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

Diego Rafael Lucio

Classificação de espécies de pássaros utilizando descritores de características visuais e acústicas

Maringá
2016

Diego Rafael Lucio

Classificação de espécies de pássaros utilizando descritores de características visuais e acústicas

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)

L938c

Lucio, Diego Rafael
Classificação de espécies de pássaros utilizando descritores de características visuais e acústicas / Diego Rafael Lucio. -- Maringá, 2016.
70 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. Reconhecimento de gêneros musicais - Padrões (Informática) - Sistema de reconhecimento. 3. Recuperação de informação por conteúdo. 4. Aprendizagem de máquina. I. Costa, 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. 006.45

MN-003856

FOLHA DE APROVAÇÃO

DIEGO RAFAEL LUCIO

Classificação de espécies de pássaros utilizando descritores de características visuais e acústicas

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

Profa. Dra. Valéria Delisandra Feltrim
Universidade Estadual de Maringá – DIN/UEM

Prof. Dr. Carlos Nascimento Silla Junior
Pontifícia Universidade Católica do Paraná – PPGIA/PUCPR

Aprovada em: 26 de agosto de 2016.
Local da defesa: Sala 101, Bloco C56, *campus* da Universidade Estadual de Maringá.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me dado força e coragem, para realizar e concluir este trabalho.

A minha família pelo apoio e incentivo, sempre acreditando na minha capacidade, e em especial à meus pais, Ana Cristina Malaguti Lucio e Jazon Lucio Sobrinho, meus exemplos de determinação, amor e caráter.

À meu irmão, Felipe Matheus Lucio, que sempre estava ao meu lado quando precisei. Agradeço aos meu professores do Programa de Pós em Ciências da Computação, principalmente ao meu orientador, Professor Doutor Yandre Maldonado e Gomes da Costa, pela colaboração, atenção, paciência, contribuição, conhecimentos transmitidos e por ter depositado sua confiança em mim para a realização desta pesquisa.

Aos membros da banca pela disposição e participação. Aqueles que de alguma forma contribuíram para a realização deste trabalho.

A todos os meus familiares e amigos, não menos importantes, que, embora não tenham sido mencionados individualmente, têm toda minha gratidão.

Classificação de espécies de pássaros utilizando descritores de características visuais e acústicas

RESUMO

Este trabalho tem por finalidade apresentar um sistema para a classificação automática de espécies de pássaros baseado em características acústicas e visuais extraídas a partir do canto dos pássaros. As características visuais foram extraídas de espectrogramas gerados a partir dos cantos, enquanto as características acústicas foram extraídas diretamente do áudio. Descritores de textura foram usados para descrever o conteúdo do espectrograma, visto que este é o principal conteúdo visual encontrado neste tipo de imagem. Os operadores de textura utilizados foram *Local Binary Pattern* (LBP), *Local Phase Quantization* (LPQ), *Robust Local Binary Pattern* (RLBP), *Gray-Scale Level Co-occurrence Matrix* (GLCM) e Filtros de Gabor. As características acústicas, por sua vez, foram descritas utilizando *Rhythm Histogram* (RH), *Rhythm Patterns* (RP) e *Statistical Spectrum Descriptor* (SSD). Com o objetivo de realizar comparações mais precisas, os experimentos realizados utilizaram uma base de dados similar a utilizada em outros trabalhos. Na etapa de classificação, foi utilizado o classificador SVM e os resultados finais foram alcançados utilizando uma validação cruzada de 10 folds.

Palavras-chave: Processamento de Sinais; Reconhecimento de Padrões; Aprendizagem de Máquina; Identificação de Espécies de Pássaros; Recuperação de Informações e Extração de Informações

Bird species classification using visual and acoustic descriptors

ABSTRACT

This work aims at presenting a system for automatic bird species classification based on acoustic and visual features extracted from the birdsong. The texture features were extracted using: Local Binary Pattern (LBP), Local Phase Quantization (LPQ), Robust Local Binary Pattern (RLBP) Gray-Scale Level Co-ocurrence Matrix (GLCM) and Gabor filters. The acoustic characteristics are in turn extracted through the descriptors: Rhythm Histogarm (RH), Rhythm Patterns (RP) and Statistical Spectrum Descriptor (SSD.) Aiming to perform more fare comparisons, the experiments performed were made over a similar database used in the work Automatic Bird Species Identification for Large Number of Species (Lopes et al., 2011a). In the classification step, SVM classifier was used and the final results were taken by using 10-fold cross validation.

Keywords: Signal processing; Pattern recognition; Machine learning; Bird species classification; Spectrogam; Information retrieval; Information extraction

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

AR: Autoregressive
DFT: Discrete Fourier Transform
DTDMFCC: Dynamic Two-dimensional Mel-frequency Cepstral Coefficients
DTW: Dynamic Time Warping
GLCM: Gray-Scale Level Co-ocurrence Matrix
GMM: Gaussian Mixture Models
LBP: Local Binary Patterns
LBP: Local Phase Quantization
HMM: Hidden Markov Model
K-NN: K Nearest Neighbors
LDA: Linear Discriminat Analisys
MFCC: Mel-frequency Cepstral Coefficients
RH: Rhythm Histogram
RLBP: Robust Local Binary Patterns
RP: Rhythm Pattern
STFT: Short-time Fourier Transform
SVD: Singular Value Decomposition
SVM: Support vector machine
SOMeJB: SOM-enhaced JukeBox
SSD: Statistical Spectrum Descriptors
TDMFCC: Two-dimensional Mel-frequency Cepstral Coefficients
TDNN: Time Delay Neural Network
VQ: Vector Quantization

SUMÁRIO

1

Introdução

O interesse para com o reconhecimento automático de espécies de pássaros baseado em sua vocalização tem aumentado e vários trabalhos sobre esse assunto têm sido publicados (Fagerlund, 2007). Tal aumento se deve pelo fato de ser essencial o conhecimento da distribuição geográfica das espécies de pássaros para o desenvolvimento sustentável da humanidade, assim como também para a conservação da biodiversidade (Goëau et al., 2014).

Para manter a conservação da biodiversidade das espécies de pássaros é necessário obter o conhecimento exato da identidade e da evolução destas, além do conhecimento da distribuição geográfica supracitado, visto que as aves desempenham papéis de grande importância para o nosso ecossistema, tais como: controle de insetos (Holmes, 1990; Holmes et al., 1979), dispersão de sementes (Snow, 1971, 1981) e polinização (Carpenter, 1978; Feinsinger e Colwell, 1978; Proctor et al., 1996).

A identificação das espécies de pássaros é um problema bem conhecido dos ornitólogos (Lopes et al., 2011a). Para realizar o reconhecimento, especialistas sugerem técnicas não invasivas para coletar dados, tal como: Técnicas baseadas em bioacústicas, que tem por finalidade identificar espécies a partir de registros de áudio capturados na natureza (Bardeli et al., 2010).

Em contraponto à técnica citada acima, há as técnicas denominadas invasivas, assim como a rede de neblina, que consiste na utilização de uma rede normalmente feita de nylon suspensa entre dois polos, assemelhado-se a uma rede de vôlei de grandes dimensões, para a captura de pássaros tendo por objetivo realizar a classificação destes (Straube e Bianconi, 2014). Entretanto há alguns problemas com o uso da rede de neblina, visto que

o uso da mesma não garante que aves de todas as espécies presentes em uma região serão capturadas, assim como também pode, em alguns casos machucar as aves capturadas para a tarefa de classificação.

Nesse contexto, é oportuno o desenvolvimento de pesquisas relacionadas ao reconhecimento automático de espécies de pássaros, baseadas em técnicas não invasivas de registro de áudio, tendo como base estes registros para a criação de uma base de dados para o desenvolvimento de um sistema de classificação. Sendo assim, esse trabalho é voltado para a tarefa de classificação automática de espécies de pássaros fazendo o uso da combinação de características acústicas e visuais, obtidas a partir dos cantos dos pássaros.

As características visuais são obtidas a partir de um espectrograma, que é uma representação visual do espectro das frequências do som. No seu formato mais comum, é representado por um gráfico em que o eixo horizontal representa o tempo e o eixo vertical representa a frequência. A amplitude é representada em uma terceira dimensão, descrita pela intensidade da cor de cada ponto da imagem.

As características acústicas são obtidas diretamente a partir do sinal de áudio tendo por finalidade analisar a essência de uma amostra, para assim descrever características como: o ritmo, o timbre e o tom da mesma.

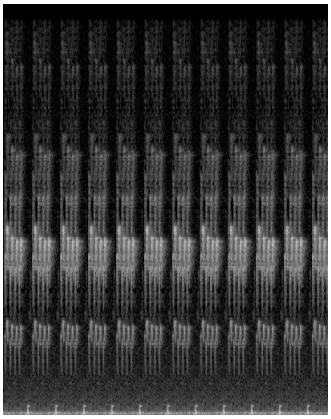
A Figura 1.1 ilustra dois exemplos de espectrogramas obtidos a partir do sinal de áudio das espécies *Automolus Leucophthalmus* e *Sittasomus Griseicapillus*, respectivamente. Como se pode observar as imagens geradas para cada uma das espécies refletem características diferentes, relacionadas inclusive a diferenças entre conteúdos harmônicos.

De acordo com a problemática supra citada foi definido que a hipótese deste trabalho é a de que é possível representar o canto de um pássaro por meio de características acústicas e visuais, com o propósito de criar um sistema de reconhecimento automático de espécies.

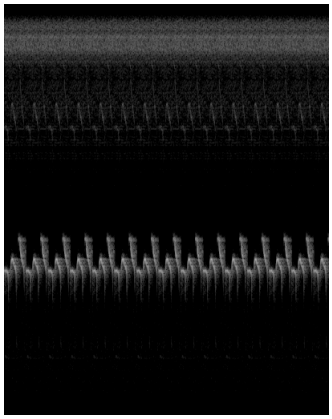
A utilização de características acústicas e visuais é justificada pelo fato de que cada tipo de característica pode capturar informações diferentes decorrente do fato de trabalharem em domínios diferentes, o que leva a crer que pode haver certa complementaridade entre as técnicas, o que pode acarretar a criação de uma metodologia de reconhecimento mais eficaz.

O desenvolvimento desse trabalho proporcionou algumas importantes contribuições no contexto de classificação automática de espécies de pássaros.

As primeiras contribuições apresentadas em (Lucio e Costa, 2015) mostraram que existem informações no conteúdo de textura presente nas imagens de espectrogramas gerados a partir do sinal do áudio com potencial para serem empregados com sucesso na tarefa de reconhecimento de espécies de pássaros. Neste trabalho, as características foram extraídas utilizando-se Filtros de Gabor, *Local Binary Pattern* (LBP) e *Local Phase*



(a) Espectrograma criado a partir de uma amostra de áudio da espécie de pássaro Automolus Leucophthalmus



(b) Espectrograma criado a partir de uma amostra de áudio da espécie de pássaro Sittasomus Griseicapillus

Figura 1.1: Exemplos de espectrogramas gerados a partir dos cantos dos pássaros

Quantization (LPQ) e, os resultados obtidos foram suficientes para confirmar a hipótese de que as imagens de espectrograma podem fornecer informações para suportar este tipo de tarefa. A partir disso, foram realizadas outras investigações variando a forma como o processo de classificação é configurado e o tipo de características utilizadas.

Em seguida foram realizados mais testes agora utilizando o *Gray Level Co-occurence Matrix* (GLCM) e o *Robust Local Binary Pattern* (RLBP) e foi constatado que esses descritores também apresentam potencial para representar o conteúdo de textura presente nas imagens de espectrogramas. Em relação os descritores de características visuais também foi constatado que o zoneamento da imagem pode proporcionar bons resultados.

Posteriormente foram realizados testes com os descritores acústicos *Rhythm Histogram* (RP), *Rhythm Pattern* (RH) e *Statistical Spectrum Descriptor* (SSD) e foi constatado que as informações referente a ritmo e timbre representadas pelos descritores, também são válidas para a tarefa de reconhecimento de espécies de pássaros.

Por fim se tem como a principal contribuição desse trabalho, a complementaridade encontrada quando utilizou-se as regras de fusão apresentadas por Kittler, Hatef, Duin e Matas para realizar a combinação dos classificadores criados com descritores acústicos e visuais.

Este trabalho encontra-se organizado da seguinte forma: no Capítulo 2 é descrita uma revisão bibliográfica acerca de classificação automática de espécies de pássaros; no Capítulo 3 são apresentados os principais fundamentos teóricos inerentes às técnicas

utilizadas para a realização dos experimentos; o Capítulo 4 descreve o método proposto para a construção do sistema de classificação de espécies de pássaros; o Capítulo 5 apresenta os resultados obtidos utilizando o protocolo para o reconhecimento de espécies de pássaros desenvolvido neste trabalho; o Capítulo 6 apresenta as conclusões obtidas com o desenvolvimento do trabalho até o presente momento.

2

Revisão da Literatura

O interesse no reconhecimento de espécies de pássaros baseado na sua vocalização tem aumentado e muitos estudos recentes têm sido publicados. O reconhecimento de espécies de pássaros é um problema típico de reconhecimento de padrões e a maioria dos estudos incluem processamento de sinais, extração de características e elaboração de sistemas de classificação.

Quanto as bases de dados utilizadas nos trabalhos apresentados neste Capítulo, não houve compartilhamento das mesmas entre os autores. Também não há um consenso sobre a forma como as amostras de áudio são obtidas, alguns trabalhos optaram por utilizar registros obtidos de forma profissional, enquanto outros optaram por trabalhar com registros de áudio obtidos de forma amadora.

Se entende como registros de áudio obtidos de forma profissional aqueles que foram adquiridos em um ambiente controlado sempre utilizando os mesmos parâmetros de gravação e sem qualquer interferência ou ruído proveniente de fatores ambientais. Por sua vez os registros amadores, são aqueles obtidos por entusiastas e criadores de pássaros, o que faz com que possa haver ruídos ruídos decorrentes de sons presentes em segundo plano.

Um fato que caracteriza os trabalhos que utilizaram os registros de áudio obtidos de forma profissional são as elevadas taxas de acerto obtidas, decorrente da ausência de ruído nas amostras de áudio. As seções seguintes apresentam respectivamente os trabalhos que fizeram o uso de registros de áudio obtidos de forma profissional e os trabalhos que utilizaram registros de áudio obtidos de forma amadora.

2.1 Trabalhos que Utilizaram Registros de Áudio Obtidos de Forma Profissional

Os trabalhos apresentados por Anderson et al. (1996) e Kogan e Margoliash (1998) estão entre as primeiras tentativas para o reconhecimento automático de espécies de pássaros por meio de sons emitidos pelos mesmos. O primeiro trabalha com *Dynamic Time Warping* (DTW) que é um algoritmo utilizado para comparar sequências que variam com o tempo, enquanto o segundo faz o uso da técnica citada, assim como também utiliza *Hidden Markov Models* (HMM) para o reconhecimento automático de duas espécies de pássaros. Em ambos os estudos são criados *templates* das amostras de áudio com base na vocalização das espécies, que é segmentada em sílabas, parágrafos e frases. Após a criação dos templates, os mesmos são utilizados como entrada em um sistema de *template matching*, que faz o uso de DTW e HMM para a classificação. Os resultados apresentados pelos trabalhos foram, respectivamente, 97,00% em uma base de dados composta por duas espécies e 82,00% em uma base de dados composta por seis espécies.

Selouani et al. (2005), Cai et al. (2007), e Chou e Liu (2009) fizeram o uso de redes neurais para realizar a classificação das espécies de pássaros. Selouani et al. (2005) utilizaram uma abordagem de rede neural chamada *Time Delay Neural Network* (TDNN) para realizar a classificação das espécies de pássaros. Como entrada do sistema foram utilizados *templates* extraídos dos registros de áudio das 16 espécies de pássaros presentes na base de dados utilizada, sendo que a melhor taxa de acerto foi de 83,00%.

O trabalho apresentado por Cai et al. (2007) utilizou *Mel-frequency Cepstral Coefficients* (MFCC) como características para o sistema de classificação para as duas bases de dados utilizadas: uma composta por 4 e a outra por 14 espécies de pássaros, as taxas de acerto para cada uma das bases foram, respectivamente, 98,70% e 86,80%.

Briggs et al. (2009) também utilizaram o MFCC como características das amostras de áudio. Além do MFCC, os autores também fizeram o uso do *Mean Frequency Bandwidth* (MFB) e da densidade do espectro como características. A etapa de classificação fez uso de K-NN, SVM e distâncias de KULLBAK-LEIBLER e HELLINGER, sendo que a melhor taxa de reconhecimento obtida para a classificação das 6 espécies presentes na base de dados foi de 91,10%.

No trabalho apresentado por Chou e Liu (2009) também foi feito o uso do MFCC como característica para as amostras de áudio das 420 espécies presentes na base de dados utilizada obtendo-se uma taxa de acerto de 18,72%.

Kwan et al. (2004) utilizaram *Guassian Mixture Model* (GMM) e também fizeram o uso de HMM para realizar a classificação das espécies. A base de dados utilizada pelos autores é composta por 4 espécies de pássaros, sendo que a melhor taxa de acerto do sistema de classificação proposto foi de 100%. O trabalho apresentado por Kwan et al. (2006) também utilizou GMM para a etapa de classificação, mas desta vez com uma base de dados composta por 11 espécies, que tiveram seus cantos representados por MFCC. Neste trabalho também foi apresentado um sistema para monitoramento automático de pássaros na natureza, com a taxa de acerto de 90%.

Chou et al. (2007) também fizeram o uso de HMM, assim como os trabalhos citados anteriormente, no entanto, este é utilizado como conjunto de características do sistema de classificação, e não como classificador. Os autores utilizaram uma base de dados composta por 420 espécies, sendo que a taxa de acerto alcançada foi de 78,20%.

Um outro exemplo da utilização do GMM na classificação foi apresentado por Lee et al. (2008), no qual foram utilizados como características o *Two-dimensional Mel-frequency Cepstral Coefficients* (TDMFCC), *Dynamic Two-dimensional Mel-frequency Cepstral Coefficients* (DTDMFCC) dos cantos das especies. A base utilizada no trabalho é composta de 28 espécies e a melhor taxa de acerto foi de 84,06%.

No trabalho apresentado por Tyagi et al. (2006) foi apresentada uma nova forma de representar as sílabas dos cantos dos pássaros, que tem como base a média do espectro do som sobre o tempo, e a classificação foi baseada na combinação de padrões. A taxa de acerto foi de 100% sobre uma base de dados de constituída por 15 espécies.

Vilches et al. (2006) apresentaram uma abordagem de classificação de espécies de pássaros baseadas em características acústicas, como: harmônia, timbre e ritmo. A base de dados utilizada era composta por três espécies de pássaros e a melhor taxa de acerto obtida foi de 98,39%.

Fagerlund (2007) também utilizou o MFCC das amostra de áudio como características para o sistema de classificação, no entanto, assim como Briggs et al. (2009), fez o uso de SVM para a classificação das amostras de áudio das duas bases de dados utilizadas no projeto: uma contendo 6 espécies e a outra 8 espécies. Os melhores resultados alcançados foram de 93,00% para a base composta por 6 espécies e de 97,00% para a base de dados composta por 8 espécies.

2.2 Trabalhos que Utilizaram Registros de Áudio Obtidos de Forma Amadora

McIlraith e Card (1997), assim como Selouani et al. (2005), Cai et al. (2007), e Chou e Liu (2009), fizeram o uso de redes neurais para realizar a classificação das espécies de pássaros. Nesse trabalho foram utilizadas como características parâmetros temporais e espectrais obtidos por meio da utilização da *Fast Fourier Transform* (FFT), sendo que sua melhor taxa de acerto foi de 82,00% em uma base de dados composta por seis espécies.

Lopes et al. (2011b) e Lopes et al. (2011a) fizeram uso de características acústicas das amostras de áudios das espécies de pássaros. As bases de dados utilizadas pelos autores são compostas por um subconjunto das amostras de áudio disponibilizadas pelo site Xeno-Canto[1]. O trabalho apresentado por Lopes et al. (2011a) utilizou 5 bases de dados compostas por 3, 5, 8, 12 e 20 espécies. Os resultados apresentados foram, respectivamente, 95,10%, 89,30%, 89,30%, 82,90% e 78,20% , para as respectivas bases de dados. Por sua vez, o trabalho apresentado por Lopes et al. (2011b) fez o uso de uma base de dados composta por 3 espécies e obtiveram como taxa de acerto os seguintes resultados, 99,70% e 98,39%, respectivamente.

2.3 Considerações Finais

O presente capítulo descreveu um histórico que mostra a evolução da pesquisa de classificação automática de espécies de pássaros, com base na sua vocalização. A partir da revisão realizada, foi construída a Tabela 2.1, na qual os trabalhos são apresentados em ordem cronológica. A tabela apresenta dados sobre os tipos de características utilizadas para a classificação, o mecanismo de classificação utilizado, o tipo de registro de áudio utilizado, a quantidade de espécies utilizadas e a melhor taxa de acerto obtida.

Tabela 2.1: Síntese da evolução do desenvolvimento de trabalhos em classificação de espécies de pássaros

Autores	**Ano**	**Características**	**Classificador**	**Tipo do áudio**	**Base/ espécies**	**Melhor Acerto**
Anderson et al.	1996	Templates criados a partir das amostras de áudio	HMM, DTW	Profissional	2 espécies	97,00%**

continua na próxima página

* Resultados apresentados com o uso de F-measure

** Resultado apresentado com o uso de acurácia

[1] http://www.xeno-canto.org/

continuação da Tabela Tabela 2.1

Autores	**Ano**	**Características**	**Classificador**	**Tipo do áudio**	**Base/ espécies**	**Melhor Acerto**
McIlraith e Card	1997	Características extraídas com FFT	Redes Neurais	Amador	6 espécies	82,00%**
Kogan e Margoliash	1998	Templates criados a partir das amostras de áudio	HMM, DTW	Profissional	2 espécies	98,70%**
Kwan et al.	2004	PCA, VQ	HMM, GMM	Profissional	4 espécies	100,00%**
Selouani et al.	2005	Templates criados a partir das amostras de áudio	TDNN com AR Backpropagation	Profissional	16 espécies	83,00%**
Kwan et al.	2006	MFCC	GMM	Profissional	11 espécies	90,00%**
Vilches et al.	2006	Características acústicas	VQ, ID3, J4.8 e Naive Bayes	Profissional	3 espécies	98,39%**
Tyagi et al.	2006	SEAV, GMM, DTW	Medida simples de distância euclidiana	Profissional	15 espécies	100,00%**
Fagerlund	2007	MFCC e parâmetros descritivos das sílabas	SVM utilizando classificação binária e por multiclasses	Profissional	6 espécies 8 espécies	93,00%** 97,00%**
Cai et al.	2007	MFCC	Redes Neurais	Profissional	4 espécies 14 espécies	98,70%** 86,80%**
Chou et al.	2007	HMM obtido sobre as sílabas	Algoritmo de Viterbi	Profissional	420 espécies	78,20%**
Lee et al.	2008	TDMFCC, DTDMFCC	GMM e VQ	Profissional	28 espécies	84,06%**
Briggs et al.	2009	Spectrum Density, Mean Frequency and Bandwidth, MFCC	K-NN, SVM, distância de KULLBAK-LEIBLER e distância de HELLINGER	Profissional	6 espécies	91,10%**
Chou e Liu	2009	MFCC	Redes neurais	Profissional	420 espécies	18,72%**
Lopes et al.	2011b	Características acústicas	Naive Bayes, KNN (k=3), J4.8, MLP, SMO(Polynomial), SMO(Pearson)	Amador	Xeno-Canto 3 espécies	98,39%*
Lopes et al.	2011a	Características acústicas	Naive Bayes, KNN (k=3), J4.8, MLP, SMO(Polynomial), SMO(Pearson)	Amador	Xeno-Canto 3 espécies 5 espécies 8 espécies 12 espécies 20 espécies	 95,10%* 89,30%* 89,30%* 82,90%* 78,20%*

* Resultados apresentados com o uso de F-measure

** Resultado apresentado com o uso de acurácia

O próximo cápitulo apresenta uma fundamentação teórica com conceitos que sustentam o desenvolvimento desta dissertação. Serão apresentadas algumas das principais abordagens presentes na literatura para a extração de textura em imagens digitais, assim

como também as principais abordagens utilizadas para a extração de características acústicas, a partir do sinal de áudio. Tais técnicas são potenciais candidatas para suportar a etapa de extração de características aqui proposto. Também é realizada a descrição de algumas das abordagens mais conhecidas para combinação das saídas dos classificadores.

3

Fundamentação Teórica

O problema de classificação pode ser definido como o processo pelo qual padrões ou sinais recebidos são distribuídos por um número prescrito de classes com aprendizado supervisionado. Estando presente em todas as áreas de atuação científica, a classificação representa um amplo conjunto de problemas de grande significado prático (Semolini, 2002).

É uma tarefa que o ser humano frequentemente executa sem dificuldades. Na qual dados são recebidos do mundo exterior por meio de nossos sentidos e assim realizamos o reconhecimento destes, em algum contexto. Essa tarefa é realizada de maneira quase que imediata praticamente sem nenhum esforço, caso o conhecimento necessário para executar a classificação já tenha sido adquirido por meio de um processo de aprendizagem (Semolini, 2002).

Todavia, nos casos em que a tarefa de classificação deve ser feita considerando dados pertencentes a espaços de grande dimensão e nos casos em que os atributos disponíveis para caracterizar cada amostra não esclarecem de forma óbvia o que diferencia um padrão pertencente a uma classe de outro pertencente a outra classe, o ser humano vai encontrar dificuldades para executar a classificação. Sendo assim, a automatização do processo de classificação passa a ser de grande interesse e a sua viabilidade aumenta conforme cresce o poder de processamento e a quantidade de memória dos computadores (Abe, 2010).

A abordagem clássica de um sistema de classificação é dividida em etapas bem definidas: pré-processamento, extração de características e classificação (Semolini, 2002)(Duda et al., 1973)(Duda et al., 2001), conforme ilustrado na Figura 3.1.

A etapa de pré-processamento compreende, em geral, tarefas como a segmentação do sinal, a fim de isolar as partes interessantes do mesmo. Adicionalmente, tarefas de

eliminação de ruídos também são comumente incluídas no pré-processamento, a fim de que a etapa de extração de características não seja afetada pelos mesmos. A etapa de extração de características depende fundamentalmente do tipo de sinal que está sendo processado, podendo este ser descrito das mais diversas formas. Imagens digitais e arquivos de áudio são exemplos de tipos de sinal que podem ser processados. Todavia é válido lembrar que cada tipo de sinal é melhor representado por descritores de características específicos. Uma imagem digital, por exemplo, necessita de descritores de atributos visuais como cor, textura e estrutura, entre outros, enquanto em um sinal de áudio as características são extraídas a partir das principais dimensões do sinal de áudio. Na última etapa, algoritmos de classificação são utilizados sobre os características extraídas a fim de se atribuir uma classe para cada padrão submetido ao sistema (Abe, 2010; Semolini, 2002).

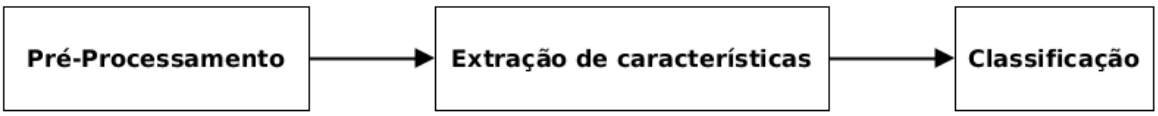


Figura 3.1: Etapas para o desenvolvimento de um sistema de reconhecimento de padrões

Cada uma das etapas apresentadas na Figura 3.1 possui uma grande quantidade de trabalhos. No entanto, na literatura, muitos esquemas diferentes vem sendo desenvolvidos para as etapas de extração de características e classificação, decorrente do fato dessa tarefa ser particularmente desafiadora.

Dos trabalhos para criação de um sistema de classificação de registros de áudio baseado em características visuais de textura extraídas a partir de um espectrograma, várias metodologias para a extração destas vem sendo aplicadas, tais como *Gray-Scale Level Co-ocurrence Matrix* (GLCM) (Haralick, 1979), *Local Binary Patterns* (LBP) (Ojala et al., 1996), *Robust Local Binary Patterns* (RLBP) (Chen et al., 2013), Filtros de Gabor (Li et al., 2010) e *Local Phase Quantization* (LPQ) (Ojansivu e Heikkilä, 2008). Além dos descritores citados também é utilizado o zoneamento das imagens para a extração de características locais, possibilitando a criação de um *pool* de classificadores. Para sistemas de classificação baseados em características acústicas tem tido grande destaque os descritores denominados *Rhythm Histogram* (RH) (Lidy e Rauber, 2005), *Rhythm Pattern* (RP) (Rauber e Frühwirth, 2001; Rauber et al., 2002) e *Statistical Spectrum Descriptors* (SSD) (Lidy e Rauber, 2005), que tem por finalidade obter informações relacionadas a harmônia, timbre e ritmo do sinal de áudio.

Em relação aos trabalhos atuais referentes a etapa de classificação, muitos tem utilizado o SVM como classificador (Briggs et al., 2009; Fagerlund, 2007). Tal escolha se

deve aos bons resultados obtidos com o mesmo em tarefas de classificação de de arquivos de áudio. Alguns trabalhos também apresentam a combinação entre classificadores por meio da utilização de regras de fusão. A representação de um sistema de fusão de classificadores pode ser vista na Figura 3.2 (Costa et al., 2013; Costa, 2013; Costa et al., 2011).

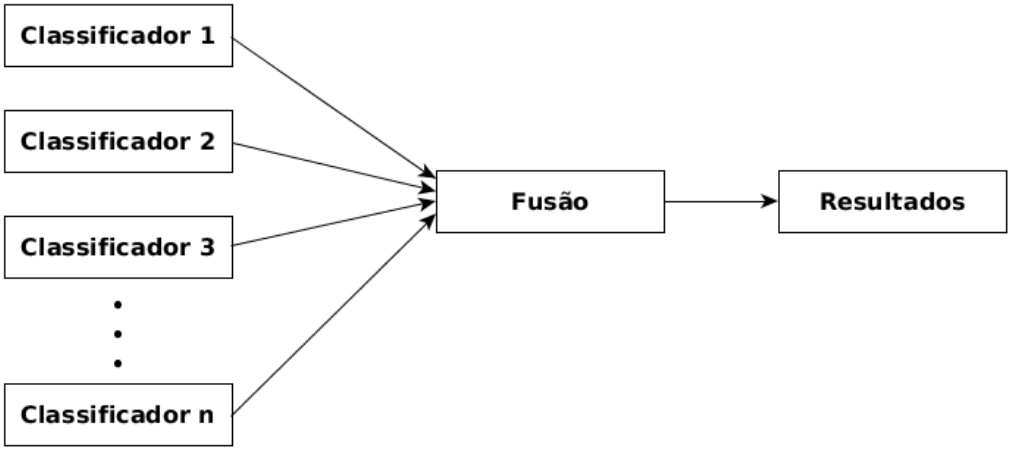


Figura 3.2: Esquematização da fusão dos resultados

Ainda sobre a etapa de classificação algumas medidas de avaliação são aplicadas sobre a fusão dos classificadores, com o objetivo de aferir o desempenho do sistema de classificação proposto (**?**)costa2013reconhecimento).

As próximas seções apresentam uma breve descrição dos conceitos de extração de características, zoneamento da imagem, classificação, fusão de classificadores e medidas de avaliação utilizados no desenvolvimento deste trabalho.

3.1 Extração de Características

A extração de características é uma etapa de grande importância para o desenvolvimento de um sistema de reconhecimento de padrões. E decorrente do objetivo deste trabalho estar relacionado à extração de características a partir do sinal do áudio dos sons de pássaros, sejam estas acústicas ou visuais, as seções seguintes apresentam algumas das abordagens apresentadas na literatura para a extração de características que podem ser utilizadas para a criação de um sistema de classificação para um domínio.

3.1.1 Características Acústicas

Rhythm Pattern

De acordo com Rauber e Frühwirth (2001) e Rauber et al. (2002), o *Rhythm Patterns* (RP) descrevem a amplitude da modulação para um intervalo de frequências de modulação presentes nas zonas críticas do sistema de audição humano. Os paragráfos seguintes descrevem o sistema de extração de características adotado pelo RP.

Em um primeiro momento, espectrogramas de segmentos de áudio de 44 kHz com duração de aproximadamente 6 segundos são processados utilizando a *Short-Time Fourier Transform* com uma janela de Hanning de 1024 amostras e um *overlap* de 50%.

Em seguida, a escala de Bark, uma escala contínua com grupos de frequência para zonas críticas para a audição humana, é aplicada ao espectrograma, agregando a este 24 zonas de frequência (Fastl e Zwicker, 2007).

Posteriormente, o espectrograma gerado pela escala de Bark é transformado na escala de Decibel, para em seguida aplicar as transformações psicoacústicas: É realizado então o cálculo da escala de Phon para incorporar curvas de ruído, que são utilizadas para calcular diferentes percepções de ruído de diferentes frequências, e transformações na escala de Sone, para calcular ruídos. O sonograma resultante da escala de Bark especifica a sensação de ruído de um seguimento de áudio para a audição humana (Fastl e Zwicker, 2007).

Em um segundo momento, a variação da energia nas zonas críticas do espectrograma na escala de Bark é considerada como uma modulação da amplitude do sinal sobre o tempo também conhecida como *cepstrum*, e é obtida por meio do uso da transformada de Fourier. O resultado é um sinal da invariante no tempo que contém a magnitude de modulação para frequência nas zonas críticas. Esta matriz representa um RP, indicando a ocorrencia de rítmo como barras verticais, mas também descrevendo pequenas flutuações em todas as zonas de frequência do sistema de audição humano.

Subsequentemente, modulações de aplitude são avaliadas de acordo com a função da sensação humana por meio da modulação da frequência acentuando valores em torno de 4 Hz, e eliminando frequência maiores que 10 Hz. A aplicação de um filtro gradiente e da suavização Gaussiana melhora a similaridade entre RPs. A matriz final de características com dimensões de 24×60 é computada pela mediana dos RP segmentados.

Rhythm Histogram

Rhythm Histogram (RH) agrega valores de modulação de amplitude de 24 zonas críticas individuais computadas por um *rhythm pattern*, apresentando a magnitude da modulação para 60 frequências de modulação entre $0,17$ e 10Hz (Lidy e Rauber, 2005).

O que acaba por caracterizar um descritor geral para características rítmicas em uma amostra de áudio. Um RH é computado para cada segmento de 6 segundos em uma amostra de áudio e o vetor de características é então calculado pela mediana dos dos valores calculados para cara um dos segmentos.

Statistical Spectrum Descriptors

O *Statistical Spectrum Descriptors* (SSD) é um descritor de características acústicas que tem por finalidade computar sensações de ruído específicas nas 24 zonas da escala de Bark, analogamente ao *Rhythm Patterns*. Subsequentemente, medidas estatísticas são calculadas para cada uma das zonas críticas, descrevendo assim variações em cada uma das zonas estatisticamente O SSD, assim, descreve flutuações nas zonas críticas e captura informação de timbre adicional que não foi coberto por outros conjuntos de características, como o *Rhythm Pattern*, dessa forma capturando e descrevendo muito bem conteúdo acústico (Lidy e Rauber, 2005).

3.1.2 Características Visuais

Nesta seção serão descritos alguns operadores bastante conhecidos na literatura de processamento de imagens para a representação de conteúdo de textura. Esses operadores foram escolhidos por estarem entre os mais utilizados e por já terem sido utilizados com sucesso em tarefas de classificação de áudio com o uso de espectrogramas (Costa et al., 2013; Costa, 2013; Costa et al., 2011; Lucio e Costa, 2015).

Adicionalmente, é importante ressaltar que as características descritas nesta seção, ditas "visuais", são assim chamadas pelo fato de que na forma como são empregadas, a natureza original do sinal é integralmente abstraída e, uma vez gerado o espectrograma, as imagens são exploradas pura e simplesmente tratando-se o problema em uma perspectiva de classificação de imagens, explorando o atributo visual mais evidente nas mesmas, que é a textura. Embora alguns descritores de características acústicas, descritos na Seção 3.1.1, também utilizem espectrogramas no processo em que são gerados, eles são explorados de uma perspectiva diferente e por isso foram classificados como "acústicos".

Filtros de Gabor

Durante muito tempo um sinal podia ser representado em função do tempo ou, alternativamente, em função da frequência por meio da transformada de Fourier. Entretanto essa abordagem possuía a limitação de permitir a extração de informações apenas no domínio da frequência e não em função do tempo. Em 1946, Denis Gabor apresentou os filtros de Gabor, que permitem extrair informações no domínio da frequência e do tempo. Em seu trabalho original Gabor buscava a síntese do sinal e preocupou-se em como um sinal poderia ser construído por meio da combinação linear de funções lineares (Li et al., 2010). Os filtros de Gabor correspondem a um conjunto de funções senoidais complexas, bidimensionais, moduladas por uma função Gausiana também bidimensional com propriedades muito úteis para a finalidade de classificação de imagens. Na análise de sinais em processamento de imagens, a extração de características exerce um papel importante no qual o principal objetivo é saber "o que está aonde". Com os princípios de Gabor, informações relacionadas a frequência pode informar "o que" , enquanto as ligadas ao tempo podem informar "aonde".

A segmentação da textura é uma tarefa difícil e muito importante em muitas aplicações de análise de imagens ou visão computacional e filtros de Gabor têm sido utilizados com êxito para estes propósitos. Existem muitas formas de se implementar filtros de Gabor apresentadas na literatura (Wu et al., 2011). Uma possível forma para filtros de Gabor bidimensionais no domínio espacial, portanto apropriados para imagens digitais, é dada pelas Equações 3.1 e 3.2.

$$\Psi(x,y) = exp\left(-\left(\frac{x^2+Y^2}{2\sigma^2}\right)\right) exp\left(\frac{j2\pi x}{\lambda}\right) \tag{3.1}$$

na qual j é a unidade imaginária, σ é o desvio padrão da função Gausiana e λ é o comprimento de onda.

Para uma imagem I de tamanho $M \times N$, e considerando $\Psi(x,y)$ conforme descrito na Equação 3.1, a saída do filtro de Gabor é obtida pela convolução da imagem de entrada com o filtro de Gabor apresentado na Equação 3.2.

$$\sum_x \sum_y I(m-x, n-y)\Psi(x,y) \tag{3.2}$$

Filtros de Gabor podem ser utilizados para detectar linhas. Uma vez que a imagem pode conter linhas com diferentes espessuras, é necessário construir filtros de Gabor com diferentes fatores de escala, variando λ. Adicionalmente, o filtro de Gabor pode detectar somente linhas verticais, o que não é suficiente em muitos casos, já que é

comum a ocorrência de linhas com diferentes orientações nas imagens. Assim, pode-se rotacionar $\Psi(x, y)$ com um ângulo θ para construir $\Psi(x', y')$ para a detecção de linhas com diferentes orientações. Nesse caso, x' e y' podem ser encontrados pelas Equações 3.3 e 3.4, respectivamente.

$$x' = x\cos\theta + y\sin\theta \tag{3.3}$$

$$y' = x\sin\theta + y\cos\theta \tag{3.4}$$

Local Binary Patterns

Foi apresentado pela primeira vez por Ojala et al. (1996) como uma medida complementar para contraste da imagem. Posteriormente foi adaptado e se tornou uma abordagem estrutural para descrição de textura (Ojala et al., 2002). A aplicação de LBP como descritor de textura tem como base o fato de que certos padrões binários locais à região de vizinhança de um pixel são propriedades fundamentais da textura de uma imagem e que o histograma de ocorrência destas características é provavelmente uma poderosa característica de textura.

O método define que a textura é descrita levando-se em consideração um pixel central C, com seus P vizinhos equidistantes considerando uma distância R, como pode ser visto na Figura 3.3. O histograma h de padrões LBP é sintetizado utilizando-se as diferenças de intensidade entre cada pixel central C e seus P vizinhos. De acordo com Ojala et al. (2002), boa parte da informação sobre características de textura é preservada na distribuição T descrita na Equação 3.5.

$$T \approx (g_0 - g_c, ..., g_{P-1} - g_c) \tag{3.5}$$

na qual g_c é a intensidade de nível de cinza do pixel central C e g_0 a g_{p-1} correspondem as intensidades de nível de cinza dos vizinhos. Quando um vizinho não corresponde exatamente á posição de um pixel, seu valor é obtido por interpolação.

Considerando o sinal resultante da diferença entre o pixel central C e cada um dos seus P vizinhos, como descrito na Equação 3.6, é definido que: se o sinal é positivo, o resultado é igual a um; caso contrário, o resultado é igual a zero, como descrito na Equação 3.7.

$$T \approx (s(g_0 - g_c), ..., s(g_{P-1} - g_c) \tag{3.6}$$

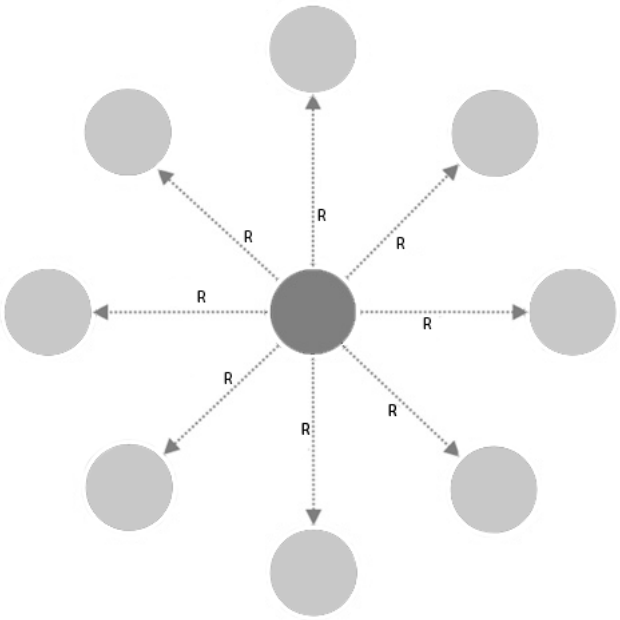


Figura 3.3: O operador LBP. O *pixel C*, escuro no centro, e os *pixels* claros são os P vizinhos adaptada de (Ojala et al., 1996)

$$s(g_i - g_c) = \begin{cases} 1 & \text{se } g_i - g_c \geq 0 \\ 0 & \text{se } g_i - g_c \leq 0 \end{cases} \tag{3.7}$$

na qual $i = [0, P]$ é o índice dos vizinhos de C.
Com isto, o valor do padrão LBP referente ao pixel C pode ser obtido por meio da multiplicação dos elementos binários por um coeficiente binomial. Associando-se um peso binomial 2^P a cada $s(g_p - g_c)$, as diferenças presentes na vizinhança são transformadas em um único código LBP, um valor $0 \leq C \leq 2^P$. A Equação 3.8 descreve como este código é obtido.

$$LBP_{P,R}(X_C, Y_C) = \sum_{P=0}^{P-1} s(g_P - g_C)2^P \tag{3.8}$$

assumindo que $X_C \in \{0, ..., No - 1\}$

O conceito de uniformidade da sequência obtida no padrão LBP, é baseado no número de transições entre zeros e uns presente na sequência associada ao padrão (Ojala et al., 2002). Um código LBP binário e considerado uniforme se o número de transições é menor ou igual a dois, considerando inclusive que o código é tratado como uma lista circular. Assim, o código representado pela sequência 00100100 não é considerado uniforme, já que

contém quatro transições. Por outro lado, o código 00100000 é considerado uniforme, já que apresenta apenas duas transições, como pode ser visto na Figura Figura 3.4.

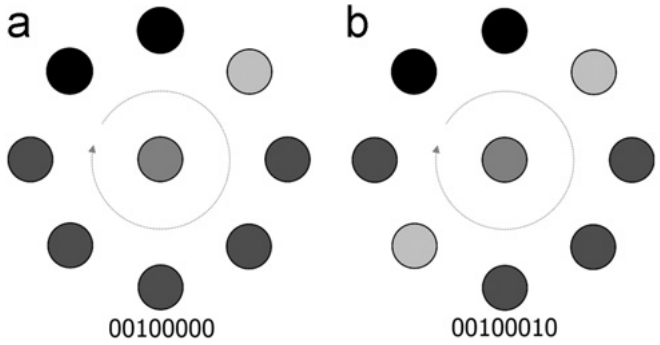


Figura 3.4: Uniformidade do padrão LBP adaptada de (Ojala et al., 1996)

Assim, em vez de utilizar integralmente o histograma de padrões LBP, cujo tamanho é 2^P, somente se faz uso dos valores associados a padrões uniformes, constituindo um vetor com menor dimensionalidade, com apenas 59 características quando é utilizado o $LBP_{8,2}$. De acordo com Ojala et al. (2002), além das 58 combinações uniformes, todos os padrões não uniformes encontrados devem participar de uma coluna adicional no histograma gerado. Devido a esse fato, o vetor de características LBP para na configuração de 8 vizinhos com distância 2 possui 59 características em sua constituição.

Robust Local Binary Pattern

O *Robust Local Binary Pattern* (RLBP) foi idealizado com a finalidade de suprir a deficiência apresentada pelo LBP quando de trata de extrair características de imagens que apresentam alta taxa de ruídos (Chen et al., 2013).

Assim como o LBP, o RLBP consiste em analisar um determinado número de pixels vizinhos P, baseado em um pixel central C, e levando-se em consideração uma distância R.

O que difere uma abordagem da outra é a forma como a uniformidade da sequência do padrão é obtida. Como foi citado anteriormente, um código LBP binário é considerado uniforme se e somente se o número de transicões é menor ou igual a dois. Tomemos então como exemplo a sequência 00100100. Ao analisarmos o número de transições entre os valores de 0 e 1, podemos ver claramente que esse valor é igual a 4, sendo assim, é plausível dizer que temos então uma sequência não uniforme.

O RBLP tem por finalidade analisar a representação binária de um pixel central para com seus vizinhos, buscando identificar *substrings* que seguem os padrões 101 e 010 e assim

substitui-ls por 111 e 000, respectivamente, e assim porteriormente realizar a análise de uniformidade.

Para entender melhor a ideia da substituição de *substrings* proposta, tomemos novamente como exemplo a representação binária 00100100, sendo que após aplicar a ideia apresentada, obtem-se a seguinte representação 00000000 e ao se realizar a análise de uniformidade, contasta-se que se trata de um padrão uniforme.

Local Phase Quantization

O borramento é uma forma de degradação de imagens que pode prejudicar consideravelmente a análise das mesmas. Esse ruído geralmente tem origem relacionada a problemas de aquisição e, em geral, o uso de algoritmos para removê-los é computacionalmente caro. Pensando nisso, Ojansivu e Heikkilä (2008) propuseram um novo método para análise de textura insensível ao borramento. É interessante observar que, embora o método tenha sido criado com esse propósito, ele também produz resultados muito bos para imagens não apresentam ruído.

O descritor, denominado *Local Phase Quantization* (LPQ) é baseado na propriedade de invariância ao borramento do espectro de fase de Fourier. Ele utiliza a informação de fase local extraída utilizando a 2D DFT computada sobre uma vizinhança retangular, chamada janela local, para cada pixel da imagem. A informação da fase local de uma imagem de tamanho $N \times N$ é dada pela *Short-time Fourier Transform* (STFT) descrita na Equação 3.9.

$$\hat{f}_{u_i}(x) = (f \times \phi_{u_i})x \tag{3.9}$$

sendo o filtro ϕ_{u_i} dado pela Equação 3.10

$$\phi_{u_i} = e^{-j2\pi u_i^T} | y \in \mathbb{Z}^2 || y || \infty \leq r \tag{3.10}$$

na qual $r = (m-1)/2$ é do tamanho da janela local e u_i é um vetor de frequências 2D.

No LPQ são considerados apenas quatro coeficientes complexos que correspondem às frequências 2D: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, $u_1 = [a, -a]^T$, em que $a = 1/m$. Por conveniência, a STFT é expressa por meio do vetor de notação conforme a Equação 3.11

$$\hat{f}_{u_i}(x) = w_{u_i}^T f(x) \tag{3.11}$$

sendo $F = [f(x_1), f(x_2), \cdots, f(x_{n^2})]$ denotado como uma matriz $m^2 \times N^2$ que compreende a vizinhança de todos os pixels da imagem e $w = [w_R, w_I]$, em que $w_R = Re[W_u1, W_u2, W_u3, W_u4]$ e $w_I = Im[W_u1, W_u2, W_u3, W_u4]$. $Re[]$ e $Im[]$ representam, respectivamente, as partes reais e imaginárias de um número complexo e a matriz de transformação $(8 \times N^2)$ é dada por $\hat{F} = wF$.

Ojansivu e Heikkilä (2008) assumem que a função $f(x)$ de uma imagem é resultado de um processo de primeira ordem de Markov, em que o coeficiente de correlação entre dois pixels x_i e x_j é relacionado exponencialmente com a sua distância L^2. Para o vetor f é definida uma matriz de covariância C de tamanho $m^2 \times m^2$, dada pela Equação 3.12. A matriz de covariância dos coeficientes de Fourier pode ser obtida por $DwCw^T$. Considerando que D não é uma matriz diagonal, os coeficientes são correlatos e podem deixar de ser por meio de $E = C^T\hat{F}$, sendo V uma matriz ortogonal derivada do valor de decomposição singular (SVD - *Singular Value Decomposition*) da matriz D, com $D' = V^TDV$.

$$C_{i,j} = \sigma^{||x_i - x_j||} \tag{3.12}$$

Os coeficientes são quantizados usando-se a Equação 3.13, em que e_{ij} são componentes de E. Esses elementos são transformados de binário para decimal por meio da Equação 3.14 e caracterizam valores inteiros compreendidos entre 0 e 255. Então, por meio de todas as posições da Imagem, é composto o vetor de 256 posições que correspondem ao histograma do LPQ.

$$q_{ij} = \begin{cases} 1 & \text{se } e_{ij} \geq 0 \\ 0 & \text{caso contrário} \end{cases} \tag{3.13}$$

$$b_j = \sum_{i=0}^{7} q_{ij} 2^i \tag{3.14}$$

Gray-Level Co-Occurence Matrix

A *Gray-Level Co-Occurence Matrix*(GLCM) é uma abordagem proposta por Haralick (1979), e, como o próprio nome sugere, faz o uso de matrizes de coocorrência para a caracterização da textura de uma imagem. Essa e realizada por meio de medidas estatísticas obtidas a partir da contagem de ocorrências dos níveis de cinza presentes nos pixels da imagem ou a obtidas por meio da forma cmo pixels de diferentes níveis de cinza se relacionam no espaço bidimensional de uma imagem.

A metodologia da extração das características consiste em se construir as matrizes de coocorrência para em seguida realizar a extração das medidas estatísticas inerentes as mesmas. As matrizes construídas a partir a imagem são da ordem $N \times N$, na qual N corresponde ao número de tons de cinza utilizados na representação da imagem, sendo que para cada posição da matriz é armazenada a probabilidade de que dois valores de intensidades de cinza estejam envolvidos por uma determinada relação espacial.

Em cada posição da matriz é armazenada a probabilidade de que dois valores de intensidades de cinza estejam envolvidos por uma determinada relação espacial. Parâmetros como a distância d entre os pixels e o ângulo θ que caracteriza a orientação de uma reta que passa pelos mesmos definem uma relação espacial. As possíveis orientações do ângulo θ são 0 °, 45 °, 90 °, 135 °, como pode ser observado na Figura 3.5.

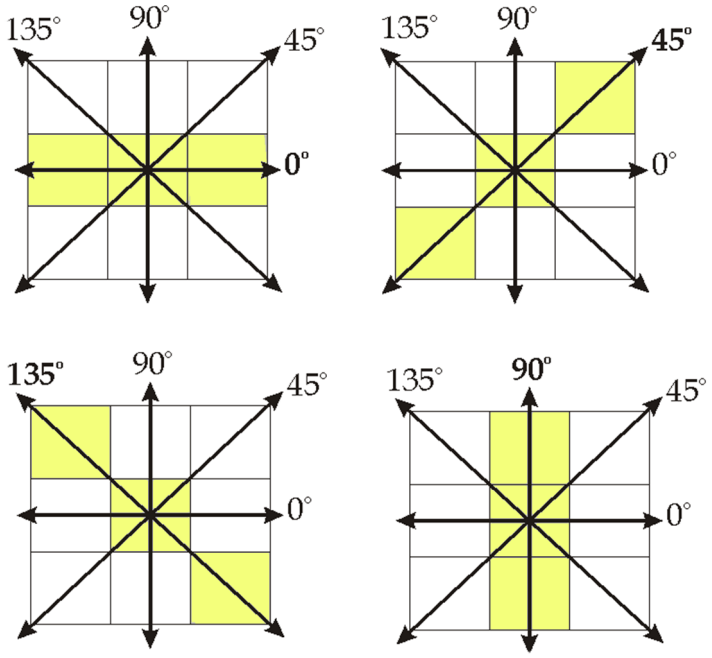


Figura 3.5: Orientações utilizadas para criação da matriz de co-ocorrência adaptada de (Haralick, 1979)

A Figura 3.6 apresenta a matriz de pixels de uma imagem com intensidades de cinza variando entre 0 e 3. Por meio da mesma iremos apresentar um exemplo da geração da matriz de co-ocorrência.

A partir da matriz de pixels apresentada na Figura 3.6, foi considerada a orientação $\theta = 0°$ e a distância $d = 1$ para criar a matriz de co-ocorrência, sendo que de acordo com o método proposto por Haralick, a matriz de co-ocorrência registra nas posições (i, j), o número de ocorrências de relação espacial entre um pixel de intensidade i e um pixel com intensidade j considerando a distância d e a orientação θ independente do

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Figura 3.6: Exemplo de matriz de pixels de uma imagem adaptada de (Haralick, 1979)

sentido da relação. Assim, a presença de um pixel de intensidade j imediatamente a direita de um pixel de intensidade i seria contabilizada na matriz com $d = 1$ e $\theta = 0°$ da mesma forma como a ocorrência da intensidade j imediatamente à esquerda de i seria contabilizada. Com isso, a matriz de co-ocorrência que se forma é simétrica. Depois de contadas as quantidades das relações espaciais, elas são transformadas em probabilidades para a relação dos processos de extração de características subsequentes, conforme mostra a Figura 3.7.

	0	**1**	**3**	**3**
0	4	2	1	0
1	2	4	0	0
2	1	0	6	1
3	0	0	1	2

→

	0	**1**	**2**	**3**
0	0,25	0,12	0,04	0
1	0,12	0,25	0	0
2	0,06	0	0,37	0,06
3	0	0	0,06	0,12

Figura 3.7: Matrix de co-ocorência de distância um e ângulo zero adaptada de (Haralick, 1979)

Haralick (1979) propuseram originalmente 14 medidas de características de texturas possíveis de se extrair das matrizes de co-ocorrência. Essas características são calculadas a partir de algumas equações que utilizam as probabilidades associadas as posições da matriz de co-ocorrência.

Das 14 características originalmente propostas, sete se consolidaram como características relevantes em processos de descrição de textura. Essas características são: contraste, energia (ou uniformidade), entropia, homogeneidade, momento de terceira ordem, probabilidade máxima e correlação. Sendo G o número de intensidade de cinza utilizado na representação da imagem e $p(i, j)$ a probabilidade de relacionamento entre as intensidades i e j, as Equações de 3.15 a 3.21 representam essas características.

$$\text{Contraste} = \sum_{i=1}^{G}\sum_{j=1}^{G}(i-j)^(2)(p(i,j)) \tag{3.15}$$

$$\text{Energia} = \sum_{i=1}^{G}\sum_{j=1}^{G}((p(i,j))^{(}2) \tag{3.16}$$

$$\text{Entropia} = \sum_{i=1}^{G}\sum_{j=1}^{G} p(i,j)\ \log\ p(i,j) \tag{3.17}$$

$$\text{Homogeneidade} = \sum_{i=1}^{G}\sum_{j=1}^{G}\left(\frac{p(i,j)}{1+(i-j)^{(}2)}\right) \tag{3.18}$$

$$\text{Momento de terceira ordem} = \sum_{i=1}^{G}\sum_{j=1}^{G} p(i,j)(i-j)^{(}3) \tag{3.19}$$

$$\text{Probabilidade máxima} = \sum_{i=1}^{G}\sum_{j=1}^{G} \max(p(i,j)) \tag{3.20}$$

$$\text{Correlação} = \left(\frac{p(i,j)-\mu_x\mu_y}{\sigma_x^2\sigma_y^2}\right) \tag{3.21}$$

nas quais $\mu_x = \sum_{i=1}^{G} i \times p_x(i)$, $p_x(i) = \sum_{j=1}^{G} p(i,j)$, $\sigma_x^2 = \sum_{i=1}^{G}(i-\mu_x)^2 p_x(i)$, $\mu_y = \sum_{j=1}^{G} j \times p_y(j)$, $p_y(j) = \sum_{i=1}^{G} p(i,j)$ e $\sigma_y^2 = \sum_{j=1}^{G}(j-\mu_y)^2 p_y(j)$.

3.2 Divisão da Imagem em Zonas

De acordo com (Costa, 2013), o zoneamento da imagem tem como objetivo preservar as informações locais presentes em regiões específicas da imagem. Decorrente do fato de que a textura gerada à partir de espectrogramas extraídos de registros de áudio não apresentarem um conteúdo uniforme ao longo dos seus eixos horizontal e vertical. Além da preservação de informações locais, a estratégia de divisão das zonas é bastante oportuna por permitir naturalmente a criação de um *pool* de classificadores, pois um classificador é criado para cada uma das zonas.

Ao se pesquisar na literatura foi constatada a existência de alguns sistemas de zoneamento que diferem entre si, podendo estes serem lineares ou não (Costa, 2013). As próximas subseções, descrevem algumas metodologias de zoneamento existentes.

Divisão em Zonas Lineares

Com a divisão linear, são estabelecidas na imagem do espectrograma zonas de igual tamanho que correspondem a bandas de frequência. Os limites de cada banda criada

dependem da quantidade de zonas definidas e do limite de frequência até o qual o sinal do áudio utilizadas apresenta informação relevante. A Figura 3.8 apresenta um espectrograma dividido linearmente em 3 zonas

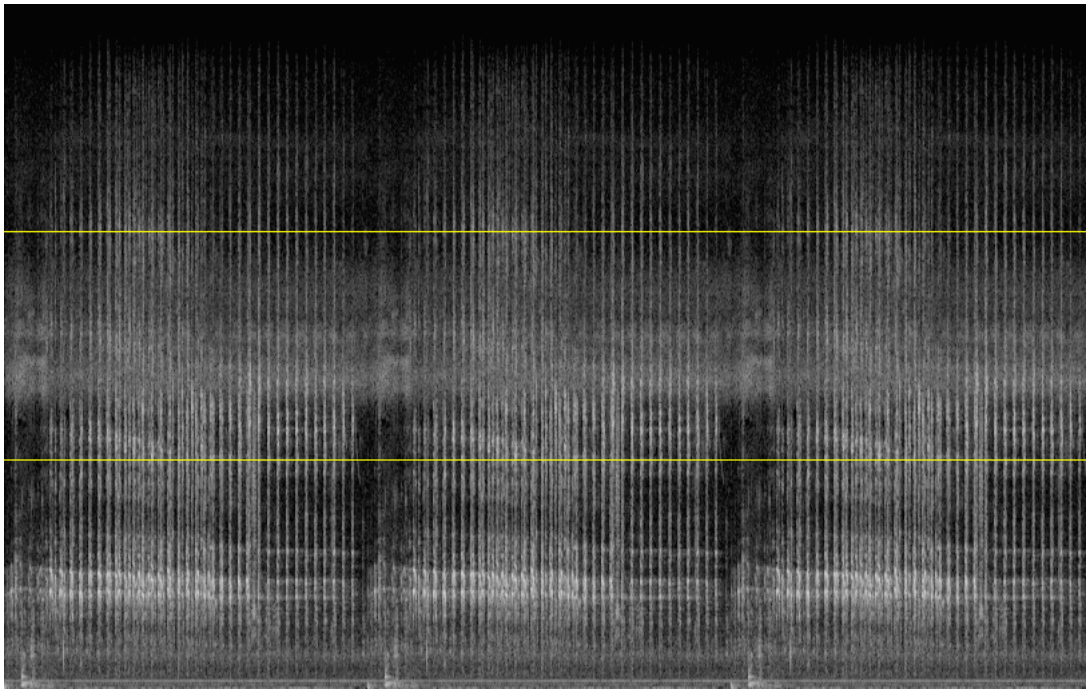


Figura 3.8: Exemplo de zoneamento utilizando escala linear, sobre um espectrograma gerado a partir de uma amostra de áudio da espécie Thamnophilus Ruficapillus

Divisão pela escala de Mel

É uma escala psicoacústica apresentada por S. S. Stevens (1940) e que esta diretamente relacionada as frequências percebidas pelos humanos, assim como a escala Bark. No entanto, são estabelecidas 15 bandas de frequência, cujos limites em Hz são: 0, 40, 161, 200, 404, 693, 867, 1000, 2022, 3393, 4109, 5526, 6500, 7743 e 12000. Nas aplicações que envolvem a classificação a partir de imagens de espectrograma, o número de zonas criadas e, consequentemente, o número de classificadores, depende do limite de frequência até o qual a imagem do espectrograma apresenta informações relevantes. A Figura 3.9 apresenta um espectrograma dividido entre os limites de frequência supracitados de baixo para cima, sendo que a primeira linha na parte inferior da imagem representa o limite de frequência mais baixo e a última linha na parte superior da imagem representa o maior limite de frequência.

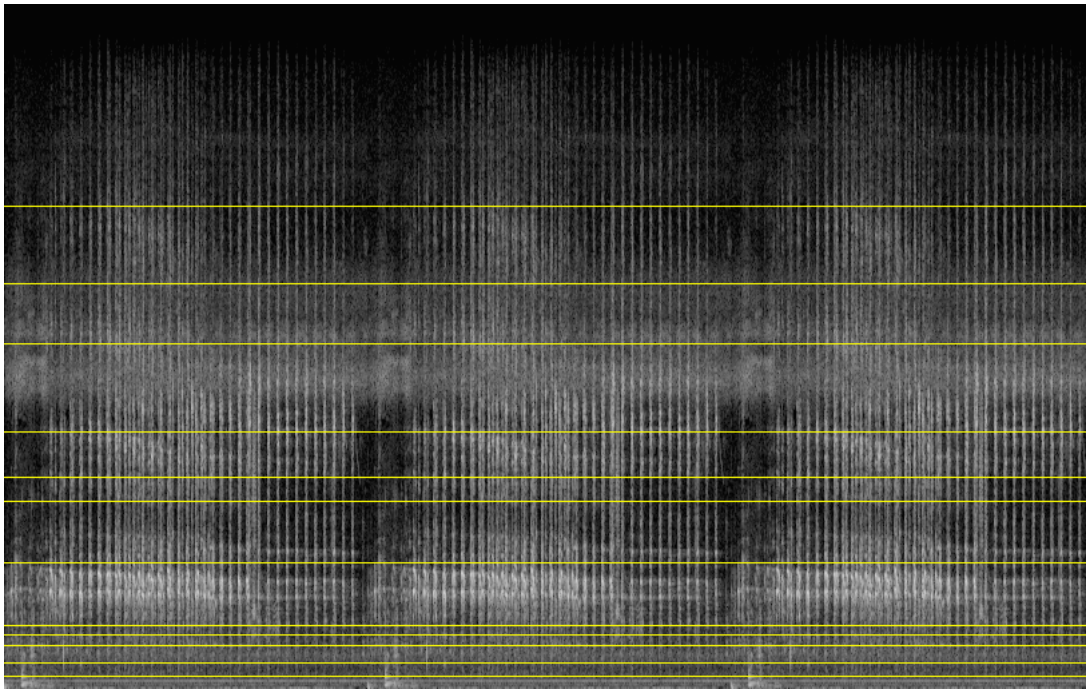


Figura 3.9: Exemplo de zoneamento utilizando escala de Mel sobre um espectrograma gerado a partir de uma amostra de áudio da espécie Thamnophilus Ruficapillus

3.3 Combinação de Classificadores

Muitos esquemas de classificação são criados tendo por base a utilização de um único classificador para resolver um determinado problema. No entanto, a qualidade das hipóteses induzidas por esses classificadores depende da quantidade dos exemplos dos conjuntos de treinamento. Por outro lado, muitos dos sistema de aprendizado de máquina conhecidos não estão preparados para trabalhar com uma grande quantidade de exemplos. Uma maneira para resolver esse problema consiste em realizar a combinação de classificadores, podendo proporcionar um melhor resultado se comparado aos resultados individuais de cada classificador (Bernardini, 2006).

O que fundamenta o ganho de desempenho ao se realizar a combinação de classificadores é que os conjuntos de padrões classificados incorretamente por diferentes classificadores não necessariamente se sobrepõem. Isso sugere que diferentes projetos de classificadores podem oferecer informação complementar sobre os padrões a serem classificados, fato este que pode acarretar na melhora do desempenho do sistema de classificação (Kittler et al., 1998).

De acordo com Kittler et al. (1998), o método mais comumente utilizado para a combinação de classificadores é dado por meio das seguintes funções matemáticas: máximo, mínimo, produto, soma e média. As Equações 3.22 a 3.26 descrevem as fórmulas utilizadas pelas funções supracitadas.

$$Regra\,do\,Maximo(v) = \arg\max_{k=1}^{c} \max_{i=1}^{n} P(\omega_k|l_i(v)) \tag{3.22}$$

$$Regra\,do\,Minimo(v) = \arg\max_{k=1}^{c} \min_{i=1}^{n} P(\omega_k|l_i(v)) \tag{3.23}$$

$$Regra\,do\,produto(v) = \arg\max_{k=1}^{c} \prod_{i=1}^{n} P(\omega_k|l_i(v)) \tag{3.24}$$

$$Regra\,da\,Soma(v) = \arg\max_{k=1}^{c} \sum_{i=1}^{n} P(\omega_k|l_i(v)) \tag{3.25}$$

$$Regra\,da\,Media(v) = \frac{1}{n} \arg\max_{k=1}^{c} \sum_{i=1}^{n} P(\omega_k|l_i(v)) \tag{3.26}$$

nas quais, v representa o padrão que será classificado, n é o número de classificadores, l_i representa a saída do i-ésimo classificador em um problema com os possíveis rótulos de classe $\Omega = \omega_1, \omega_1, ..., \omega_c$ e $P(\omega_k|l_i(v)$ é a probabilidade de que a amostra v pertença a classe ω_k encontrada pelo i-ésimo classificador.

3.4 Avaliação dos resultados

Esta seção apresenta os critérios comumente utilizados para avaliar a eficiência de sistemas de classificação, sendo estes: *precison*, *recall*, *F-measure* e *Macro-F* (Casanova, 2011). As subseções seguintes apresentas os critérios de avaliação citados.

3.4.1 Precision

É o total de exemplos corretamente classificados como classe C sobre o total de exemplos classificados como classe C. Sua formula é expressa pela Equação 3.27.

$$Precision(C_i) = \frac{M(C_i, C_i)}{M(*, C_i)} \tag{3.27}$$

onde C_i é a classe que se está sendo analizada e $M(C_i, C_i)$ é o total de exemplos classificados para a classe C_i e $M(*, C_i)$ é o total de exemplos reconhecidos como C_i.

3.4.2 Recall

É o total de exemplos corretamente classificados como classe C sobre o total de exemplos pertencentes a classe C presentes ao conjunto de dados, a desvantagem desta métrica é que ela não leva em consideração todas as medidas. Sua formula é expressa pela Equação 3.28.

$$Recall(C_i) = \frac{M(C_i, C_i)}{M(C_i, *)} = \frac{exemlos\, corretamente\, reconhecidos}{total\, de\, exemplos} \tag{3.28}$$

onde C_i é a classe que se está sendo analizada e $M(C_i, C_i)$ é o total de exemplos classificados para a classe C_i e $M(C_i, *)$ é o total de exemplos reconhecidos corretamente reconhecidos como C_i.

3.4.3 F-measure

É a média harmonica das medidas de Precison e Recall, sendo uma forma de de expressar as duas medidas com um único valor, sua formula é expressa pela Equação 3.29.

$$F - measure(C) = \frac{2 \times recall(C) \times precision(C)}{recall(C) + precision(C)} \tag{3.29}$$

onde C é a classe através do qual foram calculados o *Precision* e o *Recall*

3.4.4 Macro-F

É a média aritmética das F-measures de todas as classes presentes no conjunto de dados, sua formula é expressa pela Equação 3.30.

$$Macro - F(h) = \frac{1}{k} \sum_{i=1}^{k} F - measure(C_i) \tag{3.30}$$

onde k é o total de classes sobre o qual os testes foram realizados e C_i é a classe sobre e qual o F-Measure foi calculado.

3.5 Considerações Finais

O presente capítulo apresentou um conjunto de técnicas utilizadas em sistemas de classificação, tendo uma seção definida para cada um dos processos envolvidos em um sistema de classificação. Os próximos parágrafos descrevem sucintamente cada um dos assuntos abordados das seções presentes no capítulo.

A Seção 3.1 apresenta uma revisão de algumas das principais técnicas para a extração de características visuais e acústicas presentes na literatura.

A Seção 3.2 descreve as principais técnicas de zoneamento de imagen utilizas na literatura.

A Seção 3.3 apresenta as regras de fusão para combinação de classificadores em paralelo mais conhecidas na literatura.

E por fim a Seção 3.4 apresenta as medidas de avaliação que foram aplicadas sobre os resultados apresentados pelos classificadores.

Todos os tópicos abordados neste capítulo foram utilizados na elaboração do método proposto apresentado no Capítulo 4.

4

Método Proposto

Este capítulo tem por objetivo apresentar o método de classificação utilizado no desenvolvimento do presente trabalho. No que diz respeito especificamente ao método proposto neste trabalho, pode-se identificar as seguintes etapas para realizar a tarefa de classificação: criação da base de dados, divisão da base de dados em partições, geração dos espectrogramas, divisão das imagens em zonas, extração das características, construção de classificadores para cada zona criada e fusão das saídas dos classificadores. Cada uma dessas etapas estão apresentadas na Figura 4.1 e são descritas nas seções seguintes.

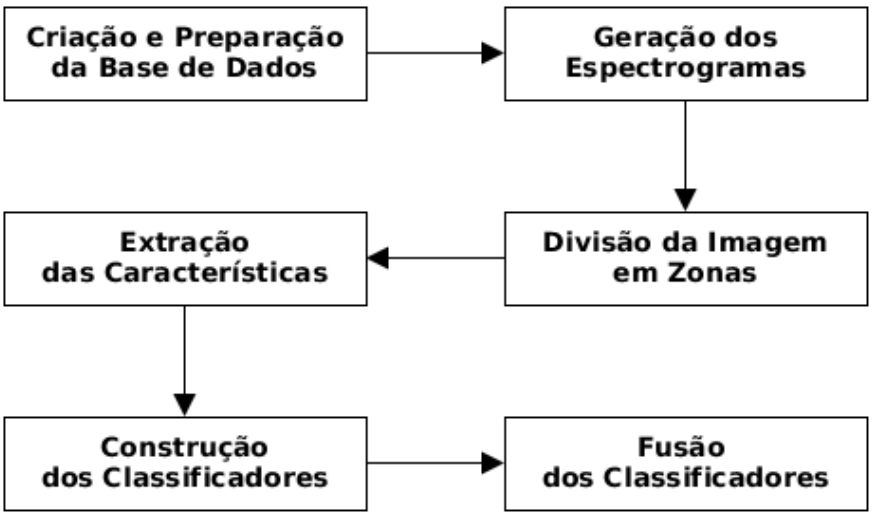


Figura 4.1: Sequência de etapas do método proposto

4.1 Criação da base de dados

Os sons emitidos por pássaros são tipicamente divididos em cantos e chamados. Geralmente, os cantos são mais longos e mais complexos, possuindo funções relacionadas ao acasalamento e a defesa territorial. Muitas espécies de pássaros cantam somente durante a temporada de acasalamento e é geralmente mais limitado aos machos. Os chamados por sua vez, são tipicamente curtos e possuem funções menos importantes, tais como: alarme, voo ou alimentação (Fagerlund, 2007).

Além da divisão entre cantos e chamados, os sons emitidos pelos pássaros também são divididos em nível hierárquico de frases, sílabas e elementos. Uma frase é composta por uma série de sílabas que ocorrem em um determinado padrão. Normalmente, sílabas em uma mesma frase são similares umas as outras, todavia, algumas vezes elas também podem ser diferentes. Sílabas são construídas por elementos, mas em alguns casos mais simples, sílabas e elementos podem ser a mesma coisa. Entretanto, sílabas complexas podem ser construídas por vários elementos. Os chamados são geralmente compostos por uma sílaba ou série de sílabas similares e o nível de frase não pode ser detectado. É comum em algumas espécies de pássaros se perder o nível da frase quando se analisa seus cantos (Catchpole, 2008).

Para a execução dos experimentos apresentados no presente trabalho foi necessária a criação de uma base de dados. Essa base foi criada aplicando-se a metodologia baseada na utilizada no trabalho de Lopes et al. (2011a). A base é composta por um subconjunto de espécies de pássaros extraídos a partir da base de dados de cantos e chamados disponibilizada pelo site Xeno-Canto[1]. As gravações disponibilizadas pelo site foram realizadas diretamente de ambientes reais, sem qualquer filtro ou pré-processamento, de modo que as gravações contém sons de outras espécies de pássaros e animais, assim como os ruídos do ambiente.

Seguindo a metodologia utilizada para a síntese da base de dados de sinais de áudio utilizada por Lopes et al. (2011a), foram realizados os seguintes passos:

- Por meio do uso da ferramenta de busca do site supracitado, foi delimitada a localização geográfica desejada, que abrangeu um raio de 250 KM nas proximidades da cidade de Curitiba, como pode ser visto na Figura 4.2;

- Dos registros apresentados como resultado, foram selecionados os de 75 espécies que apresentaram altas frequências na tabela que apresenta o resultado da busca. Como

[1] http://www.xeno-canto.org/, acessado em 02/02/2015

Figura 4.2: Localização geográfica dos registros de áudio dos pássaros (Lopes et al., 2011a)

em alguns casos o número de instâncias para a espécie era pequeno, foram utilizadas gravações da mesma espécie colhidas em outras regiões;

- Em seguida foram eliminadas as espécies que não possuíam registros de cantos, visto que estes foram escolhidos para a criação do sistema de classificação devido a sua maior complexidade se comparado aos chamados. Após esse passo, restaram 73 espécies;

- Foi efetuado o download dos registros do site por meio de um mecanismo automático de extração de informação, que dividiu as gravações em uma pasta para cada uma das espécies;

- Em seguida as gravações foram divididas em pulsos[2], devido ao fato desses segmentos caracterizarem melhor a vocalização do pássaro, visto que de acordo com a literatura a utilização desses pulsos de áudio melhoram os resultados do processo de identificação (Lopes et al., 2011a). O processo de divisão do sinal é exemplificado na Figura 4.3, na qual é apresentado o áudio original de uma espécie e os pulsos correspondentes a ele.

Após a execução dos passos apresentados no trabalho de Lopes et al. (2011a), foram efetuados os passos abaixo para adequar a base ao sistema de classificação aqui proposto.

- Foram desprezados todos os pulsos obtidos com tempo inferior a 1 segundo e, posteriormente, foram eliminadas as espécies de pássaros que continham um número de instâncias inferior a 10;
- Foi realizada a concatenação de cada um dos arquivos de áudio com ele mesmo até que a duração do mesmo fosse igual ou superior a 30 segundos. Este procedimento foi realizado para que se obtivessem espectrogramas de um mesmo tamanho;
- Foi realizada a equalização das frequências de sinal de todas as amostras de áudio para 44100 Hz com uma taxa de amostragem de 16 *bits*.

Ao realizar os passos apresentados anteriormente foi obtida a base de dados utilizada composta por um total de 2814 amostras de áudio distribuídas entre 46 espécies. A relação das espécies e da quantidade de amostras para cada espécie pode ser vista na Tabela 4.1.

Tabela 4.1: Relação de espécies apresentada na base de dados utilizada nos experimentos

Espécie	**Número de amostras**
Aegolius Harrisii	64
Amazilia Versicolor	28
Anthus Lutescens	45
Attila Rufus	10
Automolus Leucophthalmus	120
Basileuterus Leucoblepharus	50
Batara Cinerea	87
Brotogeris Tirica	28

continua na próxima página

[2]Pequeno intervalo de áudio com alta amplitude de frequência

continuação da Tabel Tabela 4.1

Espécie	**Número de amostras**
Camptostoma Obsoletum	77
Campylorhamphus Falcularius	76
Certhiaxis Cinnamomeus	112
Chiroxiphia Caudata	90
Clibanornis Dendrocolaptoides	82
Cnemotriccus Fuscatus	36
Colaptes Campestris	56
Colonia Colonus	18
Cranioleuca Obsoleta	46
Crypturellus Noctivagus	14
Culicivora Caudacuta	25
Cyanocorax Caeruleus	31
Drymophila Malura	93
Dysithamnus Mentalis	78
Emberizoides Ypiranganus	30
Gnorimopsar Chopi	56
Hemitriccus Orbitatus	14
Hypoedaleus Guttatus	71
Lathrotriccus Euleri	110
Leucochloris Albicollis	83
Mackenziaena Leachii	32
Malacoptila Striata	19
Mimus Saturninus	148
Myiodynastes Maculatus	49
Schiffornis Virescens	93
Sittasomus Griseicapillus	77
Stymphalornis Acutirostris	14
Synallaxis Spixi	85
Tangara Desmaresti	27
Thamnophilus Ruficapillus	82
Theristicus Caudatus	51

continua na próxima página

continuação da Tabel Tabela 4.1

Espécie	**Número de amostras**
Thraupis Palmarum	100
Thryothorus Longirostris	49
Trichothraupis Melanops	44
Trogon Surrucura	66
Vanellus Hilensis	70
Xenops Minutus	70
Xiphorhynchus Fuscus	108

O que justifica o procedimento de equalização da frequência é o fato de não haver nenhuma padronização nos registros de áudio utilizados na base de dados. O limite de frequência escolhido tem por base as afirmações de Gioppo e Kaestner (2011), de que os sons emitidos pelos pássaros fica/está entre 1 kHz e 5 kHz, com alguns harmônicos podendo alcançar até 18 kHz.

4.2 Divisão da Base de Dados em Folds

Após ter construído a base de dados, a mesma foi dividida em 10 partições. Tal divisão se fundamenta no fato de que durante a processo de criação da base foram desconsideradas as espécies que possuíam menos de 10 amostras de áudio. Desta forma, garantiu-se a presença de pelo menos uma amostra de cada espécie em cada fold. Para cada um dos folds as amostras começaram a ser distribuídas a partir do fold 1 até o 10, e depois do fold 10 até o 1, repetidas vezes até que todas as amostras da espécie em questão fossem distribuídas entre os folds. Esse processo foi realizado para cada uma das espécies presente na base de dados.

4.3 Geração do Espectrograma

Para a geração dos espectrogramas foi utilizado o sofware SoX 14.4.1 (*Sound eXchange*)[3], que permite a realização de conversões entre vários formatos diferentes de representação de áudio. Esse utilitário permite a utilização de alguns parâmetros que afetam a aparência do espectrograma gerado, sendo possível delimitar a altura e a largura, e o limite inferior da amplitude do sinal de áudio a ser considerada.

[3]http://sox.sourceforge.net, acessado em 05/01/2015

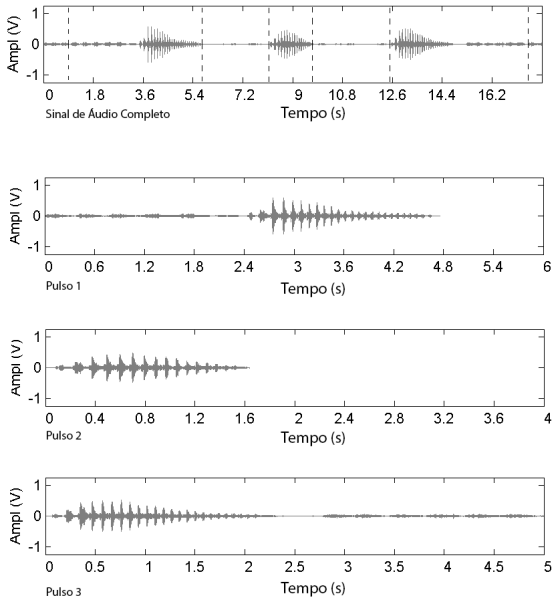


Figura 4.3: Representação da Divisão do sinal de áudio em pulsos apresentada por Lopes et al. (2011b)

O espectrograma gerado em seu eixo horizontal representa o tempo, enquanto o eixo vertical representa a frequência e a intensidade dos pixels representam a amplitude do sinal.

No protocolo de geração dos espectrogramas, foi utilizada a Transformada discreta de Fourier com o uso de uma Janela de Hanning com um tamanho de 1024, oque torna possível preservar uma boa relação entre as duas dimensões da imagem.

Como parte do trabalho é voltado para a extração de características visuais a partir da textura das imagens, se optou trabalhar com imagens em tons de cinza. visto que a informação da intensidade de energia do sinal é preservada quando se utiliza imagens nesta configuração. E também pelo fato de a maioria das técnicas de processamento de imagens empregarem a mesma metodologia. A Figura 4.4 apresenta um espectrograma gerado a partir de um sinal de áudio de 30 segundos, em tons de cinza.

A conversão do espectrograma gerado para tons de cinza foi realizada tendo por base a técnica apresentada por Gonzalez e Woods (2008). A equação 4.1 descreve o processo de conversão.

$$L = (0,2989 \times R) + (0,5870 \times G) + (0,1140 \times B) \tag{4.1}$$

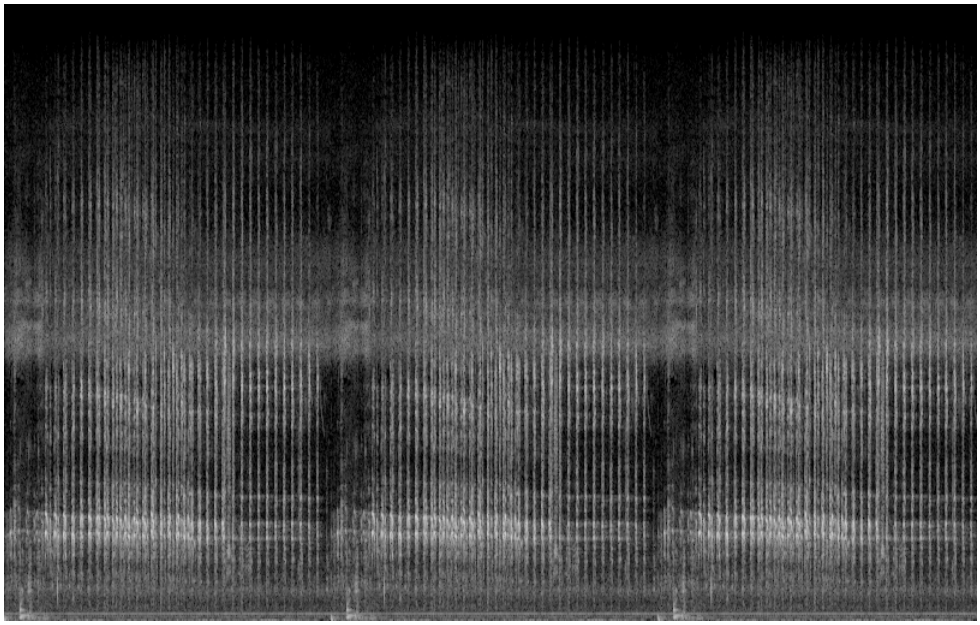


Figura 4.4: Espectrograma em tons de cinza gerado a partir de uma amostra de áudio de 30 segundos

na qual, L corresponde à luminância e consequentemente ao tom de cinza resultante, R é a intensidade do canal vermelho, G é a intensidade do canal verde e B é a intensidade do canal azul.

4.4 Divisão da Imagem em Zonas

Para o zoneamento da imagem foram adotadas as técnicas apresentadas na Seção 3.2, na qual foram realizados testes utilizando zoneamento linear, em que foi realizada a variação do número de zonas com o propósito de se verificar que quantidade de zonas apresenta uma taxa de acerto mais elevada, e não linear, fazendo-se o uso da escala de Mel. Os testes realizados com a escala de Mel foram realizados removendo as zonas inferiores da mesma, decorrente do fato de não haver uma quantidade de pixels suficiente para se extrair as características utilizadas no sistema de classificação. Sendo assim, das 15 bandas de frequência presentes na escala Mel, apenas 13 foram empregadas.

4.5 Extração de Características

4.5.1 Características Acústicas

Esta seção tem por finalidade apresentar uma descrição sobre os conjuntos de características acústicas utilizadas nesta dissertação. A mesma encontra-se dividida em subseções nas quais são apresentadas as configurações utilizadas com os descritores para gerar os vetores de características.

SSD

Os vetores de características gerados a partir do SSD possuiam 161 elementos. Os vetores foram obtidos por meio da biblioteca *Rhythm and Timbre Feature Extraction from Music* [4].

RH

Os vetores de características gerados a partir do RH possuiam 60 elementos e, assim como o SSD foram obtidos por meio da biblioteca *Rhythm and Timbre Feature Extraction from Music*.

RP

Os vetores de características gerados a partir do RP possuiam 1380 elementos e, assim como os outros dois descritores citados anteriormente também foram gerados por meio da biblioteca *Rhythm and Timbre Feature Extraction from Music* .

4.5.2 Características Visuais

Esta seção tem por finalidade apresentar uma descrição sobre os conjuntos de características visuais utilizados nesta dissertação. A mesma encontra-se dividida em subseções nas quais são apresentadas as configurações utilizadas com os descritores para gerar os vetores de características.

LBP

A geração dos vetores de características extraídos por meio do uso do LBP levou em consideração o $LBP_{8,2}$, ou seja, considerou-se 8 pixels com uma distância 2 pixels do pixel

[4] http://ifs.tuwien.ac.at/mir/musicbricks/index.html, acessado em 05/05/2016

central, fato este que proporcionou a criação de vetores de características compostos por 59 elementos.

LPQ

Para a criação dos vetorores de características a partir do LPQ se fez o uso de uma janela de dimensão 5×5 e ao se percorrer todos os pixels da imagem com com a mascará foi obtido um vetor composto por 256 elementos.

RLBP

A criação dos vetores de características com o RLBP assim como o LBP levou em consideração o $RLBP_{8,2}$, ou seja, considerou-se 8 pixels com uma distância 2 pixels do pixel central, fato este que proporcionou a criação de vetores de características compostos por 59 elementos.

Filtro de Gabor

Os vetores de características obtidos pelo filtro de gabor tem sua dimensão dependente do número de fatores de escala, da quantidade de rotações utilizadas e do número de medidas estatísticas escolhidas, sendo que o total de elementos presentes em um vetor é dado pela seguinte relação *Número de rotações* $\times$ *Número de fatores de escala* $\times$ *Quantidade características de Gabor*.

Neste trabalho foi adotado que o a quantidade de fatores de escala seria fixada em 5 e, que duas características de Gabor seriam adotadas, sendo estas: a amplitude da média e a energia quadrática da média. Coube então a variação do número de rotações influenciar no total de elementos presentes no vetor de características, decorrente do fato de se utilizar o intervalo de rotações de 5 à 20, a quantidade de características utilizadas foram respectivamente: 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, 228 e 240.

GLCM

O vetores de características obtidos com uso do GLCM eram compostos por 28 elementos, calculados a partir das 4 orientações possíveis sendo estas 0°, 45°, 90° e 135°.

4.6 Sistema de Classificação

Para realizar a tarefa de classificação foi utilizado o SVM, que é um modelo de aprendizagem supervisionado, que associa algoritmos de aprendizagem para analisar dados usados para classificação (Vapnik, 2000). Seu uso é muito difundido em toda comunidade científica em trabalhos que envolvem a análise de dados e reconhecimento de padrões, sendo que esse modelo tem apresentado bom desempenho em trabalhos publicados recentemente (Briggs et al., 2009; Fagerlund, 2007). Todavia, não foi descartado o potencial de uso de outros algoritmos de classificação, de modo a verificar variações no processo de classificação quando aplicada alguma estratégia de combinação de classificadores.

Após extraídas as características foi realizada a classificação por meio do SVM implementado na biblioteca LIBSVM (Chang e Lin, 2011), sendo que que o kernel utilizado foi o *Radial Basis Function* (RBF) e os parâmetros C (custo) e γ foram otimizados por meio do uso do procedimento *grid-search*. A técnica consiste na utilização de dois conjuntos de dados, sendo um para treino e outro para teste. Com o objetivo de obter um resultado mais consistente foi utilizada a técnica de validação cruzada, na qual um dos folds criados é utilizado como conjunto de teste e os demais para treinamento. O processo é repetido até que todos os folds criados tenham sido utilizados como conjunto de teste (Kittler et al., 1998).

Ao final, toma-se como medida de desempenho a *precision*, o *recall* e a *F-measure* obtidos entre cada uma das etapas da validação cruzada.

4.7 Considerações Finais

Este capítulo teve por finalidade apresentar todas as etapas envolvidas no método de classificação automática de espécies de pássaros aqui proposto.

A seção 4.1 apresenta as etapas envolvidas na criação da base de dados utilizada em todos os testes deste trabalho, enquanto a seção 4.2 a forma como a base de dados foi dividida em folds.

A seção 4.3 apresenta a forma como os espectrogramas foram gerados. A seção 4.4 por sua vez apresenta como as técnicas de zoneamento aplicadas sobre os espectrogramas.

A seção 4.5 apresenta os descritores de acústicos e visuais utilizados, assim como a configuração adotada para os mesmo.

A seção 4.6 apresenta qual metodologia de classificação foi utilizada sobre os vetores de características obtidos.

E por fim a seção 3.4 apresenta as medidas de avaliação que foram aplicadas sobre os resultados apresentados pelos classificadores.

O capítulo 5 apresenta os resultados obtidos com a execução de todas as etapas apresentadas neste capítulo.

5

Resultados Experimentais

O presente capítulo tem como finalidade apresentar os resultados obtidos utilizando a metodologia proposta apresentada no Capítulo 4. Este encontra-se dividido em 4 seções: 5.1, 5.2, 5.3 e 5.4. A primeira apresenta os resultados obtidos após utilizar os descritores de características visuais apresentados na Seção 3.1.2. A segunda apresenta os resultados obtidos utilizando as técnicas de zoneamento apresentadas na Seção 3.2 sobre os descritores de características visuais apresentados na Seção 3.1.2. A terceira apresenta os resultados obtidos com o uso dos descritores de características acústicas apresentados na Seção 3.1.1. A quarta apresenta os resultados obtidos com a combinação de características acústicas e visuais, por meio das regras de fusão apresentadas na Seção 3.3.

5.1 Etapa 1: Testes Utilizando Características Visuais

Esta seção apresenta os resultados obtidos com o uso dos descritores de características visuais: LBP, LPQ, RLBP, Filtros de Gabor e GLCM. Todos os testes foram realizados utilizando a base de dados apresentada na Seção 4.1. Também foi realizada a variação dos parâmetros(intervalo de tempo, limite da frequência e amplitude do sinal) de geração dos espectrogramas, tendo por objetivo verificar qual seria a melhor combinação desses parâmetros para o descritor de textura empregado.

Os resultados de todos os testes são dados por meio da utilização de uma matriz de confusão, a partir da qual são calculadas as medidas de *Precision*, *Recall*, para em seguida se calcular o *F-measure* que é a média harmônica das medidas de *Precison e Recall*.

A Tabela 5.1 apresenta as taxas de acerto obtidas. Pode-se observar que ao se elevar o limite de frequência utilizado, há um aumento nas taxas de acerto. Um aumento sensível

nas taxas de acerto também é observado ao se utilizar seguimentos de áudio de maior duração, todavia, quando houve um aumento no limite inferior da amplitude do sinal as taxas de acerto tenderam a diminuir para os descritores LBP, RLBP e LPQ. O mesmo não ocorreu com os testes realizados utilizando Filtros de Gabor e GLCM, nesses dois casos houve um pequeno aumento nas taxas de acerto.

Tabela 5.1: Resultados dos testes com descritores de características visuais

Tempo	**Frequência**	**Limite Inferior da Amplitude do Sinal**	**Recall**	**Precison**	**F-Measure**
		LBP			
00:15	16000	80	70,87%	75,87%	72,67%
00:15	18000	80	69,33%	75,69%	71,89%
00:15	20000	80	72,36%	77,70%	74,34%
00:15	22000	80	73,18%	77,64%	74,89%
00:30	**22000**	**80**	**74,70%**	**79,01%**	**76,39**%
00:30	18000	90	70,15%	75,97%	72,21%
00:30	22000	90	71,77%	76,38%	73,48
		LPQ			
00:15	16000	80	29,61%	31,62%	25,80%
00:15	18000	80	29,96%	32,38%	27,25%
00:15	20000	80	30,96%	33,81%	26,93%
00:15	22000	80	64,65%	70,23%	66,59%
00:30	**22000**	**80**	**66,65%**	**73,06%**	**68,52%**
00:30	18000	90	61,36%	66,89%	63,14%
00:30	22000	90	64,65%	70,23%	66,59
		RLBP			
00:15	16000	80	70,14%	76,06%	72,03%
00:15	18000	80	70,68%	76,21%	72,55%
00:15	20000	80	73,51%	78,24%	75,25%
00:15	22000	80	73,22%	78,48%	75,14%
00:30	**22000**	**80**	**74,94%**	**80,26%**	**76,80%**
00:30	18000	90	70,19%	75,30%	71,91%

continua na próxima página

continuação da Tabela Tabela 5.1

Tempo	Frequência	Limite Inferior da Amplitude do Sinal	Recall	Precison	F-Measure
00:30	22000	90	70,65%	75,67%	72,29%
Filtros de Gabor					
00:15	16000	80	69,70%	76,61%	71,89%
00:15	18000	80	70,62%	77,02%	72,64%
00:15	20000	80	70,78%	78,77%	72,98%
00:15	22000	80	72,31%	79,00%	74,45%
00:30	22000	80	72,49%	78,53%	74,49%
00:30	18000	90	73,59%	78,29%	75,28%
00:30	**22000**	**90**	**73,99%**	**78,73%**	**75,67%**
GLCM					
00:15	16000	80	41,52%	50,31%	41,69%
00:15	18000	80	40,91%	47,37%	41,30%
00:15	20000	80	40,55%	45,88%	40,61%
00:15	22000	80	40,84%	49,70%	41,39%
00:30	22000	80	40,71%	46,79%	41,07%
00:30	18000	90	39,17%	46,73%	39,34%
00:30	**22000**	**90**	**45,59%**	**56,34%**	**44,37%**

Após ter realizados os testes apresentados na Tabela 5.1 e de posse dos melhores parâmetros para a geração de espectrogramas quando se utilizou os Filtros de Gabor, uma nova bateria de testes foi realizada. Essa teve por objetivo verificar se a variação de rotações para a criação do vetor de características extraídas com os filtros de Gabor influenciaria as taxas de acerto.

A Tabela 5.2 apresenta os resultados dos testes realizados anteriormente citados, e por meio desta podemos constatar que ao se aumentar o número de rotações houve um aumento nas taxas de acerto. Todavia ao se ultrapassar o limite de 16 rotações foi constado que as os resultados apresentados tendiam a diminuir.

Tabela 5.2: Resultados dos testes complementares realizados com Filtros de Gabor

Rotações	**Recall**	**Precison**	**F-Measure**
5	69,62%	74,92%	71,25%
6	74,31%	78,64%	75,86%
7	74,70%	78,58%	76,19%
8	75,58%	79,63%	77,12%
9	75,45%	79,83%	77,08%
10	76,01%	81,21%	78,00%
11	76,55%	81,51%	78,36%
12	76,78%	81,44%	78,48%
13	76,17%	81,58%	78,03%
14	76,76%	81,56%	78,48%
15	76,38%	81,03%	78,06%
16	**76,44%**	**80,80%**	**79,09%**
17	76,45%	80,74%	78,06%
18	76,32%	80,65%	77,91%
19	75,99%	80,50%	77,64%
20	75,91%	80,35%	77,54%

Ao analisarmos a Tabela 5.1 e a Tabela 5.2 podemos constatar que os melhores parâmetros de geração dos espectrogramas para os descritores de textura LBP, LPQ e RLBP foram: um limite de frequência de 22000 Hz, com uma amplitude de sinal de 80 dB em seguimentos de áudio com duração igual a 30 segundos.

Para os descritores de textura GLCM e Filtros de Gabor o comportamento foi ligeiramente diferente, divergindo apenas quanto a amplitude de sinal utilizada que no caso destes deve ter um valor de 90dB. Por fim é possível afirmar que ao se utilizar descritores de características visuais, os Filtros de Gabor apresentaram as melhores taxas de acerto, sendo que os valores de *Recall*, *Precision* e *F-Measure* foram, respectivamente, 79,44%, 80,80% e 79,09%.

5.2 Etapa 2: Testes Utilizando Características Visuais com Zoneamento

Esta seção apresenta os resultados obtidos com o uso dos descritores de características visuais combinados técnicas de zoneamento dos espectrogramas, sendo apresentados apenas a regra de fusão que apresentou as mais elevadas taxas de acerto. Os resultados foram obtidos utilizando espectrogramas gerados apenas com os melhores parâmetros para cada um dos descritores empregados. Quanto ao zoneamento linear foram realizados testes com duas, três e quatro zonas. Já para o zoneamento na escala de mel foram utilizadas 13 zonas como foi falado anteriormente.

A Tabela 5.3 apresenta as taxas de acerto obtidas com o uso do zoneamento linear. E por meio dessa podemos constatar que o único descritor de textura que apresentou resultado inferior ao apresentado quando não se empregou o zoneamento do espectrograma foi o GLCM. Também constatamos que as melhores taxas de acerto foram obtidas por meio da Regra do Produto com o uso do RLBP em um espectrograma divido em 4 zonas, sendo os valores de *Recall*, *Precision* e *F-Measure* foram, respectivamente, 79,01%, 87,42% e 81,87%.

Há também os casos em que os resultados foram extremamente baixos como ocorreu com o GLCM, fato que se caracteriza por não haver informação de textura suficiente para ser empregado um descritor de textura baseado em representação estatística.

Tabela 5.3: Resultados dos testes realizados com zoneamento linear

Número de Zonas	**Regra de Fusão**	**Recall**	**Precison**	**F-Measure**
		LBP		
2	Produto	76,18%	83,05 %	78,46%
3	Produto	**78,29%**	**85,51** %	**80,74%**
4	Produto	77,66%	86,40 %	80,48%
		LPQ		
2	Produto	71,71%	81,20 %	74,32%
3	Soma	**72,10%**	**81,09%**	**77,95%**
4	Produto	72,52%	84,54 %	76,21%

continua na próxima página

continuação da Tabela Tabela 5.3				
Número de Zonas	**Regra de Fusão**	**Recall**	**Precison**	**F-Measure**
RLBP				
2	Produto	77,06%	84,20 %	79,49%
3	Produto	79,36%	86,08 %	81,76%
4	Produto	**79,01%**	**87,42%**	**81,87%**
Filtros de Gabor				
2	Produto	**76,97%**	**86,25%**	**79,56%**
3	Produto	74,82%	84,86 %	78,19%
4	Produto	76,44%	85,93 %	79,43%
GLCM				
2	Produto	2,17%	0,07 %	0,13%
3	Produto	**19,13%**	**54,98%**	**21,90%**
4	Soma	2,20%	2,24 %	0,17%

A Tabela 5.4 apresenta os resultados obtidos com o uso do zoneamento pela escala Mel. Ao realizarmos a analise da mesma, podemos verificar que todas as taxas de acerto apresentadas são inferiores à aquelas obtidas quando o espectrograma foi dividido linearmente. Também foi constatado que as mesmas também são inferiores as taxas obtidas quando não foi realizado nenhum zoneamento.

Novamente como ocorreu com o zoneamento linear ao se empregar o GLCM como descritor de textura os resultados apresentados foram extremamente baixos o que serve para reiterar a ineficácia do deste descritor quando se faz o uso do zoneamento do espectrograma.

Tabela 5.4: Resultados dos testes realizados com zoneamento pela escala Mel

Regra de Fusão	**Recall**	**Precison**	**F-Measure**
LBP			
Máximo	**73,09%**	**78,92** %	**75,03%**
LPQ			
Produto	2,17%	30,68 %	3,92%

continua na próxima página

continuação da Tabela Tabela 5.4			
Regra de Fusão	**Recall**	**Precison**	**F-Measure**
RLBP			
Máximo	72,32%	79,83%	74,66%
Filtros de Gabor			
Máximo	69,62%	78,02 %	73,18%
GLCM			
Produto	4,54%	30,68 %	7,90%

5.3 Etapa 3: Testes Utilizando Características Acústicas

Esta seção apresenta os resultados obtidos ao se utilizar os descritores de características acústicas RP, RH e SSD, que trabalham diretamente com a identificação de padrões de ritmo e timbre.

A Tabela 5.5 apresenta os resultados obtidos com a utilização dos descritores acima citados, e como pode ser visto, os melhores acertos foram obtidos com o uso do SSD, sendo que os valores de *Recall*, *Precison* e *F-Measure* foram, respectivamente, 79,68%, 83,37% e 81,33%.

Tabela 5.5: Resultados obtidos com o uso de descritores de características acústicas

Descritor	**Recall**	**Precison**	**F-Measure**
RH	28,36%	33,47 %	28,13%
RP	62,29%	67,76 %	64,00%
SSD	**79,68%**	**83,87** %	**81,33%**

5.4 Etapa 4: Testes Utilizando a Combinação de Características Acústicas e Visuais

A presente seção tem por finalidade apresentar a síntese dos resultados obtidos ao se combinar características acústicas e visuais. A combinação aqui aplicada foi realizada diretamente sobre a saída dos classificadores obtidos nos testes anteriores.

As combinações foram realizadas utilizando a saída do classificador (*predict*) obtida com o teste envolvendo o descritor de característsticas acústicas SSD, visto que este apresentou as melhores taxas de acerto quando comparado ao RP e ao RH. Quanto as saídas (*predicts*) dos classificadores de características visuais foram utilizadas aquelas que apresentaram as melhores taxas de acerto nos testes apresentados nas seções 5.1 e 5.2. A Figura 5.1 apresenta o diagrama que ilustra como o processo de fusão foi realizado.

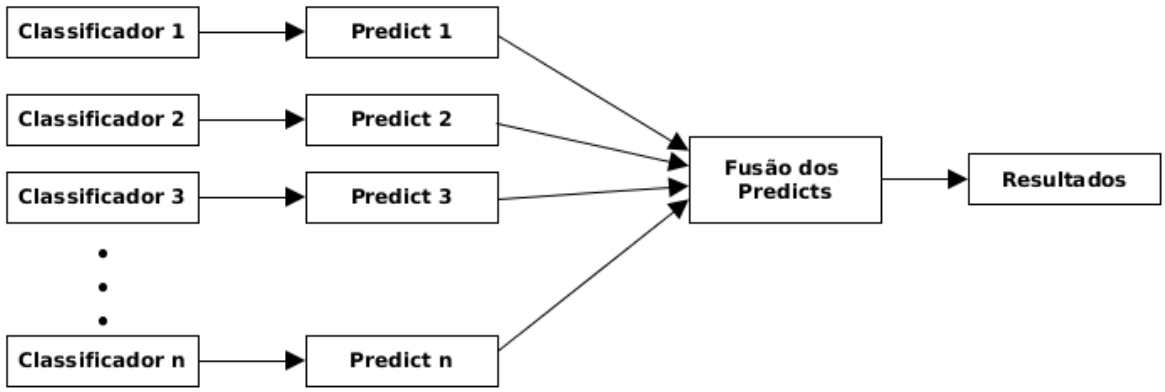


Figura 5.1: Esquematização da fusão dos resultados

A Tabela 5.6 apresenta os resultados obtidos com o uso da combinação dos melhores classificadores dos descritores de características visuais que foram apresentados na Seção 5.1, com o classificador obtido com o SSD que foi o descritor de características acústicas que apresentou as melhores taxas de acerto. Ao analisarmos essa podemos constatar que as melhores taxas de acerto foram dadas com o uso do RBLP, sendo que os valores de *Recall*, *Precision* e *F-measure* foram, respectivamente, 90,67%, 93,78% e 92,34% com o uso da Regra da Soma.

Tabela 5.6: Resultados obtidos com a combinação do descritor de características acústicas SSD com os descritores de características visuais sem zoneamento

Regra de Fusão	**Recall**	**Precison**	**F-Measure**
		LBP	
Produto	89,02%	93,09%	90,68%
		LPQ	
Produto	52,12%	60,87%	53,06%

continua na próxima página

continuação da Tabela Tabela 5.6			
Regra de Fusão	**Recall**	**Precison**	**F-Measure**
RLBP			
Soma	**91,08%**	**94,02%**	**92,34%**
Filtros de Gabor			
Soma	90,67%	93,78%	91,96%
GLCM			
Produto	84,70%	90,68%	86,92%

A Tabela 5.7 e a Tabela 5.8 apresentam os resultados obtidos com o uso da combinação dos melhores classificadores dos descritores de características visuais que foram apresentados na Seção 5.2, com o classificador obtido com o SSD que foi o descritor de características acústicas que apresentou as melhores taxas de acerto. Ao analisarmos essas podemos constatar que as melhores taxas de acerto foram dadas com o uso do RBLP. Para o zoneamento linear os valores de *Recall*, *Precision* e *F-measure* foram, respectivamente, 90,57%, 93,32% e 90,57%, com o uso da Regra da Soma. Quando se empregou o zoneamento pela escala Mel os valores de *Recall*, *Precision* e *F-measure* foram, respectivamente, 83,64%, 88,32% e 85,20% com a Regra da Soma.

Tabela 5.7: Resultados obtidos com a combinação do descritor de características acústicas SSD com os descritores de características visuais com zoneamento linear

Regra de Fusão	**Recall**	**Precison**	**F-Measure**
LBP			
Soma	**89,42%**	**88,25%**	**90,81%**
LPQ			
Soma	89,28%	92,55%	90,60%
RLBP			
Soma	90,57%	93,62%	90,57%
Filtros de Gabor			
Soma	88,93%	92,51%	90,35%
GLCM			

continua na próxima página

continuação da Tabela Tabela 5.7			
Regra de Fusão	**Recall**	**Precison**	**F-Measure**
Soma	78,99%	88,57%	82,43%

Tabela 5.8: Resultados obtidos com a combinação do descritor de características acústicas SSD com os descritores de características visuais com zoneamento pela escala Mel

Regra de Fusão	**Recall**	**Precison**	**F-Measure**
LBP			
Soma	**83,79%**	**88,25** %	**85,59%**
LPQ			
Máximo	80,96%	85,61%	82,80%
RLBP			
Soma	83,64%	88,32%	85,47%
Filtros de Gabor			
Soma	83,26%	88,23%	85,20%
GLCM			
Máximo	80,96%	85,61%	82,80%

5.5 Discussão

A Tabela 5.9 apresenta a síntese dos melhores resultados encontrados em todos os testes realizados no desenvolvimento deste trabalho.

Analisando a Tabela 5.9 é possível observar alguns aspectos envolvendo cada uma das etapas de testes realizadas. Na primeira etapa quando o espectrograma foi analisado em sua completude o melhor resultado foi dado com o uso do Filtro de Gabor como descritor de textura, e as medidas de avaliação alcançadas com a utilização do mesmo foram: *Recall* 76,44%, *Precision* 80,80% e *F-Measure* 79,09%.

Tabela 5.9: Melhores resultados

Descritor	**Zonas**	**Regra de Fusão**	**Recall**	**Precison**	**F-Measure**
LBP	Nenhum	-	74,70%	79,01 %	76,39%
LPQ	Nenhum	-	66,65%	73,06 %	68,52%

continua na próxima página

continuação da Tabela Tabela 5.9

Descritor	**Zonas**	**Regra de Fusão**	**Recall**	**Precison**	**F-Measure**
RLBP	Nenhum	-	74,94%	80,26 %	76,80%
GABOR	Nenhum	-	76,44%	80,80 %	79,09%
GLCM	Nenhum	-	45,59%	56,34 %	44,37%
LBP	Linear	Produto	78,29%	85,51 %	80,74%
LPQ	Linear	Soma	72,10%	81,09 %	77,95%
RLBP	Linear	Produto	79,01%	87,42 %	81,87%
GABOR	Linear	Produto	76,97%	86,25 %	79,56%
GLCM	Linear	Produto	19,13%	54,98 %	21,90%
LBP	Escala Mel	Máximo	73,09%	78,92 %	75,03%
LPQ	Escala Mel	Produto	2,17%	30,68 %	3,92%
RLBP	Escala Mel	Máximo	72,32%	79,83 %	74,66%
GABOR	Escala Mel	Máximo	69,62%	78,02 %	73,18%
GLCM	Escala Mel	Produto	4,54%	30,68 %	7,90%
SSD	Nenhum	-	79,68%	83,87 %	81,33%
RP	Nenhum	-	62,29%	67,76 %	64,00%
RH	Nenhum	-	28,36%	33,47 %	28,13%
LBP e SSD	Nenhum	Produto	89,02%	93,09 %	90,68%
LBP e SSD	Linear	Soma	89,42%	88,25 %	90,81%
LBP e SSD	Escala Mel	Soma	83,79%	88,25 %	85,59%
LPQ e SSD	Nenhum	Produto	52,12%	60,87 %	53,06%
LPQ e SSD	Linear	Soma	89,28%	92,55 %	90,60%
LPQ e SSD	Escala Mel	Máximo	80,96%	85,61 %	82,80%
RLBP e SSD	Nenhum	Soma	**91,08%**	**94,02** %	**92,34%**
RLBP e SSD	Linear	Soma	90,57%	93,62 %	90,57%
RLBP e SSD	Escala Mel	Soma	83,64%	88,32 %	85,47%
GABOR e SSD	Nenhum	Soma	90,67%	93,78 %	91,96%
GABOR e SSD	Linear	Soma	88,93%	92,51 %	90,35%
GABOR e SSD	Escala Mel	Soma	83,26%	88,23 %	85,20%
GLCM e SSD	Nenhum	Produto	80,47%	90,68 %	86,92%
GLCM e SSD	Linear	Soma	78,99%	88,57 %	82,43%
GLCM e SSD	Escala Mel	Máximo	80,96%	85,61 %	82,80%

Ao utilizarmos o zoneamento linear dos espectrogramas obtemos resultados melhores dos que os que foram encontrados sem a utilização do zoneamento da imagem para 4 descritores de textura, sendo estes, LBP, LPQ, RLBP e Filtro de Gabor. Todavia, quando se fez o uso do zoneamento do espectrograma da imagem pela escala Mel, todos os resultados foram inferiores ao teste em que não foi utilizada nenhuma técnica de zoamento. O melhor resultado encontrado ao utilizar o zoneamento da imagem foi dado pelo descritor

RLBP dividido em 4 zonas lineares, obtendo as seguintes medidas de avaliação: *Recall* 79,01%, *Precision* 87,42% e *F-Measure* 81,87%, quando se fez o uso da regra do produto.

Quanto aos testes em que foram utilizados descritores de caracterrísticas acústicas, podemos constatar que os melhores resultados foram encontrados ao se utilizar o SSD, que trabalha diretamente sobre o timbre das amostras de áudio, fato este que nos leva a crer que os descritores que trabalham com timbre de amostras de áudio apresentam grande relevância para tarefas que envolvem a classificação automática de espécies de pássaros. Os resultados apresentados pelo SSD mostrou taxas de acerto superiores as encontradas nos descritores de textura, sendo elas: *Recall* 79,68%, *Precision* 83,87% e *F-Measure* 81,83%.

No momento em que se combinou os resultados dos melhores classificadores baseados em características acústicas com características visuais, foi possível observar que para certas combinações de características há complementariedade entre os resultados proporcionando alcançar valores medidas de avaliação superiores a 90% nos melhores casos. Nestes experimentos o que apresentou as medidas de avaliação dos resultados mais elevadas foi aquele em que se combinou o classificador obtido com o RLBP sem zoneamento com o classificador obtido para o SSD, sendo que os valores alcançados foram: *Recall* 91,08%, *Precision* 94,02 e *F-Measure* 92,34%.

O resultados encontrados pela combinação do RLBP com o SSD não somente apresentou o melhor resultado dos testes envolvendo zoneamento como também apresentou o melhor resultado geral do trabalho, o que caracteriza que o RLBP combinado com o SSD é uma boa metodologia para a classificação automática de espéscies de pássaros.

6

Conclusão

A presente dissertação foi desenvolvida sobre a premissa da utilização de descritores de características acústicas e visuais para a criação de um sistema de classificação automático de espécies de pássaros, tendo por finalidade explorar também a existência da complementariedade entre classificadores baseados nesses diferentes tipos de características.

Nos experimentos iniciais foi constatado que o uso do Filtro de Gabor como descritor de características visuais apresentou os melhores resultados nos testes em que nenhum tipo de zoneamento foi aplicado sobre os espectrogramas, quando comparado a alguns dos principais descritores de textura, de diferentes abordagens, descritos na literatura. Nos testes seguintes, nos quais foram aplicadas técnicas de zoneamento sobre os espectrogramas e dentre todos os descritores de textura utilizados os melhores resultados foram alcançados com o uso do RLBP.

Quanto ao uso dos descritores de características acústicos, estes foram escolhidos por apresentarem bons resultados em trabalhos previamente descritos na literatura, tanto relacionados à classificação de espécie de pássaros, quanto a outros domínios de aplicação. Com o uso desses descritores foi constatado que o SSD apresentou as melhores taxas de acerto. E devido a este fato, o SSD foi empregado na combinação com características visuais. Com a execução desses experimentos foi verificado que há complementariedade entre os dois conjuntos de características distintos, sendo que o melhor resultado encontrado foi dado pela combinação do SSD com o RLBP, que apresentou as melhores taxas de acerto do trabalho com as medidas de avaliação *Recall*: 91,08%, *Precision*: 94,02% e *F-Measure*: 92,34%.

Tendo por base as explanações apresentadas, podemos verificar a eficácia do método de classificação de espécies de pássaros aqui proposto assim como também a complemen-

tariedade entre os descritores de características acústicas e visuais como pode ser visto na Seção 5.5, o que caracteriza a validade da hipótese apresentada neste trabalho.

6.1 Trabalhos futuros

Para trabalhos futuros, tem-se como objetivo sintetizar novas bases de dados com um número maior de espécies, tendo por finalidade verificar a eficácia do método de classificação aqui proposto, sobre um conjunto amostral maior. Também se estuda a possibilidade de aplicar outras técnicas de extração de características, aliadas a mecanismos de classificação diferentes do SVM, visando realizar uma comparação direta com os resultados apresentados neste trabalho, contribuindo com o desenvolvimento contínuo do estado da arte.

REFERÊNCIAS

Abe, S. *Support Vector Machines for Pattern Classification*. Advances in Pattern Recognition. London: Springer London, 2010.

Anderson, S. E.; Dave, A. S.; Margoliash, D. Template-based automatic recognition of birdsong syllables from continuous recordings. *The Journal of the Acoustical Society of America*, v. 100, n. 2 Pt 1, p. 1209–1219, 1996.

Bardeli, R.; Wolff, D.; Kurth, F.; Koch, M.; Tauchert, K.-H.; Frommolt, K.-H. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, v. 31, n. 12, p. 1524 – 1534, pattern Recognition of Non-Speech Audio, 2010.

Bernardini, F. C. *Combinação de classificadores simbólicos utilizando medidas de regras de conhecimento e algoritmos genéticos*. Tese de Doutoramento, Universidade de São Paulo, 2006.

Briggs, F.; Raich, R.; Fern, X. Z. Audio classification of bird species: A statistical manifold approach. In: 0010, W. W.; Kargupta, H.; Ranka, S.; Yu, P. S.; Wu, X., eds. *ICDM*, IEEE Computer Society, 2009, p. 51–60.

Cai, J.; Ee, D.; Pham, B.; Roe, P.; Zhang, J. Sensor Network for the Monitoring of Ecosystem: Bird Species Recognition. In: *3rd International Conference on Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007*, 2007, p. 293–298.

Carpenter, F. L. A spectrum of nectar-eater communities. *American Zoologist*, v. 18, n. 4, p. 809–819, 1978.

Casanova, M. A. *Utilizando aprendizado de máquina para construção de uma ferramenta de apoio a moderação de comentários*. Tese de Doutoramento, PUC-Rio, 2011.

Catchpole, C. *Bird song: biological themes and variations.* 2nd ed. Cambridge [England] ; New York: Cambridge University Press, 2008.

Chang, C.-C.; Lin, C.-J. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, v. 2, n. 3, p. 27, 2011.

Chen, J.; Kellokumpu, V.; Zhao, G.; Pietikainen, Z. RLBP: Robust Local Binary Pattern. In: *Proceedings of the British Machine Vision Conference*, BMVA Press, 2013.

Chou, C.-H.; Lee, C.-H.; Ni, H.-W. Bird species recognition by comparing the HMMs of the syllables. In: *Second International Conference on Innovative Computing, Information and Control, 2007. ICICIC '07*, 2007, p. 143–143.

Chou, C.-H.; Liu, P.-H. Bird species recognition by wavelet transformation of a section of birdsong. In: *Ubiquitous, Autonomic and Trusted Computing, 2009. UIC-ATC'09. Symposia and Workshops on*, IEEE, 2009, p. 189–193.

Costa, Y.; Oliveira, L.; Koerich, A.; Gouyon, F. Music genre recognition using gabor filters and lpq texture descriptors. In: Ruiz-Shulcloper, J.; Sanniti di Baja, G., eds. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, v. 8259 de *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, p. 67–74, 2013.

Costa, Y. M. *Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores.* Tese de Doutoramento, Universidade Federal do Paranï¿½, 2013.

Costa, Y. M.; Oliveira, L. S.; Koerich, A. L.; Gouyon, F. Classificação de gêneros musicais por texturas no espaço de frequência. *XXXVIII SeminÃ¡rio Integrado de Software e Hardware*, 2011.

Duda, R. O.; Hart, P. E.; Stork, D. G. *Pattern classification.* 2nd ed. New York: Wiley, 2001.

Duda, R. O.; Hart, P. E.; et al. *Pattern classification and scene analysis*, v. 3. Wiley New York, 1973.

Fagerlund, S. Bird Species Recognition Using Support Vector Machines. *European Association for Signal Processing(EURASIP) J. Appl. Signal Process.*, v. 2007, n. 1, p. 64–64, 2007.

Fastl, H.; Zwicker, E. *Psychoacoustics: facts and models.* N. 22 in Springer series in information sciences, 3rd. ed ed. Berlin ; New York: Springer, 2007.

Feinsinger, P.; Colwell, R. K. Community organization among neotropical nectar-feeding birds. *American Zoologist*, v. 18, n. 4, p. 779–795, 1978.

Gioppo, L. L.; Kaestner, C. A. A. Análise da viabilidade da construção de uma unidade de gravação autônoma de cantos de pássaros com equipamentos disponíveis no mercado. *XVI Seminário de Iniciação Cientifica e Tecnológica da UTFPR (SICITE)*, 2011.

Goëau, H.; Glotin, H.; Vellinga, W.-P.; Planqué, R.; Rauber, A.; Joly, A. Lifeclef bird identification task 2014. In: *Conference and Labs of Evaluation Forum, 2014. CLEF2014*, 2014.

Gonzalez, R. C.; Woods, R. E. *Digital image processing.* 3rd ed. Upper Saddle River, N.J: Prentice Hall, 2008.

Haralick, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE*, v. 67, n. 5, p. 786–804, 1979.

Holmes, R. T. *Avian foraging: Theory, methodology and applications. studies in avian biology 13*, cáp. Ecological and evolutionary impact of bird predation on forest insects: an overview Allen Press, Inc., p. 6–13, 1990.

Holmes, R. T.; Schultz, J. C.; Nothnagle, P. Bird Predation on Forest Insects: An Exclosure Experiment. *Science*, v. 206, n. 4417, p. 462–463, 1979.

Kittler, J.; Hatef, M.; Duin, R.; Matas, J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 3, p. 226–239, 1998.

Kogan, J. A.; Margoliash, D. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. *The Journal of the Acoustical Society of America*, v. 103, n. 4, p. 2185–2196, 1998.

Kwan, C.; Ho, K. C.; Mei, G.; Li, Y.; Ren, Z.; Xu, R.; Zhang, Y.; Lao, D.; Stevenson, M.; Stanford, V.; Rochet, C. An automated acoustic system to monitor and classify birds. *EURASIP J. Appl. Signal Process.*, v. 2006, p. 52–52, 2006.

Kwan, C.; Mei, G.; Zhao, X.; Ren, Z.; Xu, R.; Stanford, V.; Rochet, C.; Aube, J.; Ho, K. Bird classification algorithms: theory and experimental results. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*, 2004, p. V–289–92 vol.5.

Lee, C.-H.; Han, C.-C.; Chuang, C.-C. Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 16, n. 8, p. 1541–1550, 2008.

Li, W.; Mao, K.; Zhang, H.; Chai, T. Selection of gabor filters for improved texture feature extraction. In: *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, p. 361–364.

Lidy, T.; Rauber, A. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR '05)*, 2005, p. 34–41.

Lopes, M.; Gioppo, L.; Higushi, T.; Kaestner, C.; Silla, C.; Koerich, A. Automatic Bird Species Identification for Large Number of Species. In: *2011 IEEE International Symposium on Multimedia (ISM)*, 2011a, p. 117–122.

Lopes, M.; Lameiras Koerich, A.; Nascimento Silla, C.; Alves Kaestner, C. Feature set comparison for automatic bird species identification. In: *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2011b, p. 965–970.

Lucio, D. R.; Costa, Y. M. G. Bird species classification using spectrograms. In: *Computing Conference (CLEI), 2015 Latin American*, IEEE, 2015, p. 1–11.

McIlraith, A. L.; Card, H. C. Birdsong recognition using backpropagation and multivariate statistics. *Signal Processing, IEEE Transactions on*, v. 45, n. 11, p. 2740–2748, 1997.

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, 1996.

Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 971–987, 2002.

Ojansivu, V.; Heikkilä, J. Blur Insensitive Texture Classification Using Local Phase Quantization. In: *Proceedings of the 3rd International Conference on Image and Signal Processing*, Berlin, Heidelberg: Springer-Verlag, 2008, p. 236–243 (*ICISP '08*, v.1).

Proctor, M.; Yeo, P.; Lack, A.; et al. *The natural history of pollination.* HarperCollins Publishers, 1996.

Rauber, A.; Frühwirth, M. Automatically Analyzing and Organizing Music Archives. In: Goos, G.; Hartmanis, J.; van Leeuwen, J.; Constantopoulos, P.; SÃ¸lvberg, I. T., eds. *Research and Advanced Technology for Digital Libraries*, v. 2163, Berlin, Heidelberg: Springer Berlin Heidelberg, p. 402–414, 2001.

Rauber, A.; Pampalk, E.; Merkl, D. Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by sound similarity. In: *Proc. ISMIR*, 2002, p. 71–80.

S. S. Stevens, J. V. The relation of pitch to frequency: A revised scale. *The American Journal of Psychology*, v. 53, n. 3, p. 329–353, 1940.

Selouani, S.; Kardouchi, M.; Hervet, E.; Roy, D. Automatic birdsong recognition based on autoregressive time-delay neural networks. In: *2005 ICSC Congress on Computational Intelligence Methods and Applications*, 2005, p. 6 pp.–.

Semolini, R. *Support vector machines, inferência transdutiva e o problema de classificação.* Tese de Doutoramento, Universidade Estadual de Campinas, 2002.

Snow, D. W. Evolutionary aspects of fruit-eating by birds. *Ibis*, v. 113, n. 2, p. 194–202, 1971.

Snow, D. W. Tropical Frugivorous Birds and Their Food Plants: A World Survey. *Biotropica*, v. 13, n. 1, p. 1, 1981.

Straube, F. C.; Bianconi, G. V. Sobre a grandeza e a unidade utilizada para estimar esforço de captura com utilização de redes-de-neblina. *Chiroptera Neotropical*, v. 8, n. 1-2, p. 150–152, 2014.

Tyagi, H.; Hegde, R. M.; Murthy, H. A.; Prabhakar, A. Automatic identification of bird calls using Spectral Ensemble Average Voice Prints. In: *Signal Processing Conference, 2006 14th European*, 2006, p. 1–5.

Vapnik, V. N. *The nature of statistical learning theory.* Statistics for engineering and information science, 2nd ed ed. New York: Springer, 2000.

Vilches, E.; Escobar, I.; Vallejo, E.; Taylor, C. Data mining applied to acoustic bird species recognition. In: *18th International Conference on Pattern Recognition, 2006. ICPR 2006*, 2006, p. 400–403.

Wu, M. J.; Chen, Z. S.; Jang, J. S. R.; Ren, J. M.; Li, Y. H.; Lu, C. H. Combining visual and acoustic features for music genre classification. In: *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, 2011, p. 124–129.