CONSTRUCTION OF A BIRD IMAGE DATASET FOR ECOLOGICAL INVESTIGATIONS

Ryota Yoshihashi, Rei Kawakami, Makoto Iida and Takeshi Naemura

The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan

ABSTRACT

Wind turbines have become significant collision risks to endangered birds. To estimate the impact of bird strikes and take action against them, we need an automated bird monitoring system that detects and categorizes bird species. Towards this goal, we constructed an image dataset as a benchmark for ecological investigations of birds. The dataset consists of time-series images of a scene in a wind farm, bounding boxes around the aerial objects in the images, and species annotations. Bird experts searched for and annotated the images of birds, and thus, even birds that appeared to be very small in the whole image could be specified in detail. In total, 32,000 birds are annotated. In collecting the images, we had to monitor birds in the distance to cover enough area, but even with a telephoto setup, the average size of the birds in the image is around 25 pixels. This means we need a means of recognition that works on very low-resolution images. As an application of the dataset, we conducted experiments in two-class categorization (birds and non-birds), to serve as basis for detection and further categorization. We found that the dataset is a challenging recognition task because of its low resolution and hard negative samples.

***Index Terms*—** Image recognition, object detection, image dataset, ecological conservation

1. INTRODUCTION

Wind farms are becoming more widespread as a result of the recent rise in demand for clean energy. However, birds often strike the blades of wind turbines, which is of particular concern when it comes to endangered species. To avoid such strikes, wind farms can be built in regions inhabited by fewer birds, or their turbines can be made to decelerate or stop when birds approach. Moreover, to estimate the risk of a bird strike, we need to detect birds around the wind farm and identify their species, numbers, and flying routes. However, as manual investigation is expensive and potentially inaccurate, automatic detection and categorization of birds would be preferable.

The development of an automated bird monitoring system poses novel problems in the field of image processing and computer vision, i.e., detection and fine-grained categorization of small objects in a large image. This is due to two reasons: First, the system needs to have a wide monitoring area, since it has to notice the approach of birds well before they collide with the wind turbine, and the observation area has to be big enough to ascertain the distribution of birds. Second, it needs to categorize bird species in a fine-grained way, since the ability to discriminate between an endangered species and visually similar non-endangered ones is important to assess the severity of their situation.

Towards the development of an automated system, we have constructed a wild-bird dataset and evaluated it using basic object detection methods. The dataset is designed to fulfill the requirements of automatic bird monitoring, namely, precise and fine-grained monitoring of a wide area. The dataset consists of sequences of images around a wind farm, in which each bird is annotated by experts with a bounding box and a tree-structured label of species, e.g., bird - hawk - black kite. Over 32,000 birds and 4,900 non-birds are annotated. Thus, it offers a means of training and evaluating image recognition algorithms that was derived from very natural circumstances.

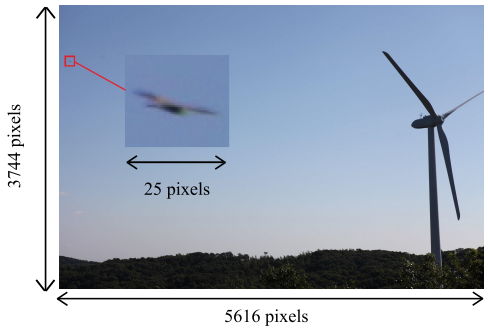

Fig. 1. A typical scene captured with our telephoto setup and stored in the database. Although the resolution of the images is as large as 5616×3744 pixels, the birds look small.

Ours is the first practical dataset for ecological investigations which monitor wide area of natural environments, and we expect that it will be useful for detecting birds or as a benchmark for fine-grained object categorization. The average size of birds in the dataset is around 25 pixels, whereas the size of the scene is larger than a 4K image. This means we need a detection algorithm that can work with very low resolution images. Furthermore, the dataset includes insects, helicopters and airplanes, which are hard to differentiate from birds. It will be a challenge to develop an algorithm that can handle such a difficult categorization. To evaluate the dataset, we performed a two-class categorization (birds and non-birds), to serve as a basis for understanding the characteristics of the low-resolution detection problem. We believe our findings can be applied to other problems such as those in wide area surveillance. Our dataset is now available at http://bird.nae-lab.org/dataset.html.

Below, we briefly review some of the related work. It is well known that the resolution of an image affects recognition performance. In detection benchmarks on other objects, it has been shown that smaller objects are more difficult to detect [1]. Classification is also difficult with low-resolution images. Scene classification and generic object classification at low resolution have already been tested on famous benchmarks such as Tiny Images [2] and CIFAR-100 [3]. However, it is not clear how to perform fine-grained classification at low resolution in a task with subordinated classes such as bird species. Although studies on fine-grained classification of birds with rich image datasets and rich models are ongoing [4, 5], monitoring objects over a wide area will require different algorithms, as this paper will show. Thus, we believe that our dataset will be a useful benchmark for low-resolution detection and classification.

Despite the recent development of high-performance detection methods including efficient features and powerful classifiers, the number of studies focusing on wild bird detection is small. Some trials on bird detection have been based on anomaly detection by segmentation [6] or background subtraction [7]. The methods therein succeeded at detecting birds; however, they would misrecognize other objects such as wind turbines, insects, or trees appearing in the scenes and, hence, are impractical for natural environments. One of the few exceptions is the method of Qing et al. [8], wherein a boosted HOG-LBP was used on a dataset of 1,000 bird images. Moreover, the existing datasets that include bird images are insufficient for our purposes of detecting flying birds at a distance; for instance, ImageNet [9], Caltech-UCSD Birds 200 [4] and Birdsnap [5] mainly contain images taken by cameras set close to the birds and contain no time-series sequences. Thus, we believe our dataset would be useful for evaluating novel bird recognition methods.

2. CONSTRUCTION OF THE DATASET

2.1. Image Capture

We designed a camera setup for capturing birds in the distance. The setup is able to capture a bird with a one-meter wing span 580 meters away that would cover an area of 20 pixels in the image, by taking into account the distance between the cameras location and the wind turbine. Specifically, we used a digital still camera (Canon EOS Mark II 5D) controlled by a laptop and equipped with a telephoto lens (Canon EF70-200mm F4L USM). The resolution of the sensor is 5616×3744 pixels, the focal length of the lens is set to 70 mm, and the field of view is $27^\circ \times 19^\circ$.

The images were recorded near a wind turbine for three days. The capture system took a picture every two seconds for seven hours from 9:00 to 16:00. We obtained 10,814 images per day, 32,442 images in total. The frame rate was only 0.5 fps, because of the large data size and data transfer speed. Figure 1 shows examples of the captured images. An example of a bird's appearance is also shown. Image variances other than birds include movements of clouds, the spinning blades of the wind turbine, shaking of the bushes by the wind, and illumination changes. Such variances pose a challenge when we try to detect birds from image differences.

2.2. Labeling

Each bird in the dataset is enclosed by a bounding box labeled with its species by experts. The labeling format is similar to those of other detection datasets such as the Caltech Pedestrian Detection Benchmark [10], which includes bounding boxes on time-series images. In addition, ours has fine-grained category annotations on each bounding box. Negative samples of other flying objects such as planes and bugs are also labeled.

For annotating the kind of bird, we designed a tree-structured list of categories so that an expert can annotate the bounding box with labels consisting of the names listed in the tree. The names of the kinds of birds in the list were selected on the basis of the results of a preparatory field survey. The granularity of the label can be selected depending on how clear the image of the bird is. For example, when a black kite appears, we may categorize it, depending on clarity, as a black kite, a kind of hawk, as a bird, or as an unclear flying object. These options become the nodes of the tree, and the depth of the tree corresponds to the level of detail. We made the list updatable, so that when an expert finds a bird that is not listed, he or she can add it to the list.

Besides birds, other flying objects, such as airplanes, helicopters, insects, and fallen leaves are also recorded. In so doing, these objects can be distinguished from birds that might have been missed by the experts. Non-bird images can also be used as negative samples for machine learning. Objects that are too ambiguous for experts to distinguish are also recorded and labeled. Thus, the dataset contains three types of object: birds, non-birds, and unclear flying objects.

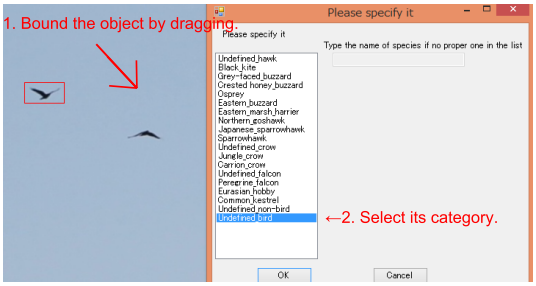

Fig. 2. The interface for labeling. Bird experts used this interface to check the images, find birds, create their bounding boxes, and annotate the boxes.

Manually labeling sequential images of the dataset faces several issues. First, manual labeling is time consuming, since the images number as many as 32,442. To efficiently process the data, we developed a user interface, a screenshot of which is shown in Figure 2. It enables us to check images sequentially and label a bird with two actions, i.e. by making a bounding box by dragging a mouse and selecting a category from the list. The user's actual procedure is as follows: a user goes through a sequence frame by frame and checks if there is any flying object. When a flying object is found, he or she inputs the bounding box and selects the category from the given list, or else types in the category if it is not listed. The procedure is iterated until the end of the sequence.

Second, it is often difficult for non-experts to confirm an image to be of a bird, because of their small size in the images. We thus requested dozens of members of a wild bird society to inspect the images and input the data. Their efforts ensured that the labeling is precise and fine-grained.

Third, due to the large size of the images (5616 by 3744 pixels), we cannot display its full size on an ordinary display. Therefore, we divided the original image into 30 (6 by 5) parts. One segment was assigned to a user, who then went through a one-day sequence of it (a total of 10,814 images). Birds that are on the boundaries of segments may be missed easily. To stop this from happening, we asked the users to check the images twice, and we divided the images differently in each instance. In the first check, the images were divided into 30 (6 by 5) segments. In the second check, we shifted the dividing lines by half a segment.

2.3. Dataset statistics and Image Examples

The dataset contains 32,973 bird images, each of which is specified by a bounding box annotated with its category whenever identified. It also contains 4,911 non-birds, and 1,907 unclear flying objects.

Figure 3 shows the categories and their proportions for each day. Hawks are the most frequent, with crows being second; 30% and 10% of the overall observations were of hawks and crows, respectively. The percentage of unspecified birds is about 40%. Other birds include swallows, sparrows, meadow buntings, and so forth. Table 1 lists the numbers of appearances of each species. Black kites, the most common Japanese hawk, make up most of the bird appearances. Species whose appearances number more than 10 are listed. These are candidates of classes for fine-grained categoriza-

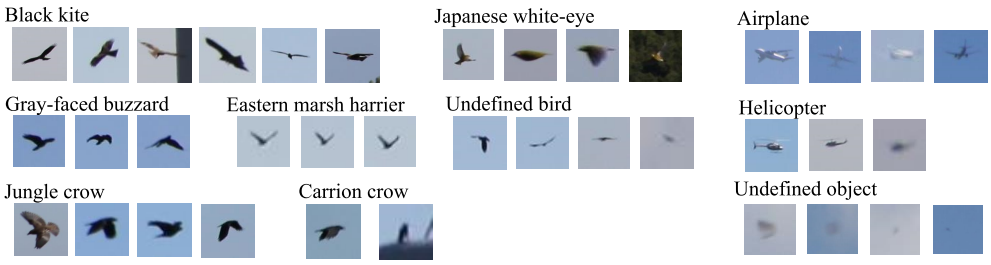

Fig. 4. Examples of found birds and other objects. The images of different sizes are resized to the same size.

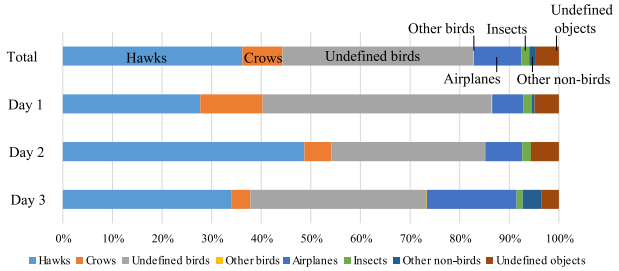

Fig. 3. Proportions of categories of found objects. Hawks were the most frequently observed, with crows being the second most frequent among the specified birds. Because there were no remarkable differences as to the proportion of birds observed during the three days, these figures can be regarded as typical for this location.

Table 1. Numbers of appearances of main species. Black kites, the most common kind of hawk in Japan, make up the major portion of the specified birds. The listed species are candidate classes for fine-grained categorization.

species	kind	number of appearance
Black kite	Hawk	2921
Jungle crow	Crow	53
Gray-faced buzzard	Hawk	36
Crested honey buzzard	Hawk	30
Osprey	Hawk	16
Japanese white-eye	Other bird	15

tion; categorization with fewer images is difficult.

Figure 4 shows examples of birds found by the users. Some images are relatively clearer, and thus, they can be specified in detail. Even some of the not-so-clear images are specified in detail. For example, the eastern marsh harriers in Figure 4 are not so clear. These birds however could be specified according to their actions. The three images are a sequence of a single individual, and it kept a V pose while flying during the sequence. This is a characteristic feature of eastern marsh harriers that made it possible to specify the species of the individual.

Figure 5 shows the size distribution of birds and non-bird objects in the images. The peak of the distribution is around 25 pixels. Distant birds are found less often because of their small size, while nearby birds do not come in the narrow angle of view often. The birds can not be distinguished from non-birds on the basis of their apparent size.

3. APPLICATION: BIRD RECOGNITION

We conducted experiments to clarify the characteristics of this dataset. We examined the nature of the dataset in relation to recognition precision; namely, the objects imaged at low resolution and the difficulty of distinguishing similar objects such as birds and

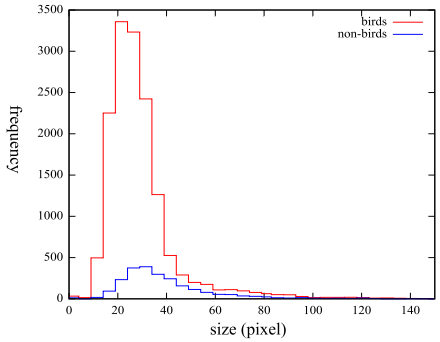

Fig. 5. Distribution of birds and non-birds in the dataset. The peak bird size is 20 pixels square and that of non-birds is 35 pixels. The birds and non-birds are not distinguishable by their apparent size.

airplanes or insects. At the same time, we compared the precision of recognition of the most popular detection methods that use Haar-like [11] and histogram of orientated gradients (HOG [12]) features. We conducted a two-class categorization into birds and non-birds by using Haar-like and HOG features in order to examine the features of the low-resolution bird recognition task.

3.1. Method of Bird Recognition

We used Haar-like [11] or HOG [12] features and the AdaBoost [13] learning algorithm. The procedure was as follows.

1. Regularize the images to the same size (24 $\times$ 24 pixels)
2. Extract Haar-like or HOG features from the regularized images
3. Train AdaBoost with the feature vectors of the training images.
4. Input the feature vectors of test images and classify birds and non-birds.

In this method, we can classify images of different sizes with one classifier by extracting the same dimensional feature vectors from the regularized images.

As black and white patterns for Haar-like features, we used the same ones that Viola and Jones [11] used. The sizes of these patterns were two, six, and ten pixels. The cell size of HOG was four pixels square, and the block size was 3 cells. The size of the images after regularization was 24 $\times$ 24 pixels, and the dimension of the feature vectors was 5567 for Haar-like and 1296 for HOG features.

3.2. Experiment

We generated bird images as positive samples and other images as negative samples for this recognition experiment from the constructed dataset. First, we clipped moving objects out of the images in the dataset by background subtraction. Clipped regions recorded in the dataset were divided into bird regions or non-bird regions

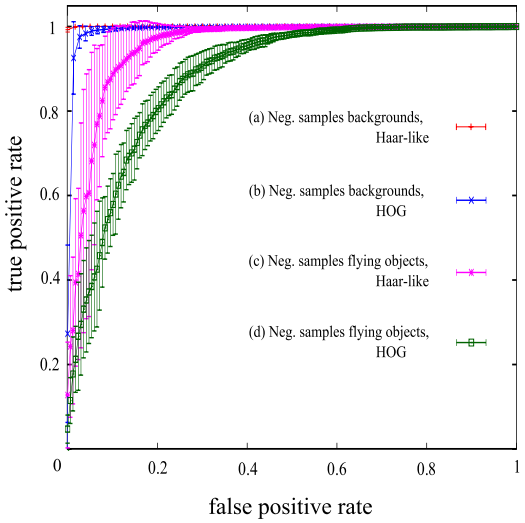

Fig. 6. ROC curves of bird recognition. The red plot is for Haar-like results with negative samples of backgrounds (a), blue is for HOG with backgrounds (b), purple for Haar-like with non-birds (c), and green for HOG with non-birds (d). The curve at the upper left has better recognition precision. Non-bird samples include hard negatives such as trees, insects and airplanes. The results (c) and (d) show the challenging characteristics of the dataset.

according to the species and type annotations. When previously unrecorded regions were detected by background subtraction, these were categorized into background regions, because these are movements of the wind turbine, bushes, or clouds. The bird regions numbered 8,969, non-bird regions 1,118 and background regions 18,688.

We conducted two recognition experiments with different negative samples. First, we experimented on recognition of bird regions and background regions. The purpose of this experiment was to evaluate whether birds can be detected amidst the background. Second, we experimented on recognition of bird regions and non-birds regions. The purpose was to determine whether birds and other flying objects such as insects or airplanes can be classified.

We conducted a five-fold cross-validation to calculate the precision of bird recognition and plotted the results as receiver operation characteristic (ROC) curves (Figure 6). The figure plots the true positive rates (TPR) against false positive rates (FPR) and shows the trade-off between the ratio of correct bird recognitions to misrecognitions. The parameter of the curve is the threshold value in the final stage of AdaBoost. Each plotted point shows a TPR and FPR with one threshold value. We calculated the false positive rate as

$$\mathrm{FPR} = \frac{\#\{\text{false positives}\}}{\#\{\text{negative samples}\}} \tag{1}$$

and the true positive rate as

$$\mathrm{TPR} = \frac{\#\{\text{true positives}\}}{\#\{\text{positive samples}\}} \tag{2}$$

from the test data.

3.3. Results

Figure 6 shows the ROC curves. Among the four curves, the red one (a) shows the Haar-like results with negative samples of backgrounds, the blue one HOG with backgrounds (b), the purple one Haar-like with non-birds (c), and the green one HOG with non-birds (d). The curve at the upper left has better recognition precision. A shared trend in the two experiments with different negative samples is the better precision given by the Haar-like features than by HOG. This suggests that Haar-like is a proper feature for detecting birds at low resolution. This seems to be because low resolution and blurry images have larger effects on HOG features than on Haar-like features.

Fig. 7. Examples of misrecognition in the experiments.

Haar-like features performed well with the negative samples of backgrounds. TPR was 0.99 when FPR was 0. In other words, it missed 1% of the birds when it did not misrecognize the backgrounds as birds. This shows that it is possible to detect birds from this background. However, the precision when including the negative samples of non-birds is worse. TPR is 0.9 when FPR is 0.2 with Haar-like. In other words, it missed 10% of the birds when it misrecognized 20% of the backgrounds as birds. This shows the difficulty of detecting birds from other flying objects, such as insects or airplanes. Categorization between these visually similar classes is still a challenging task.

The better performance of the Haar-like features is an unusual result since HOG has been shown to perform better in various tasks, including nesting seabird detection [8]. This suggests that simple features work better on unclear images such as ours. Considering the better performance with the simple feature, it is worth exploring recognition using low-resolution images.

Figure 7 shows examples of false positives in the experiment. With the negative samples of backgrounds, trees and the blades of the wind turbine are misrecognized as birds. The background subtraction algorithm detected them when the illumination changed or when the wind blew and moved them. To avoid these false positives, we need more robust background subtraction methods and more powerful classifiers.

These results show the unique characteristics of our dataset, which is derived from the natural environment. The recognition performances depend much on both the algorithms and the negative samples as showed in Figure 6. These results may be overlooked in evaluation on more generalized benchmarks. Considering this, it seems to be important to evaluate bird monitoring systems in the natural environments where they actually operate. In this term, our dataset describes the problems in realizing the natural monitoring system and offers a preferable benchmark toward it.

4. CONCLUSION

We constructed an image dataset for ecological investigations of birds in wind farms. This large-scale dataset, including 32,442 sequential images and 32,973 birds appearing in them labeled by experts, is now available. As an application, we conducted bird recognition experiments and showed the unique characteristics of the low-resolution bird detection task. Haar-like features outperformed HOG features on this dataset, and this suggests that the task has different characteristics from others in which HOG has been shown to perform well. Moreover, it became clear that objects similar to birds, such as blades of the wind turbine or insects, can be misrecognized. We believe that the task of low-resolution image recognition presented by our dataset will be one of the next challenges for image processing and computer vision.

5. REFERENCES

[1] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *PAMI*, vol. 34, no. 4, pp. 743–761, 2012.

[2] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," *PAMI*, vol. 30, no. 11, pp. 1958–1970, 2008.

[3] A. Krizhevsky, "Learning multiple layers of features from tiny images," *Computer Science Department, University of Toronto, Tech. Rep.*, 2009.

[4] P. Welinder, S. Branson, T. Mita, Wah C., Schroff F., S. Belongie, and P. Perona, "Caltech-ucsd birds 200," 2010.

[5] T. Berg, J. Liu, S. W. Lee, D. W. Alexander, M. L.and Jacobs, and P. N. Belhumeur, "Birdsnap: Large-scale fine-grained visual categorization of birds," *Proc. CVPR*, 2014.

[6] W.W. Verstraeten, B. Vermeulen, J. Stuckens, S. Lhermitte, D. Van der Zande, M. Van Ranst, and P. Coppin, "Webcams for bird detection and monitoring: A demonstration study," *Sensors*, vol. 10, pp. 3480–3503, 2010.

[7] W.K. Poon, C.J. Wong, K. Abdullah, E.S. Lim, and C.K. Teo, "Development of migratory birds population monitoring system using digital single reflex camera," *Proc. CGIV*, pp. 136–140, 2011.

[8] C. Qing, P. Dickinson, S. Lawson, and R. Freeman, "Automatic nesting seabird detection based on boosted hog-lbp descriptors," *Proc. ICIP*, pp. 3577–3580, 2011.

[9] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," *Proc. CVPR*, pp. 248–255, 2009.

[10] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," *Proc. CVPR*, pp. 304–311, 2009.

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. CVPR*, vol. 1, pp. I–511–I–518, 2001.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proc. CVPR*, vol. 1, pp. 886–893, 2005.

[13] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learing and an application to boosting," *Computational Learning Theory*, vol. 904, pp. 23–37, 1995.