a zero para todos os experimentos que tiveram a base africana envolvida, como observado na tabela 18.

Foram realizados os experimentos envolvendo a combinação dos vetores de estimativas dos classificadores de sistemas utilizando LBP ou LPQ. Posteriormente, a combinação desses valores com as estimativas do descritor de características acústicas SSD. A combinação de métodos de classificação parte do princípio da complementaridade, na qual, classificadores são ditos complementares se não cometem os mesmos erros na tomada de decisão. Na combinação com o descritor acústico, as taxas obtidas não se mostraram promissoras, decorrente do impacto causado pelo baixo desempenho individual desse descritor. Já nos resultados obtidos pela combinação dos descritores de textura, como descrito na tabela 29, as taxas foram superiores as obtidos por cada um dos métodos individualmente, tabela 5.

Por fim, a aplicação das variações do KNORA sobre a combinação dos descritores de textura LBP e LPQ, resultando na melhor taxa de reconhecimento presente neste trabalho com o KNORA UNION-W, apresentado na tabela 35. Embora o resultado obtido com a aplicação do KNORA tenha se mostrado ligeiramente superior à combinação dos descritores de textura, a complexidade incorporada ao sistema com a estratégia de seleção não justifica o ganho pouco significativo,

Ainda que tanto as proposições fazendo uso do KNORA quanto as de classificadores combinados tenham obtido resultados comparáveis aos presentes na literatura, na forma de taxas de reconhecimento, o cálculo do limite superior em todas as bases envolvidas na proposta ainda se mostra distante dos resultados alcançados. O limite superior, apresentado na tabela 34, serve de referência para o resultado da aplicação dos métodos em um cenário ideal, no qual o classificador que consegue classificar corretamente a amostra é sempre selecionado para a decisão da classe atribuída. E ainda garante a supremacia da sua decisão. Ciente dessa garantia imposta pelo limite superior e pela dificuldade de construir um classificador universal, capaz de atuar em qualquer cenário com mesmo desempenho, é possível afirmar que o potencial realista de reconhecimento é inferior ao calculado e depende de fatores externos ao sistema para que seja atingido, tal como a qualidade da informação presente extraída das amostras analisadas.

Tabela 36 – Melhores resultados apresentados na experimentação e seus correspondentes na literatura

	Melhor resultado	Método utilizado	Literatura
GTZAN	74,2%	LPQ seleção dinâmica de vizinhança KNORA UNION regra do produto	85,49% (SEYERLEHNER <i>et al.</i> , 2010)
LMD	87,89%	KNORA UNION-W regra da soma sobre LBP + LPQ	86,1% (NANNI <i>et al.</i> , 2016)
País	89,1%	LPQ KNORA ELIMINATE regra da soma	82,21% (LIDY <i>et al.</i> , 2010)
Função	50,2%	LPQ KNORA ELIMINATE regra da soma	48,25% (LIDY <i>et al.</i> , 2010)
Instrumento	74,9%	LPQ KNORA UNION-W regra da soma	69,06% (LIDY <i>et al.</i> , 2010)
Etnia	87,4%	LBP KNORA ELIMINATE regra da soma	88,57% (LIDY <i>et al.</i> , 2010)

Diferente das bases mais presentes na literatura para tarefas de reconhecimento de padrões, como é o caso da GTZAN, a base africana se beneficiou da utilização do KNORA, obtendo para todas as quatro estruturas de categorização taxa de reconhecimento próximas ou superiores as de resultados já publicados. A aplicação dos métodos que alcançaram melhores resultados para a base africana podem ser encontrados nas tabelas 18, 15 e 17. O desempenho mais relevante apresentado neste capítulo se deve à combinação dos classificadores associados aos descritores de textura, tabela 29. O valor obtido para a LMD é comparável a outros já descritos na literatura, levando em consideração as características incorporadas no sistema de classificação, como restrição do *artist filter*.

6 Considerações Finais

Neste trabalho apresentamos uma série de abordagens para classificação de gêneros musicais a fim de otimizar as estimativas previstas pelos classificadores na tomada de decisão.

Inicialmente foram definidas características para o domínio visual, na forma de atributos de textura em espectrograma. Já para o domínio acústico as características foram extraídas por meio de descritores aplicados diretamente no sinal de áudio. Os descritores de características foram avaliados individualmente, seção 5.2, de modo que as experimentações seguintes se beneficiassem dos componentes mais promissores. O método KNORA representa um mecanismo de seleção de classificadores, na forma de manipulação do impacto das estimativas de um subconjunto de classificadores na tomada de decisão, com base na similaridade entre a amostra e valores conhecidos já classificados em uma base de validação. Embora tenha sido bem explorado em publicações recentes (VRIESMANN *et al.*, 2012), foi observada a possibilidade de adaptação do critério de construção de vizinhança, uma variação não experimentada do método baseada no princípio de que um número pré estabelecido de vizinhos selecionado não necessariamente contribue para aplicação de peso nos classificadores. Assim se mostra cabível gerenciar a vizinhança dinamicamente construída a partir de critérios de aceitação, como apresentado na subseção 5.3.3.

Na sequência, seção 5.4, foi tratado o conceito de complementaridade entre classificadores de métodos distintos, experimentos nos quais os métodos combinados são capazes de produzir melhores resultados do que cada um individualmente. Foram apresentadas duas abordagens para combinação dos sistemas baseado em único descritor de característica. A primeira, a partir da soma dos vetores de estimativas de cada um dos classificadores com o vetor do classificador correspondente no outro método. Combinadas as estimativas são aplicadas as regras de fusão para obter a decisão final. Na segunda abordagem é selecionado somente o vetor de estimativas do classificador de maior acurácia para cada um dos pares de classificadores entre os métodos utilizados. Selecionadas as estimativas dos classificadores mais acurados entre os métodos envolvidos são aplicadas as regras de fusão para atribuição de

uma classe para a amostra.

Considerando que no experimento no qual foram combinados classificadores de dois diferentes métodos, os resultados obtidos com a omissão do classificador menos acurado em cada par, apresentados na tabela 30, são próximos aos obtidos com a soma de estimativas de todos os classificadores envolvidos, tabela 29, é possível levantar a hipótese de que existam classificadores que não contribuam para obter uma decisão associada à amostra. Com base nessa suspeita, propõe-se como trabalho futuro a exploração de estratégias de seleção de classificadores a fim de verificar a contribuição individual quando houver combinação de métodos complementares.

Ainda como trabalho posterior, a possibilidade de desenvolver critérios para gerenciamento de vizinhança dinâmica no KNORA para seleção de vizinhos. Essa proposta se baseia no fato de que a utilização de um tamanho fixo de vizinhança para um conjunto de amostras analisadas obriga na seleção de vizinhos que sejam utilizadas amostras na base de validação que prejudiquem as estimativas obtidas para decisão final na etapa de seleção de classificadores.

Referências

- BONET, J. S. D.; VIOLA, P. A. Structure driven image database retrieval. In: *NIPS*. [S.l.: s.n.], 1997. Citado na página 52.
- BREIMAN, L. Bias, variance, and arcing classifiers. Tech. Rep. 460, Statistics Department, University of California, Berkeley, CA, USA, 1996. Citado 2 vezes nas páginas 36 e 60.
- CORNELIS, O.; CALUWE, R. D.; TRÉ, G. D.; HALLEZ, A.; LEMAN, M.; MATTHÉ, T.; MOELANTS, D.; GANSEMANS, J. Digitisation of the ethnomusicological sound archive of the royal museum for central africa (belgium). *IASA journal*, n. 26, p. 35–43, 2005. Citado na página 68.
- COSTA, C. H.; JR, J. D. V.; KOERICH, A. L. Automatic classification of audio data. In: IEEE. *Systems, Man and Cybernetics, 2004 IEEE International Conference on*. [S.l.], 2004. v. 1, p. 562–567. Citado na página 50.
- COSTA, Y.; OLIVEIRA, L.; KOERICH, A.; GOUYON, F. Music genre recognition based on visual features with dynamic ensemble of classifiers selection. In: IEEE. *Systems, Signals and Image Processing (IWSSIP), 2013 20th International Conference on*. [S.l.], 2013. p. 55–58. Citado 3 vezes nas páginas 23, 25 e 58.
- COSTA, Y.; OLIVEIRA, L.; KOERICH, A.; GOUYON, F. Music genre recognition using gabor filters and lpq texture descriptors. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. [S.l.]: Springer, 2013. p. 67–74. Citado 4 vezes nas páginas 25, 29, 54 e 57.
- COSTA, Y. M. Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores. 2013. Citado na página 14.
- COSTA, Y. M.; OLIVEIRA, L.; KOERICH, A. L.; GOUYON, F.; MARTINS, J. Music genre classification using lbp textural features. *Signal Processing*, Elsevier, v. 92, n. 11, p. 2723–2737, 2012. Citado 3 vezes nas páginas 23, 25 e 55.
- DESHPANDE, H.; SINGH, R.; NAM, U. Classification of music signals in the visual domain. In: SN. *Proceedings of the COST-G6 Conference on Digital Audio Effects*. [S.l.], 2001. p. 1–4. Citado na página 19.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: *Multiple classifier systems*. [S.l.]: Springer, 2000. p. 1–15. Citado 5 vezes nas páginas 17, 36, 37, 48 e 56.

DOUGHERTY, G. *Pattern Recognition and Classification: An Introduction*. [S.l.]: Springer New York, 2012. (SpringerLink : Bücher). ISBN 9781461453239. Citado na página 50.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012. Citado 2 vezes nas páginas 16 e 26.

EFRON, B.; EFRON, B. *The jackknife, the bootstrap and other resampling plans*. [S.l.]: SIAM, 1982. Citado na página 59.

GEORGE, J.; SHAMIR, L. Computer analysis of similarities between albums in popular music. *Pattern Recognition Letters*, Elsevier, v. 45, p. 78–84, 2014. Citado na página 51.

GIACINTO, G.; ROLI, F. Methods for dynamic classifier selection. In: IEEE. *Image Analysis and Processing, 1999. Proceedings. International Conference on*. [S.l.], 1999. p. 659–664. Citado na página 57.

GJERDINGEN, R. O.; PERROTT, D. Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, Taylor & Francis, v. 37, n. 2, p. 93–100, 2008. Citado na página 20.

GONZALEZ; WOODS, R.; EDDINS, S. *Digital Image Processing Using MATLAB*. [S.l.]: Pearson Education, 2004. ISBN 9788177588989. Citado 2 vezes nas páginas 27 e 29.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, n. 6, p. 610–621, 1973. Citado na página 28.

JAIN, A. K.; DUIN, R. P.; MAO, J. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 22, n. 1, p. 4–37, 2000. Citado na página 36.

KITTLER, J.; HATEF, M.; DUIN, R. P.; MATAS, J. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 20, n. 3, p. 226–239, 1998. Citado 3 vezes nas páginas 22, 37 e 59.

KO, A. H.; SABOURIN, R.; JR, A. S. B. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, Elsevier, v. 41, n. 5, p. 1718–1731, 2008. Citado 7 vezes nas páginas 15, 40, 41, 43, 45, 46 e 58.

KOERICH, A. L.; POITEVIN, C. Combination of homogeneous classifiers for musical genre classification. In: IEEE. *Systems, Man and Cybernetics, 2005 IEEE*

International Conference on. [S.l.], 2005. v. 1, p. 554–559. Citado 3 vezes nas páginas 15, 21 e 35.

LI, T.; OGIHARA, M. Music genre classification with taxonomy. In: IEEE. *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05)*. IEEE *International Conference on*. [S.l.], 2005. v. 5, p. v–197. Citado 2 vezes nas páginas 22 e 25.

LI, T.; OGIHARA, M.; LI, Q. A comparative study on content-based music genre classification. In: ACM. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. [S.l.], 2003. p. 282–289. Citado 3 vezes nas páginas 20, 22 e 25.

LIDY, T.; RAUBER, A. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: *ISMIR*. [S.l.: s.n.], 2005. p. 34–41. Citado 3 vezes nas páginas 25, 34 e 71.

LIDY, T.; SILLA, C. N.; CORNELIS, O.; GOUYON, F.; RAUBER, A.; KAESTNER, C. A.; KOERICH, A. L. On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections. *Signal Processing*, Elsevier, v. 90, n. 4, p. 1032–1048, 2010. Citado 5 vezes nas páginas 24, 25, 71, 77 e 88.

MARTINS, J. G. Identificação de espécies florestais utilizando seleção dinâmica de classificadores no espaço de dissimilaridade. 2014. Citado 4 vezes nas páginas 42, 43, 44 e 72.

MCKAY, C.; MCENNIS, D.; FUJINAGA, I. A large publicly accessible prototype audio database for music research. In: *ISMIR*. [S.l.: s.n.], 2006. p. 160–163. Citado na página 67.

NANNI, L.; COSTA, Y. M.; LUMINI, A.; KIM, M. Y.; BAEK, S. R. Combining visual and acoustic features for music genre classification. *Expert Systems with Applications*, Elsevier, v. 45, p. 108–117, 2016. Citado 6 vezes nas páginas 24, 25, 55, 70, 82 e 88.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, v. 29, n. 1, p. 51–59, jan. 1996. ISSN 00313203. Citado na página 29.

OJANSIVU; VILLE; HEIKKILÄ; JANNE. Blur insensitive texture classification using local phase quantization. In: *Image and signal processing*. [S.l.]: Springer, 2008. p. 236–243. Citado 2 vezes nas páginas 32 e 33.

PAMPALK, E.; FLEXER, A.; WIDMER, G. *et al.* Improvements of audio-based music similarity and genre classificaton. In: LONDON, UK. *ISMIR*. [S.l.], 2005. v. 5, p. 634–637. Citado na página 22.

PIETIKÄINEN, M.; HADID, A.; ZHAO, G.; AHONEN, T. *Computer Vision Using Local Binary Patterns*. [S.l.]: Springer London, 2011. ISSN 1381-6446. Citado na página 31.

PIETIKÄINEN, M.; OJALA, T.; XU, Z. Rotation-invariant texture classification using feature distributions. *Pattern Recognition*, v. 33, p. 43–52, 2000. Citado na página 32.

SEYERLEHNER, K.; SCHEDL, M.; POHLE, T.; KNEES, P. Using block-level features for genre classification, tag classification and music similarity estimation. *Submission to Audio Music Similarity and Retrieval Task of MIREX 2010*, 2010. Citado 3 vezes nas páginas 23, 25 e 88.

SHAH, M.; JETHAVA, G. B. A literature review on hand written character recognition. *Indian Streams Research Journal*, v. 3, n. 2, p. 1–19, 2013. Citado 2 vezes nas páginas 48 e 51.

SILLA, C. N.; KAESTNER, C. A.; KOERICH, A. L. Classificação ao de gêneros musicais utilizando vetores de característica híbridos. *13o Simpósio Brasileiro de Computacã Musical (SBCM 2011)*, p. 32–44, 2011. Citado 2 vezes nas páginas 23 e 56.

SILLA, C. N.; KOERICH, A. L.; KAESTNER, C. A. The latin music database. In: *ISMIR*. [S.l.: s.n.], 2008. p. 451–456. Citado 3 vezes nas páginas 22, 25 e 67.

STURM, B. L. A survey of evaluation in music genre recognition. In: *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation*. [S.l.]: Springer, 2012. p. 29–66. Citado na página 14.

SUROWIECKI, J. *The wisdom of crowds*. [S.l.]: Anchor, 2005. Citado 2 vezes nas páginas 21 e 71.

TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on, IEEE*, v. 10, n. 5, p. 293–302, 2002. Citado 8 vezes nas páginas 14, 19, 20, 21, 25, 34, 66 e 67.

UMESH, S.; COHEN, L.; NELSON, D. Fitting the mel scale. In: *IEEE. Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*. [S.l.], 1999. v. 1, p. 217–220. Citado na página 54.

VAPNIK, V. N.; KOTZ, S. *Estimation of dependences based on empirical data*. [S.l.]: Springer-verlag New York, 1982. Citado na página 58.

VRIESMANN, L. M.; JR, A. de S. B.; OLIVEIRA, L. E. S. de; SABOURIN, R.; KO, A. H.-R. Improving a dynamic ensemble selection method based on oracle information. *International Journal of Innovative Computing and Applications* 17, Inderscience Publishers Ltd, v. 4, n. 3-4, p. 184–200, 2012. Citado 2 vezes nas páginas 74 e 89.

VRIESMANN, L. M.; JR, A. S. B.; OLIVEIRA, L. S.; KOERICH, A. L.; SABOURIN, R. Combining overall and local class accuracies in an oracle-based method for dynamic ensemble selection. 2015. Citado 2 vezes nas páginas 24 e 25.

WANG, L.; HE, D.-C. Texture classification using texture spectrum. *Pattern Recognition*, v. 23, n. 8, p. 905 – 910, 1990. ISSN 0031-3203. Citado 2 vezes nas páginas 27 e 29.

WOODS, K.; BOWYER, K.; JR, W. P. K. Combination of multiple classifiers using local accuracy estimates. In: IEEE. *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*. [S.l.], 1996. p. 391–396. Citado 2 vezes nas páginas 41 e 65.