Bird Species Classification Using Spectrograms

Diego Rafael Lucio
Programa de Pós Graduação em Ciência da Computação
Universidade Estadual de Maringá
Avenida Colombo, 5790 - Jardim Universitário,
Maringá - Paraná - Brasil
Email: diegorafaellucio@gmail.com

Yandre Maldonado e Gomes da Costa
Programa de Pós Graduação em Ciência da Computação
Universidade Estadual de Maringá
Avenida Colombo, 5790 - Jardim Universitário,
Maringá - Paraná - Brasil
Email: yandre@din.uem.br

***Resumo*—This paper describes a system for automatic bird species classification based on features taken from the textural content of spectrogram images. The texture features are extracted using three of the most common texture operators described in the Digital Image Processing literature: Local Binary Pattern (LBP), Local Phase Quantization (LPQ) and Gabor Filters. Aiming to perform more fare comparisons, the experiments were performed over a database already used in other works presented in the literature. In the classification step, SVM classifier was used and the final results were taken using 10-fold cross validation. The experiments were performed over a challenger dataset composed of 46 classes, and the best accuracy rate obtained is about 77.65%.**

***Index Terms*—Bird species classification. Spectrogam. Pattern recognition.**

I. Introdução

O interesse em reconhecimento automático de espécies de pássaros tem aumentado e várias técnicas de reconhecimento tem sido exploradas, tais como: características visuais de imagens de pássaros e característica acústicas a partir da vocalização das espécies. Alguns exemplos de características visuais de imagens de pássaros podem ser vistos nos trabalhos aprentados por Marini et al. [1], Pang et al. [2], Huang et al. [3] e Cullinan et al. [4]. Os trabalhos apresentados por Fagerlund [5], Anderson et al. [6] e Selouani et al. [7]. 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 [8].

Para manter a conservação da biodiversidade das espécies de pássaros é necessário obter o conhecimento supracitado, visto que as aves desempenham papéis de grande importância para o nosso ecossistema. [9]–[14].

A identificação das espécies de pássaros é um problema bem conhecido dos ornitólogos. Para realizar o reconhecimento especialistas sugerem técnicas não invasivas para coletar dados, devido a preocupação com o bem estar das espécies dos pássaros. O uso de técnicas de bioacústica para identificação das mesmas a partir dos registros de áudio capturados na natureza tem sido uma das técnicas utilizadas, devido ao fato desta ter se mostrado muito eficaz [15] [16].

No entanto, o emprego destas técnicas somente foi possível com o desenvolvimento tecnológico, pois o uso de dispositivos de tamanho reduzido possibilitou a gravação de sons emitidos pelos pássaros de forma menos invasiva, ou seja o contato direto com as aves não se tornou mais necessário, para realizar a identificação. Isso é o que caracteriza uma técnica de monitoramento não invasiva.

De forma contrária à técnicas descritas anteriormente, há as técnicas denominadas invasivas, 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 [17]. Independente da técnica ser invasiva ou não, com base nos registros de áudio armazenados é possível realizar a classificação das espécies aplicando técnicas de processamento de sinais aliadas a técnicas de aprendizagem de máquina [18]. Contudo é importante considerar alguns aspectos quando se realiza a coleta de amostras de sinal de áudio de espécies de pássaros. De acordo com Bardeli el al. [19] e Agranat [20], em ambientes reais há alguns problemas como o ruído causado pelo vento, assim como também a sobreposição de sinais de áudio, que acabam por afetar mais de cinquenta porcento das gravações.

Diante do exposto, o objetivo deste trabalho é investigar um novo método para o reconhecimento de espécies de pássaros, fazendo o uso características visuais, obtidas de espectrogramas que são representações visuais do espectro das frequências do som pelo tempo,cujo principal atributo visual é a textura. Esses são gerados a partir do sinal de áudio colhido do canto dos pássaros. A Figura 1 é um exemplo de espectrograma gerado a partir de uma amostra de áudio.

Os experimentos baseados no sistema de classificação proposto alcançaram, no melhor caso uma taxa de acerto de 77,65%. Sendo esta uma taxa comparável a de outros trabalhos recentes aplicados sobre o subconjunto de espécies retirados do mesmo repositório.

Este trabalho encontra-se organizado da seguinte forma: seção 2 descreve os trabalhos e conceitos envolvidos em sistemas de classificação automática de pássaros; na seção 3 é apresentada a fundamentação teórica a cerca de extração de características de textura das imagens, assim como também são apresentadas algumas medidas de avaliação comumente utilizadas em sistemas de classificação; a seção 4 apresenta a base de dados utilizada; a seção 5 descreve o protocolo experimental empregado no trabalho; a seção 6 apresenta os resultados obtidos e a discusão acerca destes; a seção 7

978-1-4673-9143-6/15/$31.00

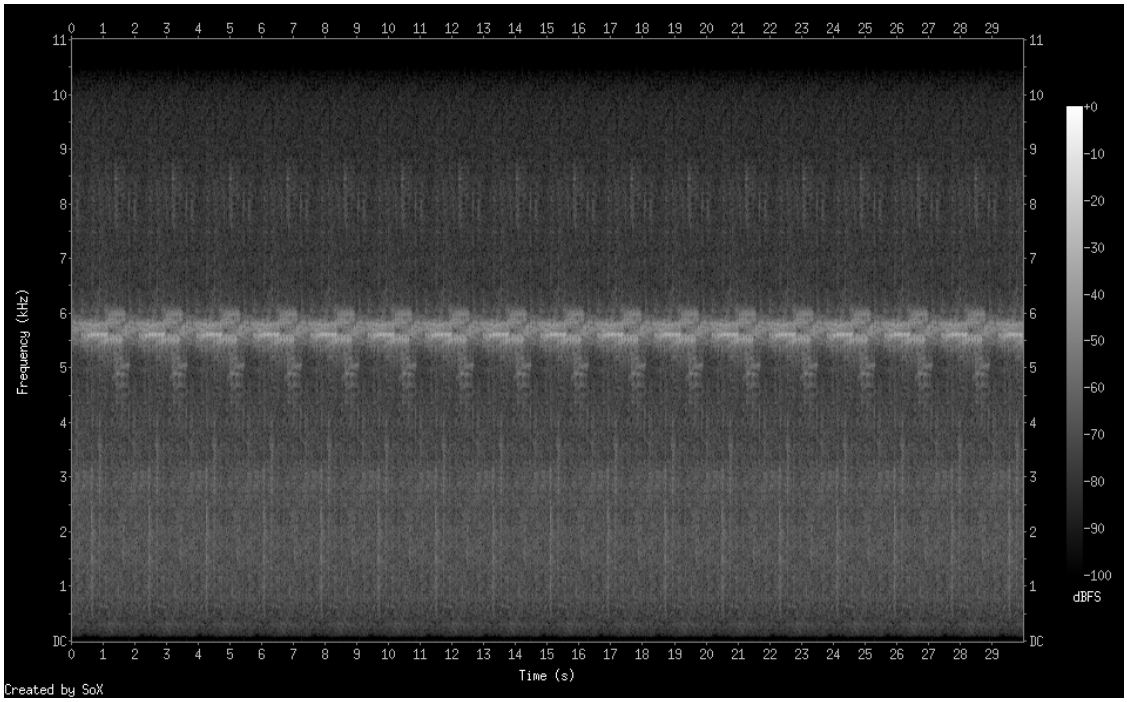


Figura 1: Exemplo de espectrograma gerado a partir de amostra de áudio de 30 segundos

apresenta as consideração finais sobre o trabalho e possíveis trabalhos futuros.

II. Conceitos e Trabalhos Relacionados

O interesse no reconhecimento de espécies de pássaros baseado na sua vocalização tem aumentado e muitos estudos têm sido publicados recentemente. O reconhecimento de espécies de pássaros é um típico problema em que se pode aplicar reconhecimento de padrões, processamento de sinais, extração de características e elaboração de um esquema de classificação.

Nos trabalhos apresentados por Anderson et al. [6] e Kogan e Margoliash [21] foram descritas as primeiras tentativas de reconhecimento automático de espécies de pássaros por meio de seus sons. Em ambos os trabalhos foram aplicados *Dynamic Time Warping* (DTW) e *Hidden Markov Models* (HMM) para o reconhecimento automático de duas espécies de pássaros. Nesses estudos, as amostras de som foram divididas em sílabas que são representadas pelos DTW e HMM das mesmas e a classificação é dada pela distância euclidiana da amostra em questão para com a base de dados. As taxas de acerto apresentadas nos trabalhos foram respectivamente 97,00% e 82,00%.

Chou et al. [22], também fazem o uso de HMM, no entanto, este é utilizado como o conjuntos 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%. É importante observar que neste trabalho as amostras utilizadas foram obtidas de sinal gravado em estúdio, não acometido por ruídos ambientais. Adicionalmente, todas as sílabas de cada classe foram extraídas de uma mesma amostra.

Fagerlund [5] utilizou o *Mel Frequency Cepstral Coefficient* (MFCC) das amostra de áudio como características para o sistema de classificação. No entanto, assim como em [23], fizeram o uso de SVM para a classificação das amostras de áudio das bases de dados utilizados no projeto uma contendo 6 espécies e a outra contendo 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.

Nos trabalhos [18], [24], [25] e [26], foram utilizadas 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. Os melhores resultados encontrados são respectivamente 95,10%, 99,70%, 99,70% e 95,10%.

A Tabela I apresenta a lista de alguns trabalhos mais relevantes encontrados na literatura relacionados ao reconhecimento automático de espécies de pássaros, assim como também o mecanismos de reconhecimento empregado, a quantidade de espécies utilizadas e suas respectivas taxas de acerto.

III. Fundamentação Teórica

O problema de classificação, pode ser descrito como o processo pelo qual padrões ou sinais recebidos são distribuídos por um número prescrito de classes com o uso de alguma técnica de aprendizagem. Esta presente em muitas áreas de atuação, a classificação representa um amplo conjunto de problemas de grande significado prático [27].

Aplica-se a tarefas que o ser humano frequentemente executa sem dificuldades. Em que dados são recebidos do mundo

Tabela I: Relação de trabalhos e seus respectivos conjuntos de características e classificadores

Trabalho	Características	Classificador	Qtde de espécies	Melhor Acerto
Anderson el al.	Sílabas e amostras de áudio	HMM e DTW	2 espécies	97,00%*
Kogan e Margoliash	Sílabas e amostras de áudio	HMM e DTW	2 espécies	98,70%*
Chou et al.	HMM de sílabas	Algoritmo de Viterbi	420 espécies	78,20%*
Fagerlund	MFCC e parâmetros descritivos das sílabas	SVM	6 espécies	93,00%*
			8 espécies	97,00%*
Lopes e Kaestner	Características acústicas obtidas com MARSYAS	Naive Bayes, KNN (k=3), J4.8, MLP, SMO(Polynomial) e SMO(Pearson)	Xeno-Canto 3 espécies	95,10%**
			5 espécies	89,30%**
			8 espécies	89,30%**
			12 espécies	82,90%**
			20 espécies	78,20%**
Lopes e Kaestner	Características acústicas obtidas com MARSYAS, IOIHC e Sound Ruler	Naive Bayes, KNN (k=3), J4.8, MLP, SMO(Polynomial) e SMO(Pearson)	Xeno-Canto 3 espécies	99,70%**
Lopes et al.	Características acústicas obtidas com MARSYAS, IOIHC e Sound Ruler	Naive Bayes, KNN (k=3), J4.8, MLP, SMO(Polynomial) e SMO(Pearson)	Xeno-Canto 3 espécies	98,39%**
Lopes et al.	Características acústicas obtidas com MARSYAS, IOIHC e Sound Ruler	Naive Bayes, KNN (k=3), J4.8, MLP, SMO(Polynomial) e SMO(Pearson)	Xeno-Canto 3 espécies	95,10%**
			5 espécies	89,30%**
			8 espécies	89,30%**
			12 espécies	82,90%**
			20 espécies	78,20%**

*Taxa de acerto
**Resultados calculados com F-measure

exterior através de nossos sentidos e assim realizamos o reconhecimento destes, em algum contexto. Esta é realizada de maneira quase que imediata e com praticamente nenhum esforço, caso o conhecimento necessário para executar a classificação já tenha sido adquirido através de um processo de aprendizagem [27].

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 muitas dificuldades para executar a classificação, sendo assim, a automatização do processo de classificação passa a ser um grande interesse e a sua viabilidade aumenta conforme cresce o poder de processamento e memória dos computadores [27].

A abordagem clássica de um sistema de classificação é dividida em etapas bem definidas, sendo as principais: pré-processamento, extração de características e classificação [27]–[29], conforme ilustrado na Figura 2.

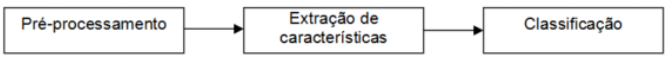


Figura 2: Principais etapas para o desenvolvimento de um sistema de reconhecimento de padrões

A etapa de pré-processamento dos itens 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 comummente 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. Em geral, o sinal está descrito em forma de imagem e, por isso, procura-se extrair descritores de atributos visuais como cor, textura e estrutura entre outros. Na última etapa, algoritmos de classificação bastante conhecidos são utilizados sobre os descritores extraídos a fim de se atribuir uma classe para cada padrão submetido ao sistema [27].

Cada uma das etapas apresentadas na Figura 2, possui uma grande quantidade de trabalhos abordando-as. No entanto, na literatura muitos esquemas diferentes vem sendo desenvolvidos para as etapas de extração de características e classificação, decorrente do fato, destas serem particularmente desafiadoras e decisivas no desempenho final do esquema de classificação.

Dos trabalhos para criação de um sistema de classificação baseado em características visuais, várias metodologias para a extração destas vem sendo aplicadas, tais como Filtros de Gabor, *Local Binary Pattern* (LBP) e *Local Phase Quantization* (LPQ).

Na etapa de classificação são empregados algoritmos projetados para atribuir uma classe, ou predições para classes previstas no problema, aos padrões que lhe são fornecidos como entrada. Existem muitos algoritmos projetados para este fim. Neste trabalho, optou-se pelo uso de *Support Vector Machine* (SVM) por ser um algoritmo de classificação que vem obtendo, sistematicamente, bom desempenho em diferentes domínios de aplicação.

As próximas seções apresentam uma breve descrição dos conceitos de extração de características, classificação e medidas de classificação utilizados no desenvolvimento deste trabalho

A. *Extração de Características*

A extração de características é uma etapa de grande importância para o desenvolvimento de um sistema de reconhecimento de padrões. Em sistemas de classificação de sinais de

áudio com conteúdo musical, as características devem ser relacionadas às principais dimensões da música, incluindo melodia, harmonia, ritmo, timbre e localização espacial. Embora o conjunto de características citado anteriormente seja utilizado originalmente em trabalhos de recuperação de informação músical, também é comummente utilizado em trabalhos que reconhecem canto de pássaros.

Considerando que neste trabalho serão exploradas características obtidas de imagens de espectrogramas gerados a partir do sinal do áudio dos sons de pássaros, cujo principal atributo visual é a textura, a seção III-A1 apresenta algumas das abordagens apresentadas na literatura para a extração de características de textura que foram utilizadas nos experimentos descritos neste trabalho.

1) Características Visuais:

- **Local Binary Pattern (LBP):** é uma técnica de representação de textura apresentada pela primeira vez por [30] como uma medida complementar para contraste da imagem. Posteriormente foi adaptado e se tornou uma abordagem estrutural para descrição de textura [31]. 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 cada um pixel central C, com seus P vizinhos equidistantes considerando uma distância R, com pode ser visto na Figura 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 [31], boa parte da informação sobre características de textura é preservada na distribuição T descrita na Equação 1.

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

  onde 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 à 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 2, 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.

$$T \approx (s(g_0 - g_c), ..., s(g_{P-1} - g_c) \tag{2}$$

  na qual

$$s(g_i - g_c) = \begin{cases} 1 & \text{se } g_i - g_c \geq 0 \\ 0 & \text{se } g_i - g_c \leq 0 \end{cases} \tag{3}$$

  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 através 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 4 descreve como este código é obtido.

$$LBP_{P,R}(X_C, Y_C) = \sum_{P=0}^{P-1} s(g_P - g_C)2^P \tag{4}$$

  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 [31]. 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 4.
  Assim ao invés de utilizar integralmente o histograma de padrões LBP, cujo tamanho é 2^P, é possível utilizar apenas os valores associados a padrões uniformes, constituindo um vetor com menor dimensionalidade, com apenas 59 características. De acordo com [31], além das 58 combinações uniformes, todos os padrões não uniformes encontrados devem participar de uma coluna adicional no histograma gerado. Devido a este fato, o vetor de características LBP para na configuração de 8 vizinhos com distância 2 possui 59 características em sua constituição.
- **Local Phase Quantization (LPQ):** é uma técnica para descrição de textura apresentada por Ojansivu, Ville and Heikkilä [32] originalmente utilizadas em imagens que apresentam borramento, todavia, è interessante observar que, embora o método tenha sido criado com este propósito, ele também produz resultados muito bons para imagens que não apresentam borramento.

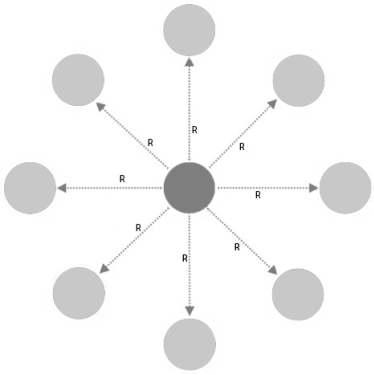
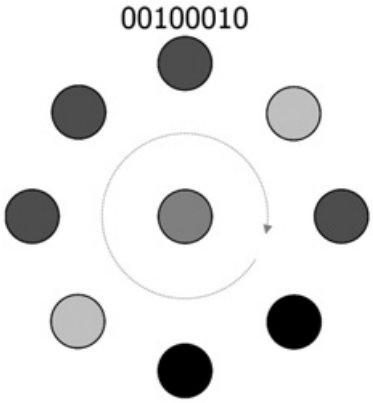
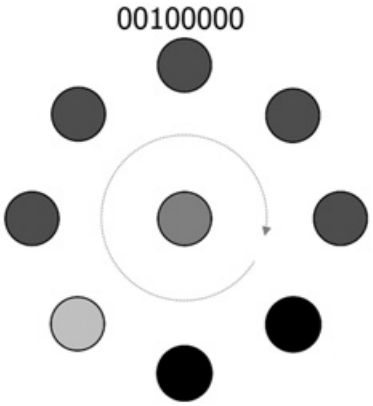


Figura 3: O operador LBP. O *pixel C*, escuro no centro, e os *pixels* claros são os *P* vizinhos



(a) Representação de um padrão não uniforme



(b) Representação de um padrão uniforme

Figura 4: Uniformidade do padrão LBP

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 5.

$$\hat{f}_{u_i}(x) = (f \times \phi_{u_i})x \tag{5}$$

sendo o filtro ϕ_{u_i} dado pela equação

$$\phi_{u_i} = e^{-j2\pi u_i^T} | y \in \mathbb{Z}^2 || y || \infty \leq r \tag{6}$$

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(equação)5 é expressa através do vetor de notação conforme a equação 7

$$\hat{f}_{u_i}(x) = w_{u_i}^T f(x) \tag{7}$$

sendo $F = [f(x_1), f(x_2), \cdots, f(x_{x^2})]$ 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 $w_I = Im[W_u1, W_u2, W_u3, W_u4]$. O $Re[]$ 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 $\hat{F} = wF$.

[32] 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 relacionada 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 8. 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 através 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 $D' = V^T DV$.

$$C_{i,j} = \sigma^{||x_i - x_j||} \tag{8}$$

Os coeficientes são quantizados usando-se a equação 9, em que e_{ij} são componentes de E Estes elementos são transformados de binário para decimal através da equação 10 e caracterizam valores inteiros compreendidos entre 0 e 255. Então, através de todas as posições da Imagem, é composto o vetor de 256 posições que correspondem ao histograma do LPQ.

$$q_{ij} = \begin{cases} 1 & \text{se } e_{ij} \geq 0 \\ 0 & \text{caso contrário} \end{cases} \tag{9}$$

$$b_j = \sum_{i=0}^{7} q_{ij} 2^i \tag{10}$$

- **Filtros de Gabor:** durante muito tempo um sinal podia ser representado em função do tempo ou, alternativamente, em função da frequência através da transformada de Fourier. Entretanto esta 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 tempo. Em seu trabalho original Gabor buscava a síntese do sinal e preocupou-se em como um sinal poderia ser construído através da combinação linear de funções lineares [33]. Os filtros de Gabor correspondem à 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. Uma possível forma para filtros de Gabor bidimensionais no domínio espacial, portanto apropriados para imagens digitais, é dada pelas equações 11 e 11.

$$\Psi(x,y) = exp\left(-\left(\frac{x^2+Y^2}{2\sigma^2}\right)\right) exp\left(\frac{j2\pi x}{\lambda}\right) \quad (11)$$

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 11, a saída do filtro de Gabor é obtida pela convolução da imagem de entrada com o filtro de Gabor apresentado na equação 12.

$$\sum_x \sum_y I(m-x, n-y)\Psi(x,y) \quad (12)$$

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, pose-se rotacionar $\Psi(x,y)$ com um ângulo θ para construir $\Psi(x',y')$ para a detecção de linhas com diferentes orientações. Neste caso, x' e y' podem ser encontrados pelas equações 13 e 14 respectivamente.

$$x' = x\cos\theta + y\sin\theta \quad (13)$$

$$y' = x\sin\theta + y\cos\theta \quad (14)$$

B. Medidas de Avaliação

Esta seção apresenta os critérios comumente utilizados para avaliar a eficiência de sistemas de classificação, sendo estes: *precison*, *recall*, *F-measure* e *Macro-F*. As subseções seguintes apresentam os critérios de avaliação citados [34] .

1) Precision: É o total de exemplos corretamente classificados como uma classe C sobre o total de exemplos classificados como a classe C, a desvantagem desta métrica pe que ela não leva em consideração os exemplos de deveriam ter sido reprovados, mas foram aprovados. Sua formula é expressa pela equação 15.

$$Precision(C_i) = \frac{M(C_i, C_i)}{M(*, C_i)} \quad (15)$$

2) Recall: É o total de exemplos corretamente classificados como uma classe C sobre o total de exemplos pertencentes a classe C presentes co conjunto de dados, a desvantagem desta métrica é que ela não leva em consideração todas as medidas. Sua formula é expressa pela equação 16.

$$Recall(C_i) = \frac{M(C_i, C_i)}{M(C_i, *)} \quad (16)$$

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 17.

$$F - measure(C) = \frac{2 \times recall(C) \times precision(C)}{recall(C) + precision(C)} \quad (17)$$

em que C é a classe sobre a qual o valor está sendo calculado.

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 18.

$$Macro - F(h) = \frac{1}{k}\sum_{i=1}^{k} F - measure(C_i) \quad (18)$$

em que k é o total de classes presentes no conjunto de amostras.

IV. Base de Dados dos Cantos

Os sons emitidos por espécies de pássaros são tipicamente divididos em cantos e chamados, dependendo da sua função. Geralmente as cantos são mais longos e mais complexos do que os chamados e ocorrem espontaneamente. A principal função dos cantos é relacionada ao acasalamento e a defesa territorial. Muitas espécies de pássaros cantam somente durante a temporada de acasalamento e essa ação é originalmente mais limitada aos machos. Os chamados são tipicamente curtos e carregam uma função, por exemplo, um alarme, voo, ou alimentação [5].

Além da divisão entre cantos e chamados, os sons emitidos pelos pássaros também são divididos em nível hierárquico de frases, sílabas e elementos. Embora haja uma certa imprecisão na forma como essas entidades se organizam, pode-se fazer algumas considerações acerca delas. 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 [35].

A base de dados utilizada nos experimentos descritos neste trabalho foi inspirada na metodologia utilizada no trabalho

Automatic Bird Species Identification for Large Number of Species [18]. A base é composta por um subconjunto da base de dados de cantos e chamados disponibilizada pelo site Xeno-Canto [1]. As gravações disponibilizadas pelo site foram realizadas diretamente em ambientes reais, sem qualquer filtro ou pré-processamento, sendo assim as amostras contém sons de outras espécies de pássaros e animais, assim como os ruídos do ambiente.

Os tópicos seguites descrevem o critério utilizado para a síntese da base de dados de sinais de áudio utilizada por baseado no esquema proposto por Lopes et al. [18] :

- Por meio do uso da ferramenta de busca do site supracitado, foi delimitada a localização geográfica, esta abrangeu um raio de 250 KM a partir da cidade de Curitiba;
- Dos registros apresentados como resultado, foram selecionados os de 75 espécies que apresentaram os registros de áudio com as mais altas frequências acústicas, como 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, com isso restaram 73 espécies;
- Foi efetuado o *download* dos registros do site através 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 destes segmentos caracterizarem melhor a vocalização do pássaro, visto que, de acordo com a literatura a utilização destes pulsos de áudio melhoram os resultados do processo de identificação [18]. O processo de divisão do sinal é exemplificado na Figura 5, onde é apresentado o áudio original de uma de uma espécie e os pulsos correspondentes a este;
- Foram desprezados todos os pulsos obtidos com tempo inferior a 0,1 segundo, e posteriormente foram eliminadas as espécies de pássaros que continham um número de instâncias inferior a 10, visto que este é o numero de folds aplicado no sistema de classificação, onde se obteve um total de 2814 amostras de de áudio distribuídas entre 46 espécies.
- Foi realizada a concatenação de cada uma das amostras de áudio com ela mesma, até esta possuir uma duração igual ou superir a 30 segundos.
- Foi realizada a equalização da frequência do sinal de cada uma das amostras para 22050 Hz.

Após ter compilado a base de dado, ela foi distribuída em folds mantendo o melhor balanceamento possível entre as classes (espécies). Como durante o processo de criação da base de dado se fez o uso de um parâmetro de corte para espécies com menos de 10 amostras, garante-se a presença de pelo menos uma amostra de cada espécie em cada fold. Alguns detalhes acerca fa base de dados utilizada neste trabalho estão descritos na Tabela A, apresentada no Apêndice A deste trabalho.

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

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

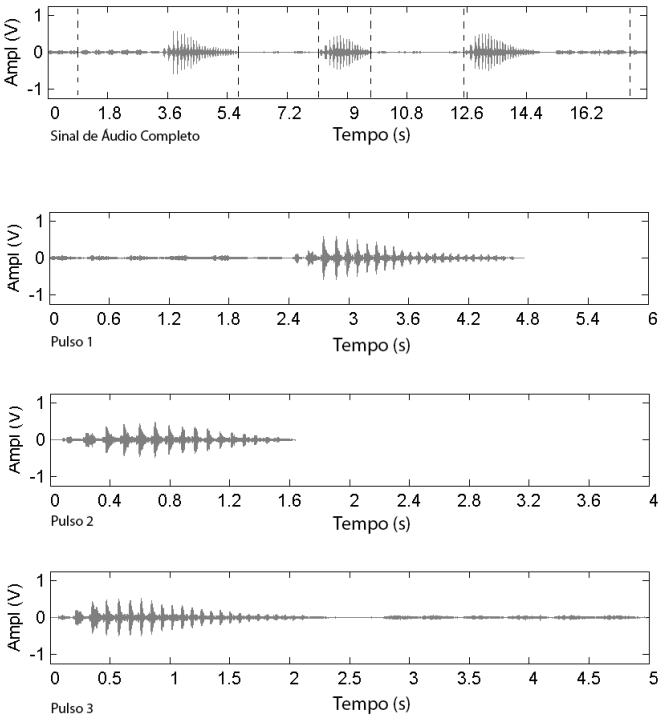


Figura 5: Representação da Divisão do sinal de áudio em pulsos adaptada de adaptada de Lopes et al. [36]

V. Método Proposto

Esta seção apresenta o método de classificação utilizado no desenvolvimento do trabalho até o presente momento. No que diz respeito especificamente ao método proposto neste trabalho, foram realizadas as seguintes etapas para realizar a tarefa de classificação: divisão da base de dados em folds, geração dos espectrogramas, extração das características, treinamento utilizando SVM para a criação de modelos de classificação com otimização de parâmetros do classificador. Cada uma das etapas apresentadas anteriormente podem ser vistas na seções seguintes.

A. Divisão da Base de Dados em Folds

Após ter sintetizado a base de dados, foi definida a forma como esta seria dividia em folds para poder realizar a tarefa de classificação. Durante o processo de criação da base de dados se fez o uso de um parâmetro de corte para espécies com menos de 10 amostras, para que fosse possível dividir a base em 10 folds. Desta forma, garante-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 distribuídos entre os folds, este processo foi realizado para cada uma das espécies presente nas versões da base de dados.

B. Geração do Espectrograma

Para a geração dos espectrogramas foi utilizado o sofware Sox 14.4.1 (*Sound eXchange*), um utilitário disponível em http://sox.sourceforge.net que permite a realização de conversões entre vários formatos diferentes de representação de áudio. Este permite a utilização de alguns parâmetros que irão impactar na aparência do espectrograma gerado, por meio deste pode-se delimitar a altura e a largura, e a amplitude do sinal do sinal de áudio a ser considerada. Alguns parâmetros deste software foram empiricamente ajustados a fim de produzir espectrogramas com conteúdo de textura destacado.

C. Extração de Características

O processo de obtenção dos vetores de características é similar aos encontrados em [37]–[40] e serão empregadas estratégias para a representação da textura da imagem do espectrograma detalhadas na seção III-A.

Os vetores descritores de textura de LBP foram extraídos considerando-se 8 vizinhos à uma distância de 2 *pixels* do *pixel* central. Com isso o veor utilizado tem um total de 59 características.

Com LPQ foi utilizada uma janela de tamanho 7, e o vetor utilizado tem 256 características. Na extração de características com Filtros de Gabor foram utilizadas 6 rotações e 5 fatores de escala. Destas, foram extraídas energia quadrada e amplitude da média, que resultou em um total de 60 características.

Vale ressaltar que tais configurações foram utilizadas decorrente do fato destas apresentarem os melhores resultados no trabalho apresentado por Costa [41].

D. Sistema de Classificação

Para realizar as tarefas de classificação será utilizado o SVM, um modelo de algoritmos de aprendizagem apresentado por em [42], seu uso é muito difundido em toda comunidade científica em trabalhos que envolvem a análise de dados e reconhecimento de padrões, sendo que este tem apresentado bom desempenho em vários trabalhos publicados recentemente. Para a construção dos modelos de classificação SVM, foi utilizado o *kernel Radial Basis Function* (RBF) e os parâmetros C e γ foram otimizados utilizando um procedimento *grid-search*.

O esquema de classificação proposto consiste na seguinte sequência de passos: divisão da base de dados em folds, geração dos espectrogramas, extração das características. Após extraídas as características foi realizada a classificação por meio do SVM através da biblioteca LIBSVM [43]. 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, sendo que o processo é repetido até que todos os folds criados tenham sido utilizados como conjunto de teste [44].

Ao final, toma-se como medida de desempenho a taxa de acerto média obtida entre todas as situações testadas.

VI. Resultados e Discussão

Conforme já descrito, foram utilizados os descritores de textura LBP, LPQ e Filtros de Gabor na extração de características dos espectrogramas, a configuração utilizada para tais descritores foi apresentada na seção V-C. Alguns parâmetros para a geração de espectrogramas foram testados com diferentes valores, sendo estes: o intervalo de tempo, o limite de frequência adotado e a amplitude do sinal. Os resultados para os descritores de textura podem ser vistos na tabelas II, III e IV.

A tabela II apresenta as taxas de acerto obtidos com o uso do LBP como descritor de textura. Pode-se constatar que ao se utilizar um trecho de áudio de 15 segundos temos uma progressão nas taxas de acerto, com o aumento do limite de frequência utilizada na geração dos espectrogramas. Também foi possível constatar o mesmo fato quando foram utilizados segmentos de áudio de 30 segundos, assim como também foi possível verificar-se uma inflexão na curva dos resultados quando ultrapassamos os 80 dB visto que os resultados tendem a diminuir.

Tabela II: Resultados obtidos com o uso do LBP

Tempo	Frequência	Amplitude do Sinal	Acerto	F-measuare
00:15	14000	80	70,19%	67,93%
00:15	16000	80	71,78%	69,35%
00:15	18000	80	73,45%	71,89%
00:15	18000	90	73,63%	72,21%
00:15	20000	80	75,66%	74,34%
00:15	22000	80	76,30%	74,89%
00:30	22000	70	73,67%	72,13%
00:30	22000	80	**77,33%**	**76,39%**
00:30	22000	90	75,02%	73,48%

A tabela III apresenta as taxas de acerto obtidos com o uso do LPQ como descritor de textura. Pode-se constatar que ao se utilizar um trecho de áudio de 15 segundos não temos taxas de acerto uma uma progressão linear assim como quando se fez o uso do LBP como descritor de textura. Todavia a inflexão da curva dos resultados também se mostrou presente quando utilizou-se uma amplitude de sinal de 90 dB assim como quando se empregou o LBP.

Tabela III: Resultados obtidos com o uso do LPQ

Tempo	Frequência	Amplitude do Sinal	Acerto	F-measuare
00:15	14000	80	67,77%	64,47%
00:15	16000	80	30,56%	24,53%
00:15	18000	80	31,06%	25,87%
00:15	18000	90	65,96%	63,14%
00:15	20000	80	31,73%	26,93%
00:15	22000	80	71,68%	68,52%
00:30	22000	70	33,12%	28,12%
00:30	22000	80	**71,93%**	**69,09%**
00:30	22000	90	69,15%	66,59%

A tabela IV apresenta as taxas de acerto obtidos com o uso de Filtros de Gabor como descritor de textura. Pode-se constatar que ao se utilizar um trecho de áudio de 15 segundos há uma progressão linear dos resultados quando foi utilizada uma mesma amplitude so sinal na geração dos espectrogramas. E diferentemente do que ocorreu quando utilizou-se os descritores de textura LBP e LBP não houve uma inflexão da curva dos resultados quando foi utilizada a aplitude do sinal de 90 dB, e devido a isto se fez um teste adicional utilizando-se uma amplitude de sinal de 100 dB, e a partir dos resultados deste teste adicional pode-se ver que a curva de inflexão quando se utiliza Filtros de Gabor é diferente da apresentada nos testes em que foram utilizados o LBP e o LPQ.

Tabela IV: Resultados obtidos com o uso de Filtros de Gabor

Tempo	Frequência	Amplitude do Sinal	Acerto	F-measuare
00:15	14000	80	74,84%	71,49%
00:15	16000	80	75,27%	71,89%
00:15	18000	80	75,59%	72,64%
00:15	18000	90	77,47%	75,28%
00:15	20000	80	75,98%	72,98%
00:15	22000	80	76,55%	74,45%
00:30	22000	70	76,01%	71,72%
00:30	22000	80	76,83%	74,49%
00:30	22000	90	**77,65%**	**75,67%**
00:30	22000	90	77,51%	74,31%

Como pode ser visto pela variação dos parametros citados anteriormente, foi constatado que os melhores resultados são obtidos com um espectrograma gerado utilizando o intervalo de tempo de 30 segundos, com limite de frequência de 22000 Hz e amplitude de sinal de sinal de 80 dB para os descritores de textura LBP e LQP. Com o uso de Filtros de Gabor o melhor resultado foi apresentado utilizando o intervalo de tempo de 30 segundos, com limite de frequência de 22000 Hz e amplitude de sinal de sinal de 90 dB. Esses valores foram escolhidos empiricamente pelo fato de permitir a produção de imagens de espectrograma com conteúdo de textura bem caracterizado, o que favorece a representação do mesmo com os descritores aqui empregados. Os resultados dos experimentos realizados até o momento podem ser vistos nas Tabelas II, III e IV.

Como análise final dos resultados pode-se ver que o melhor resultado foi obtido utilizando-se o Filtro de Gabor com uma taxa de acerto de 77,65% , utilizando uma amostra de aúdio de 30 segundos. Para se obter tal resultado foram utilizados espectrogramas gerados com um limite de frequência de 22000 Hz e uma amplitude de sinal de 90 dB. Todavia não foi possível obter exatamente as bases, com as mesmas classes e amostras, utilizadas em outros trabalhos, o que permitiria uma comparação mais justa dos resultados. Entretanto, as amostras utilizadas foram tiradas do mesmo repositório utilizado em outros trabalhos. Nesses casos, obteve-se desempenho similar aos apresentados no presente trabalhos, mesmo utilizando um conjunto de classes maior, fato este que demonstra que o método adotado tem potencial para ser utilizado em classificação automática de espécies de pássaros.

VII. Considerações Finais e Trabalhos Futuros

Com a realização dos experimentos, se pode verificar que os resultados apresentados, se equiparam com os resultados da apresentados na literatura fazendo o uso de outros sistemas de classificação, como pode ser visto na seção VI. Todavia ainda há alguns aspectos a serem analisados, em trabalhos futuros, como utilização do zoneamento dos espectrogramas, técnicas para redução do ruído tanto para imagem quanto para som, assim como também a utilização de características acústicas das características aqui investigadas, de modo a se obter taxas de acerto ainda melhores. Pretende-se também utilizar a base empregada neste trabalho com os descritores utilizados em outros trabalhos para uma comparação mais justa do desempenho do seu método aqui empregado.

Adicionalmente, aventa-se a possibilidade de se investigar uma metodologia de classificação baseada em *Convolutional Neural-Network* (CNN), na medida em que se defina novas bases de dados com número de espécies bastante superior ao aqui utilizado. Ainda neste contexto, há a possibilidade de se investigar o potencial do uso de dissimilaridade em busca de boas medidas de desempenho.

Apêndice A

Tabela V: Relação de espécies e amostras da base de dados

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
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

Continua na próxima coluna

Continuação da coluna anterior

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

AGRADECIMENTOS

Os autores agradecem ao Departamento de Informática da UEM, pela infraestrutura disponibilizada e a CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, por fornecer o apoio financeiro necessário para o desenvolvimento do presente trabalho.

REFERÊNCIAS

[1] A. Marini, J. Facon, and A. L. Koerich, "Bird species classification based on color features." IEEE, Oct. 2013, pp. 4336–4341.

[2] C. Pang, H. Yao, and X. Sun, "Discriminative features for bird species classification." ACM Press, 2014, pp. 256–260.

[3] C. Huang, F. Meng, W. Luo, and S. Zhu, "Bird breed classification and annotation using saliency based graphical model," *Journal of Visual Communication and Image Representation*, vol. 25, no. 6, pp. 1299–1307, Aug. 2014.

[4] V. I. Cullinan, S. Matzner, and C. A. Duberstein, "Classification of birds and bats using flight tracks," *Ecological Informatics*, vol. 27, pp. 55–63, May 2015.

[5] S. Fagerlund, "Bird species recognition using support vector machines," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[6] S. E. Anderson, A. S. Dave, and D. Margoliash, "Template-based automatic recognition of birdsong syllables from continuous recordings," *The Journal of the Acoustical Society of America*, vol. 100, no. 2 Pt 1, pp. 1209–1219, Aug. 1996.

[7] S. Selouani, M. Kardouchi, E. Hervet, and D. Roy, "Automatic birdsong recognition based on autoregressive time-delay neural networks," in *2005 ICSC Congress on Computational Intelligence Methods and Applications*, 2005, pp. 6 pp.–.

[8] H. Goëau, H. Glotin, W.-P. Vellinga, R. Planqué, A. Rauber, and A. Joly, "Lifeclef bird identification task 2014," in *CLEF2014*, 2014.

[9] R. T. Holmes, *Ecological and evolutionary impact of bird predation on forest insects: an overview*, 1990, pp. 6–13.

[10] R. T. Holmes, J. C. Schultz, and P. J. Nothnagle, "Bird predation on forest insects: an exclosure experiment," vol. 206, pp. 462–463, 1979.

[11] D. W. Snow, "Evolutionary aspects of fruit-eating by birds," *Ibis*, vol. 113, no. 2, pp. 194–202, Apr. 1971.

[12] F. L. Carpenter, "A spectrum of nectar-eater communities," *American Zoologist*, vol. 18, no. 4, pp. 809–819, 1978.

[13] P. Feinsinger and R. K. Colwell, "Community organization among neotropical nectar-feeding birds," *American Zoologist*, vol. 18, no. 4, pp. 779–795, 1978.

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

[15] F. Straube, "Newsletter bocev (bird observers club of the european valley)," *Oct*, 2005.

[16] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1524 – 1534, 2010, pattern Recognition of Non-Speech Audio.

[17] F. C. Straube and G. V. Bianconi, "Sobre a grandeza e a unidade utilizada para estimar esforço de captura com utilização de redes-de-neblina," *Chiroptera Neotropical*, vol. 8, no. 1-2, pp. 150–152, 2014.

[18] M. Lopes, L. Gioppo, T. Higushi, C. Kaestner, C. Silla, and A. Koerich, "Automatic Bird Species Identification for Large Number of Species," in *2011 IEEE International Symposium on Multimedia (ISM)*, Dec. 2011, pp. 117–122.

[19] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K. H. Tauchert, and K. H. Frommolt, "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1524–1534, Sep. 2010.

[20] I. Agranat, "Automatically identifying animal species from their vocalizations," in *Fifth International Conference on Bio-Acoustics, Holywell Park*, 2009.

[21] J. A. Kogan and D. Margoliash, "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*, vol. 103, no. 4, pp. 2185–2196, 1998.

[22] C.-H. Chou, C.-H. Lee, and H.-W. Ni, "Bird species recognition by comparing the HMMs of the syllables," in *Second International Conference on Innovative Computing, Information and Control, 2007. ICICIC '07*, Sep. 2007, pp. 143–143.

[23] F. Briggs, R. Raich, and X. Z. Fern, "Audio classification of bird species: a statistical manifold approach," 2009.

[24] T. L. Marcelo and C. A. A. Kaestner, "Identificação automática de pássaros através de cantos e chamados," in *SICITE'2010*, 2010.

[25] M. T. Lopes and C. A. A. Kaestner, "IDENTIFICAcCão automática de pássaros através de cantos e chamados," 2011.

[26] M. T. Lopes, A. Lameiras Koerich, C. Nascimento Silla, and C. Alves Kaestner, "Feature set comparison for automatic bird species identification," in *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2011, pp. 965–970.

[27] R. Semolini, "Support vector machines, inferência transdutiva e o problema de classificação," Ph.D. dissertation, Universidade Estadual de Campinas, 2002.

[28] R. O. Duda, P. E. Hart *et al.*, *Pattern classification and scene analysis*. Wiley New York, 1973, vol. 3.

[29] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[30] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[31] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[32] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and signal processing*. Springer, 2008, pp. 236–243.

[33] W. Li, K. Mao, H. Zhang, and T. Chai, "Selection of gabor filters for improved texture feature extraction," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 361–364.

[34] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.

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

[36] M. Lopes, A. Lameiras Koerich, C. Nascimento Silla, and C. Alves Kaestner, "Feature set comparison for automatic bird species identification," in *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2011, pp. 965–970.

[37] J. G. Martins, Y. Costa, D. Bertolini, and L. Oliveira, "Uso de descritores de textura extraıdos de glcm para o reconhecimento de padroes em diferentes domınios de aplicaçao," *XXXVII Conferencia Latinoamericana de Informática*, pp. 637–652, 2011.

[38] Y. M. G. Costa, L. S. Oliveira, A. L. Koericb, and F. Gouyon, "Music genre recognition using spectrograms," in *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*, June 2011, pp. 1–4.

[39] Y. Costa, L. Oliveira, A. Koerich, and F. Gouyon, "Music genre recognition using gabor filters and LPQ texture descriptors," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, J. Ruiz-Shulcloper and G. Sanniti di Baja, Eds. Springer Berlin Heidelberg, 2013, vol. 8259, pp. 67–74.

[40] Y. M. G. Costa, O. L. S., K. A. L., G. F., and M. J. G., "Music genre classification using LBP textural features," *Signal Processing*, vol. 92, no. 11, pp. 2723 – 2737, 2012.

[41] Y. M. G. Costa, "Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores," Ph.D. dissertation, Universidade Federal do Paraná, 2013.

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

[43] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[44] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, Mar 1998.