Bird Species Classification Based on Color Features

Andréia Marini*, Jacques Facon* and Alessandro L. Koerich*†

*Postgraduate Program in Computer Science (PPGIa)
Pontifical Catholic University of Paraná (PUCPR)
Curitiba PR, Brazil 80215–901
Email: {marini,facon,alekoe}@ppgia.pucpr.br

†Department of Electrical Engineering
Federal University of Paraná (UFPR)
Curitiba PR, Brazil, 81531-990
Email: alessandro.koerich@ufpr.br

Abstract—This paper presents a novel approach for bird species classification based on color features extracted from unconstrained images. This means that the birds may appear in different scenarios as well may present different poses, sizes and angles of view. Besides, the images present strong variations in illuminations and parts of the birds may be occluded by other elements of the scenario. The proposed approach first applies a color segmentation algorithm in an attempt to eliminate background elements and to delimit candidate regions where the bird may be present within the image. Next, the image is split into component planes and from each plane, normalized color histograms are computed from these candidate regions. After aggregation processing is employed to reduce the number of the intervals of the histograms to a fixed number of bins. The histogram bins are used as feature vectors to by a learning algorithm to try to distinguish between the different numbers of bird species. Experimental results on the CUB-200 dataset show that the segmentation algorithm achieves 75% of correct segmentation rate. Furthermore, the bird species classification rate varies between 90% and 8%, depending on the number of classes taken into account.

Index Terms—pattern recognition; color features; color image segmentation; machine learning; bird species classification.

I. INTRODUCTION

Bird species identification from images is an important and challenging problem with many applications in the real world such as environment protection and endangered animal rescue [1]. There are also some other practical reasons to monitor birds. In order to evaluate the quality of our living environment it is important to obtain reliable information about the population of wild animals. Birds are numerous and sensitive to environmental changes; also, and are easier to monitor than other species. Therefore, the use of automated methods for bird species identification is an effective way to evaluate the quantity and diversity of the birds which appear in a region [2], [3]. The practical reasons previously mentioned justify the study of mechanisms for bird species identification.

Bird identification is a well-known problem to ornithologists, and is considered as a scientific task since antiquity. Ornithologists study birds; their existence in nature, their biology, their songs, their distribution, and their ecological impact. Bird classification is usually done by ornithology experts based

http://www.birding.com

1. Birds
The main objective of this paper is to evaluate simple feature descriptors in bird images, and what is the expected classification performance that can be achieved when dealing with a great number of bird species. Given the complexity of the problem, a scenario which is unconstrained, a high number of classes, a high visual similarity between some bird species, there is a need of novel methods to deal with each step of the problem and to provide results that are more reliable than those currently achieved. This paper presents a first approach to bird species classification which is based only on color features. The choice of employing only colors is that it is simple and intuitive and recurrent to the birds of the same species. Furthermore, color is also perceived by the human visual system, and is one of the main features used by ornithologists to discriminate between bird species. On the other hand, it is possible to represent the pixel color in different color spaces such as RGB and HSV to enhance some specific characteristics in the images.

This paper is organized as follows. Section II presents some visual clues that are frequently used by ornithologists to identify bird species. Section III describes the proposed method for bird species classification. Section IV presents the evaluation of the proposed method on the CUB-200 dataset. Finally, conclusions and the perspective of future work are stated in the last section.

II. VISUAL IDENTIFICATION OF BIRD SPECIES

Basic bird identification clues are acoustic and visual. The main visual clues comprise the bird’s silhouette, its plumage and coloration. However, we must take into account the time of year because bird’s plumage will change during different seasons. The acoustics clues comprise the songs and calls that birds make. Furthermore, the bird’s behavior and habitat it is found in are also possible clues. Field marks, the marks that distinguishing one bird from another are also important, such as breast spots, wing bars which are described as thin lines along the wings, eye rings, or crowns, eyebrows which are described as lines over the eyes, eye lines, which are lines through the eyes, and many others. The shape of the beak is often an important clue. Size can usually be a good indicator of the species of the bird. As an example, most songbirds fit into a certain size group. Bird shape and posture are the most important characteristics used to identify birds. Most experts can identify a bird from its shape or silhouette because this is the least likely characteristic to change. The tail of a bird can have many variations. The tail can be notched, long and pointed, or rounded. The legs can be long, or short.

In spite of having many different features that can take into account to distinguish between bird species, from the pattern recognition point of view, relying on many of such features to identify different bird species is not trivial mainly due to the difficulty in obtaining standard bird images. The acquisition of bird images is somewhat out of the control of the researchers and this imposes an extraordinary variability which is very difficult to be handled by most of the image processing, feature extraction methods and learning algorithms.

III. PROPOSED APPROACH

Many of the characteristics suggested in Section II are not practical to be extracted from real-life bird images since they require that images follow some standard, which is difficult to obtain in practice. The size of a bird in an image depends on the resolution, distance between the birds and the acquisition equipment, and the focal distance of the lens. Therefore, based on a practical observation of a large number of images, at a first sight, color seems to be the most discriminative feature that can be observed in many of real-life bird images. For this

http://www.all-birds.com
reason, the proposed method is based solely on color features as a means of building a baseline for bird species identification.

Figure 1 shows an overview of the proposed method. The environment where the birds are found when the image are acquired, acts as an unfavorable factor. Therefore, any attempt to eliminate the background before extracting features should be considered. Although, eliminating the background could be considered a problem as difficult as the species identification itself. The segmentation step is based on the assumption that all available images are in colors, that the birds are at the central position in the images, and that the bird edges are far away from the image borders. Therefore, there are some strips at the image borders that can be considered as belonging to the image background. The size of these strips is chosen to be a percentage, usually between 2% and 10% of the image horizontal and vertical dimensions. First, these strips are scanned and the colors that are found into them are stored in a ranked list according to the color frequency. Next, a search procedure is carried out on the remaining of the image and the pixels that have similar colors to those found in the strips are labeled as background; otherwise they are labeled as "bird". At the end we have all pixels in the image labeled either as background or bird. Figure 2 presents some details of the proposed segmentation approach. The proposed segmentation step has some similarity with the work of Das and Manmatha [17] which have employed a similar approach to segment the region of interest from the background. Further, the images are indexed in domain specific databases using colors computed from the object of interest only, instead of the whole image.

Color is an important feature in the image classification process. Several works are based on color features or use them together other feature types [13], [14], [16], [18]–[20]. To obtain relevant information about the relationship of the color in the image, both the red, green and blue (RGB) and the hue, saturation and value (HSV) spaces are considered. RGB color space rearranges the geometry of RGB so that it could be more relevant to human perception, because it is more natural to think about a color in terms of hue and saturation than in terms of additive color components. Regardless which the color space is employed, three-color histograms are computed, one for each channel. The histograms are converted into feature vectors by a binning process where each channel is represented by a fixed number of bins.

The feature vectors are labeled with the corresponding bird species and used in a supervised learning process. Support vector machine (SVM) was chosen as the supervised learning model and the Platt’s sequential minimal optimization algorithm as the learning algorithm. The choice of the SVM is due to the fact that it can efficiently perform non-linear classification using the kernel trick which implicitly maps their inputs into high-dimensional feature spaces. The one-against-one approach is used to handle the bird species classification since it is a multiclass classification problem. Therefore, $k(k - 1)/2$ classifiers are constructed and each one trains data from two classes, where $k$ is the number of classes. In classification, the voting strategy, where each binary classification is considered to be a voting where votes can be cast for all data points and at the end a point is designated to be in a class with the maximum number of votes. The SVMs
classifiers where implemented with a Radial Basis Function kernel and the gamma and cost parameters were optimized through a grid search.

Two different approaches were employed to handle the features vectors since at the end of the feature extraction three feature vectors for each image is available. In the first approach the three feature vectors are concatenated and handled by a single classification algorithm as shown in Figure 3. In the second approach, each feature vector is handled by a different classifier and at the end, the outputs of the classifiers are combined to reach a decision on the bird species (Figure 4).

Figure 3. Classification scheme based on the concatenation of the color planes features

Figure 4. Classification scheme based on combination of classifiers

IV. EXPERIMENTAL RESULTS

This section presents the evaluation of the proposed approach for bird species classification based on color features on the Caltech-UCSD Birds 200 (CUB-200) dataset [13]. This is a challenging image dataset annotated with 200 bird species. CUB-200 includes 6,033 annotated images of birds, belonging to 200, mostly North American, bird species. Each image is annotated with a rough segmentation, a bounding box, and binary attribute annotations. There are between 20 and 40 samples for each bird species. The images in this dataset were obtained directly from real environments, without any filtering or preprocessing, and thus contain environmental elements at the background as well as occluding the birds, such as leaves and branches. Furthermore, there is no normalization in the image acquisition process. This means that the birds may occupy a small portion of the image, or even the whole image. All experiments described in this section were carried out considering a 5-fold cross-validation procedure, that is, the results are obtained from five randomly independent experiment repetitions, unless otherwise noted.

How effective is the proposed approach in segmenting the bird images? The segmentation procedure described in Section III was applied to the bird images. Since the images in the CUB-200 dataset may have different dimensions, the width of the strip chosen to represent the background colors depends on them. Several experiments were carried with stripes width of 2% to 10% of the image dimensions. The CUB-200 dataset includes a rough segmentation of the birds in each image, so we can compare the image segmented by the proposed algorithm with such a rough segmentation to evaluate the segmentation. For such an aim, we count the number of true positives (TP) which is defined as the number of pixels labeled by the segmentation algorithm as belonging to the bird which are also labeled as belonging to the bird in the rough segmentation. Similarly, we compute the number of false positives (FP) which is defined as the number of pixels labeled by the segmentation algorithm as belonging to the bird but are labeled as belonging to the background in the rough segmentation, true negatives (TN), which is defined as the number of pixels labeled by the segmentation algorithm as belonging to the background which are also labeled as belonging to the background in the rough segmentation, and the false negatives (FN), which is defined as the number of pixels labeled by the segmentation algorithm as belonging to the background but are labeled as belonging to the bird in the rough segmentation. These four measures are used to compute the segmentation rate by Equation 1.

\[
\text{Segmentation Rate} = \frac{TP + TN}{TP + FP + TN + FN}
\] (1)

Table I shows the segmentation results. The segmentation results compare favorably to those presented by Das and Manmatha [17] which achieved 87% of segmentation on a dataset of 450 images.

<table>
<thead>
<tr>
<th>Color Space</th>
<th>Segmentation Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>71.0</td>
</tr>
<tr>
<td>HSV</td>
<td>75.0</td>
</tr>
</tbody>
</table>

One of the goals of the experiments is to evaluate the impact of the segmentation on the classification as well as the segmentation method itself. For such an aim, first the experiments that takes into account the whole image. This means that the color features were extracted from the whole image which includes the bird and the background. Table II shows the results for the HSV and RGB color spaces, with and without segmentation. The results are also show in terms of the number of classes. Table II shows that the correct classification
rate is relatively high for the experiments dealing with few classes, but it falls off for two hundred classes. This is a clear indication that the color features are not discriminative enough to deal with a high number of classes. Regarding the segmentation, it is also clear that it has a favorable impact on the classification rates, providing an increasing of 8.82% in the classification rate for two classes, but such an impact is not as meaningful as the number of classes increases, achieving 0.43% for two hundred classes. Furthermore, when comparing the results achieved by the color spaces HSV and RGB, the best results were always achieved using HSV color space. The difference in classification rate ranges from 10.29% for two classes to 1.74% for two hundred classes. Notice that all results reported in Table II are for the first classification scheme (Fig. 4), where a single classifier is used.

Table III shows the results of the second classification approach, where the features extracted from each image plane are used to train different classifiers, one for each image plane. In Table III, $P_H$, $P_S$, and $P_V$ denote the vectors generated from the color histograms of channel $H$, $S$, and $V$ respectively. Only the results for the HSV space are shown in this table since such a space always provided the best correct classification rates. The results presented in Table III are very interesting because they use less information to classify the bird species than the previous approach that employs a feature vector which concatenates the three image channels. However, the results in Table III do not overcome any result shown in Table II. The second classification approach implies the combination of the output of the individual classifiers which provide at the output values between 0 and 1. These values can be considered as estimations of the \textit{a posteriori} probabilities, then they can be combined using the max, product, sum, weighted sum and weighted product rules. Table IV shows the results achieved by combining the individual classifiers by several rules. If we compare the result shown in Table IV with those shown in Table II, the second classification approach only overcomes the first classification approach when five classes are taken into account. For all other cases, the first classification approach achieves the best results. This indicates that the feature vector splitting and the combination of classifier is not suitable to handle this problem.

V. CONCLUSION

This paper deals with the automatic bird species identification from bird images. We present a series of experiments conducted in a dataset composed by more than 6,000 images from 200 different bird species. The experimental scenarios employ two different classification approaches where in the first one, the feature vectors generated from the split image planes are concatenated and feed into the a single classifiers. In the second approach each feature vector is treated separately by a different classifier. Furthermore, we have also considered two different color spaces, RGB and HSV, and a different number of species to be classified to evaluate the scalability of the proposed approach.

Several experiments were carried out to evaluate the performance and the impact of the color in the image segmentation. It is very clear the impact of the segmentation on the classification results. But even if for both HSV and RGB color spaces, more than 70% of the pixels were correctly segmented, the impact on the bird species classification was not very impressive, ranging from 8.82% to 0.43%. This suggests that the segmentation does not play an important role in such a problem, in particular when the number of classes is high. Although, this conclusion is only valid for the color features that we have employed. We can not extend this conclusion to other types of features. Furthermore, we can conclude that the proposed approach does not present a good scalability since

### Table II

**Correct classification rates with and without segmentation for 2, 5, 17, and classes for the HSV and RGB color spaces**

<table>
<thead>
<tr>
<th>Color Space</th>
<th>Without Segmentation</th>
<th>With Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV</td>
<td>83.82 47.02 25.05 8.17</td>
<td>92.64 48.34 25.63 8.60</td>
</tr>
<tr>
<td>RGB</td>
<td>73.53 39.07 16.96 4.16</td>
<td>77.94 40.39 18.49 6.86</td>
</tr>
</tbody>
</table>

### Table III

**Correct classification rates for the HSV color space**

<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>2</th>
<th>5</th>
<th>17</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_H$</td>
<td>82.35</td>
<td>45.03</td>
<td>20.42</td>
<td>6.60</td>
</tr>
<tr>
<td>$P_S$</td>
<td>77.94</td>
<td>42.38</td>
<td>15.22</td>
<td>4.34</td>
</tr>
<tr>
<td>$P_V$</td>
<td>79.41</td>
<td>42.38</td>
<td>10.60</td>
<td>2.87</td>
</tr>
</tbody>
</table>

### Table IV

**Correct classification rates for the fusion of the classifiers output, with segmentation for a different number of classes for the HSV color space**

<table>
<thead>
<tr>
<th>Fusion Rule</th>
<th>2</th>
<th>5</th>
<th>17</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX</td>
<td>85.29</td>
<td>47.68</td>
<td>19.65</td>
<td>6.76</td>
</tr>
<tr>
<td>SUM</td>
<td>86.76</td>
<td>49.01</td>
<td>22.16</td>
<td>7.16</td>
</tr>
<tr>
<td>PROD</td>
<td>88.24</td>
<td>49.67</td>
<td>22.54</td>
<td>7.25</td>
</tr>
<tr>
<td>WSUM</td>
<td>89.71</td>
<td>51.66</td>
<td>23.89</td>
<td>7.59</td>
</tr>
<tr>
<td>WPROD</td>
<td>91.18</td>
<td>51.66</td>
<td>23.70</td>
<td>8.03</td>
</tr>
</tbody>
</table>
we expected much better results for a high number the classes.

Comparing different works in the literature is not a straightforward task because of different experimental protocols. Table V summarizes some recent works on bird species classification that have used the CUB-200 dataset. Based on the results presented in this study and the performance of the related works, we can assert that color features are interesting alternatives for bird species identification problem, however, the best results reported have been achieved with SIFT features. Notice that the experimental protocol adopted by Chai et al. [14], Branson et al. [21], and Moghimi [20] is slightly different since they use a training (20 images per class split) and a testing set.

TABLE V

<table>
<thead>
<tr>
<th>Reference</th>
<th>Features</th>
<th>Method</th>
<th>Correct Classification Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welinder et al. [13]</td>
<td>SIFT + Spatial Pyramids</td>
<td>SVM</td>
<td>19</td>
</tr>
<tr>
<td>Chai et al. [14]</td>
<td>SIFT + BiCoS-MT Segmentation</td>
<td>SVM</td>
<td>16.2</td>
</tr>
<tr>
<td>Moghimi [20]</td>
<td>Color + Segmentation</td>
<td>kNN</td>
<td>18.9</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENTS

This research is partially supported under CNPq grants 301.653/2011-9, 306.703/2010-6 and 472.238/2011-6, CAPES/Fulbright grant BEX1770/712-9, and FA grant 203/12.

REFERENCES