

AUTOMATIC NESTING SEABIRD DETECTION BASED ON BOOSTED HOG-LBP DESCRIPTORS

Chunmei Qing, Patrick Dickinson*, Shaun Lawson* and Robin Freeman†*

*School of Computer Science, University of Lincoln, UK

Email: [cqing](mailto:cqing@lincoln.ac.uk), [pdickinson](mailto:pdickinson@lincoln.ac.uk), [slawson](mailto:slawson@lincoln.ac.uk)

† Microsoft Research Cambridge, UK, Email: a-robfr@microsoft.com

ABSTRACT

Seabird populations are considered an important and accessible indicator of the health of marine environments: variations have been linked with climate change and pollution [1]. However, manual monitoring of large populations is labour-intensive, and requires significant investment of time and effort. In this paper, we propose a novel detection system for monitoring a specific population of Common Guillemots on Skomer Island, West Wales (UK). We incorporate two types of features, Histograms of Oriented Gradients (HOG) and Local Binary Pattern (LBP), to capture the edge/local shape information and the texture information of nesting seabirds. Optimal features are selected from a large HOG-LBP feature pool by boosting techniques, to calculate a compact representation suitable for the SVM classifier. A comparative study of two kinds of detectors, i.e., whole-body detector, head-beak detector, and their fusion is presented. When the proposed method is applied to the seabird detection, consistent and promising results are achieved.

Index Terms— seabird detection, HOG, LBP, AdaBoost, SVM

1. INTRODUCTION

A long-term study of Skomer's guillemot population is being conducted by ecologists. Currently, field researchers use manual methods to gather data, including population size estimates, samples of chick survival rates, and leg-ringing. However, it is not feasible to manually gather more detailed data about the daily activity of birds; for example, how long individuals spend foraging for food. We are developing a system which will help ecologists by automatically extracting detailed information from image sequences of the nesting area. In this system, accurate detection of guillemots in each image is a fundamental task. For example, a researcher may wish to determine the attendance at a specific nest from a corpus of video/image data, and correlate this with manual observations.

Detection in the wild field is a hard problem, because of variations in viewpoint, size, and pose. The problem is made more difficult by the fact that most birds are inclined to group together, or merge into their natural backgrounds.

Visual and audio properties (e.g., shape, colour, songs) are all important keys for bird recognition [2] [3]. Among all visual/acoustic properties, shape features are one of the most distinctive properties [4]. The guillemots' black and white plumage presents both strong image gradients, and distinctive and persistent localized shape characteristics which appear relatively insensitive to orientation. HOG has been widely accepted as one of the best features to capture the edge or local shape information. It has shown great success in object detection and recognition [5] [6]. However, HOG performs poorly when the background is clustered with noisy edges. LBP is complementary in this aspect. It can filter out noises using the concept of uniform pattern [7]. LBP is an exceptional texture descriptor and has been used in various applications and has achieved very good results in face recognition [8]. Our premise is that the appearance of a guillemot can be better captured if we combine HOG and LBP together like in human detection [9] [10].

The large size of the feature vector limits the number of training samples and increases the computation cost in SVM classification. Introduced by Freud and Schapire [11], the boosting algorithm has been successfully used to select Haar-like features for face detection [12] and for learning HOG features for human detection [13]. In this paper, Discrete AdaBoost is used to reduce the dimension of the HOG-LBP feature space and obtain the most discriminating features of seabirds. We then train a linear SVM [14] to perform the classification with the obtained feature.

To further improve the detection performance, two detectors are trained based on the whole-body database and the head-beak database. Detections are applied separately to the testing images/videos and finally, the results are combined.

Our contributions stem from the data domain we are studying: examples of similar systems have been developed for detecting humans in crowded situations [9] [10], but no

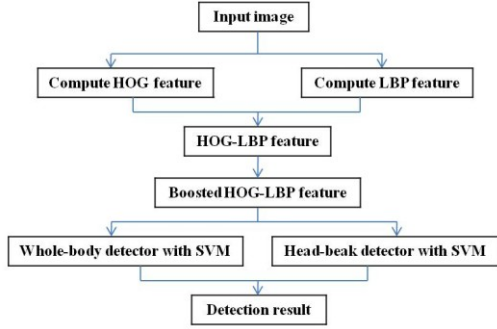


Figure 1. The framework of seabird detection.

other work has developed this as a technique for feature-based detection of seabirds. In addition, we propose a novel augmented feature, boosted HOG-LBP, which boosts four level HOG and LBP to provide the global and local description of objects. We also present a comparative study of two kinds of detectors, i.e., whole-body detector, head-beak detector, and results obtained by fusing these are also provided.

2. THE SEABIRD DETECTION FRAMEWORK

The proposed seabird detection procedure based on the HOG-LBP feature is shown in Figure 1.

A. HOG feature

HOG is an excellent descriptor for capturing the edge direction or the distribution of local intensity gradients of objects [5]. In order to encode both the appearance and the spatial relations of seabirds, we use multilevel HOG features to describe. Figure 2 illustrates the procedure of the HOG feature extraction.

Firstly, for each detection window we compute its gradient magnitude using 1-D masks, i.e., $[-1 \ 0 \ 1]$ (Figure 2(b)). Secondly, we divide the gradient magnitude of the image into non-overlapping blocks at four levels (Figure 2(c)). The first level regards the whole detection window as one block to build a holistic description of a seabird. The other three levels describe the seabird on different level of locality. Each block at each level consists of four rectangle cells. The gradient magnitude of each pixel in the cell is voted into 9 bins according to the orientation of the pixel's gradient. The nine orientation bins are evenly spaced over $0^\circ - 180^\circ$ ("unsigned" gradient). Thirdly, each block is represented by a 36-D feature vector that is normalized by the L2-norm to reduce the influence of the local variation in illumination and foreground-background contrast. Then, the feature vectors of the blocks are concatenated into the feature vector of each level (Figure 2(d)). Finally, four feature vectors corresponding to four levels are concatenated into the final HOG feature.

B. LBP feature

LBP is an excellent texture descriptor for its invariance to gray-scale and rotation [8]. The same as the HOG feature,

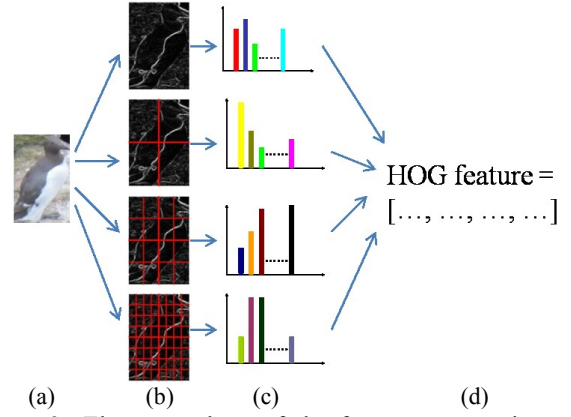


Figure 2. The procedure of the feature extraction of the four level HOG. (a) Input image. (b) Block division at four levels of the gradient magnitude. (c) Histograms of each level. (d) Final HOG feature

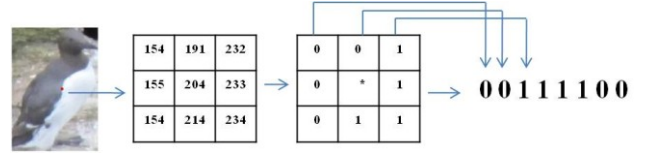


Figure 3. An illustration of the $LBP_{8,1}^{u2}$ pattern calculation.

we use four level block structured LBP to describe the seabird.

The LBP patterns we used is $LBP_{8,1}^{u2}$, where the subscript denotes that 8 points with radius 1 are sampled for each pixel, and the superscript stands for using only uniform patterns. A binary pattern is called uniform pattern if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. For each block at one level, pixels in the block with different uniform patterns are voted into different bins and all of the nonuniform patterns are voted into one bin. We then use the L2-norm to normalize the histograms of the blocks. Finally, the four LBP feature vectors corresponding to the four levels are concatenated into the final LBP feature.

C. Boosted HOG-LBP feature

The size of the feature set is much greater than the number of pixels in the detection window, and many do not contain information useful for classification. AdaBoost algorithm [11] has shown its capability to improve the performance of various classification and detection systems. In order to compute and select local descriptors the discrete AdaBoost algorithm is used. This algorithm selects a small set of discriminative HOG-LBP descriptors in order to achieve robust detection results and reduce the high dimensionality.

The resulted HOG-LBP feature contains important information on how to separate guillemots from other objects, yet redundant information may also be included in the feature. In this paper the AdaBoost is applied to learn a new feature from the HOG-LBP feature at hand. As the HOG-LBP is a histogram with bins indicating local gradient



Figure 4. "The Amos" Skomer Island, and hide position.



Figure 5. Some of the positive samples from our whole-body dataset.



Figure 6. Some of the positive samples from our head-beak dataset.

distribution or texture information, we compare the value on one bin with a threshold to determine whether the image contains guillemots. This forms our weak classifiers in AdaBoost, which are decision stumps. In the iteration processing, AdaBoost select a small number of weighted HOG-LBP features, i.e. weak classifiers, to integrate into a strong classifier. This forms the new feature in our approach, which we name as boosted HOG-LBP feature. A linear SVM is used to train on the feature vector for the final classification.

3. EXPERIMENTS

A. Data Collection

Guillemots monitored on the Skomer Island nest on a cliff face known as "The Amos" (see Figure 4). The position of this cliff face is particularly convenient in that it is viewable from an opposing cliff position with minimal disturbance. The camera was located at the hide position. Firstly, still image sequences were captured using multiple digital still image cameras positioned around the nest site. Camera trap devices were used (which are weather-proofed), and configured to capture a still image approximately every 5 seconds. Secondly, a consumer high-definition camcorder was used to capture video data for several hours a day, over a period of approximately one week. Data was collected at different levels of detail, and from cliff sections with both low and high densities of nests. Reference images of the site were also be captured prior to the start of the breeding season (when no birds are present) [15].

B. Training Processing

We created two dataset for training: the whole-body dataset and the head-beak dataset. For the whole-body dataset, we randomly selected and manually cropped 1000 seabirds from

TABLE 1: DETECTION RESULTS COMPARISONS

		WHOLE -BODY	HEAD- BEAK	FUSION
HOG-LBP	DR	64.8%	67.1%	76.7%
	FN	8075	8075	8075
Boosted HOG-LBP	DR	69.5%	71.9%	79.1%
	FN	300	300	300

the still image sequences and videos. Then 5000 positive seabird patterns were derived from 1000 original seabird patterns by random rotation about ± 10 degree, random scaling about $\pm 10\%$, random mirroring and random shifting up to ± 1 pixel. The 6000 negative examples are randomly sampled from bird-free images. They were all normalized to a size of 48×64 in this dataset. Some of the positive samples are shown in Figure 5. In the head-beak dataset, there are 2000 positive samples and 6000 negative samples. We add some lower body images to the negative set to reduce false alarm. All the images in this dataset were normalized to 32×32 pixels. Some of the head-beak images are shown in Figure 6. We use the linear SVM ($c = 1$) to train and classify on two datasets.

C. Detection Results

During testing, each image sequence is densely scanned from the top left to the bottom right with rectangular sliding windows in different scales. For each sliding window, certain features such as HOG, LBP, HOG-LBP and boosted HOG-LBP are extracted and fed to linear SVM, which is trained offline using labeled training data. We use the SVM to classify the sliding window as enclosing a guillemot (a positive detection), or not. We perform the detection for each single image at 10 scales without considering any temporal smoothing.

We use the detection rate (DR) and feature number (FN) to evaluate the detection results as shown in Table 1. The fusion of the whole-body detector and the head-beak detector get the best performance based on the boosted HOG-LBP feature. Some detection results on a guillemot image are illustrated in Figure 7. In this image, some of guillemots are crowded together. Guillemots are often under some degree of occlusion and exhibit large shape and head pose variations, which challenge the detection. It can be seen that there are false positive samples existing. This is because we just use the original image sequences for detection. No any foreground segmentation processing is adopted for elimination of background.

4. CONCLUSION

In this paper, we propose a solution to the problem of seabird detection, which combines the whole-body detector and the head-beak detector by using the boosted HOG-LBP features. Consistent and promising results are achieved. However, our detector gives a number of false positive

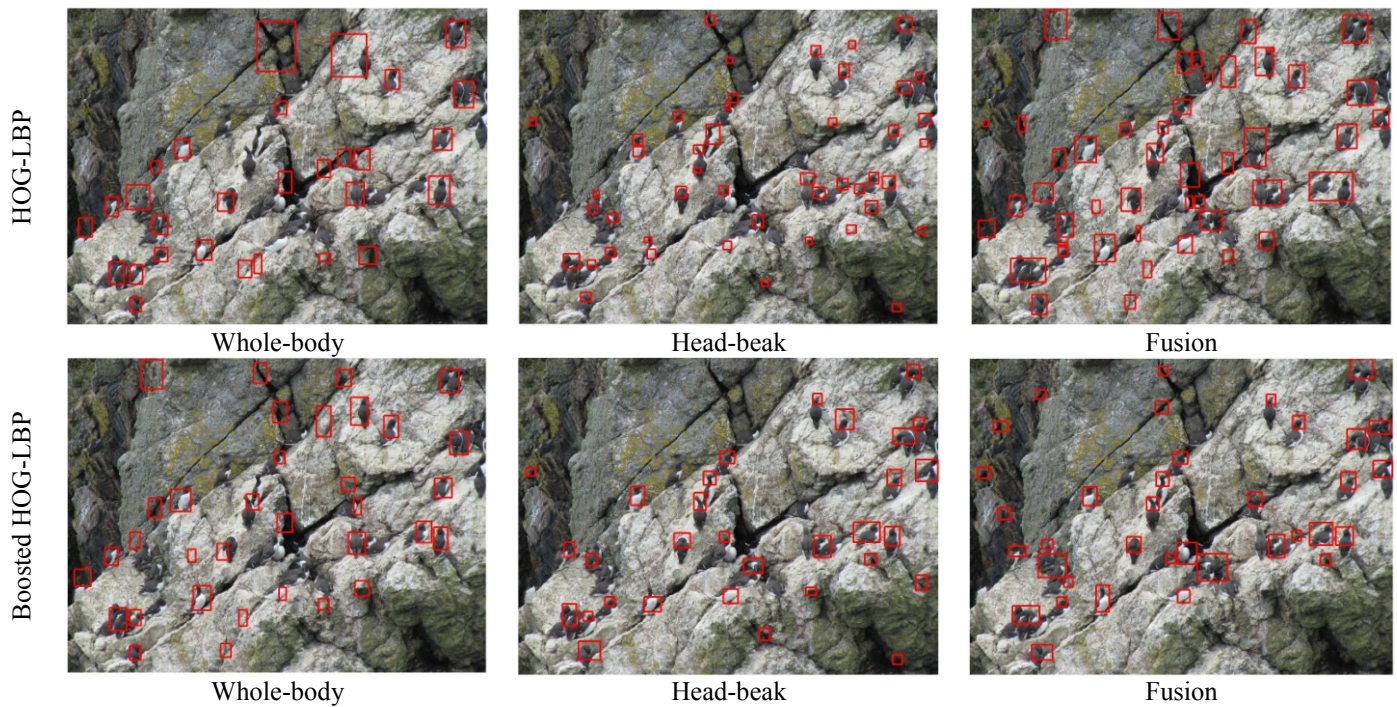


Figure 7. Detection results on a natural guillemot image based on four different types of features for three kinds of detectors.

samples especially in the background area. In our ongoing work, we will integrate a foreground segmentation module [16] to obtain the active areas in the observed area and the detection module to detect guillemots from the detected foreground areas. We believe that false positive samples can be largely eliminated and detection can be speeded up.

Future work includes incorporating motion information using block matching or optical flow fields, looking for image descriptors that are more robust against illumination and other factors, and provisioning detailed population-level data about nesting seabirds, in a non-intrusive manner.

ACKNOWLEDGMENTS

The work presented in this paper is supported by EPSRC grant reference EP/H017143/1.

REFERENCES

[1] S. Votier, T. Birkhead, D. Oro, M. Trinder, M. Grantham, J. Clark and R. McCleery, and B. Hatchwell, "Recruitment and survival of immature seabirds in relation to oil spills and climate variability". *Journal of Animal Ecology*, 77(5), pp. 974-983, 2008.
 [2] Y. Zhang, C. Kwan, D. Lao, and Y. Deng, "Bird classification in noisy environments: Theory, results and comparative studies," in *The proceedings of the 8th Bird Strike Committee USA/Canada Annual Meeting*, St. Louis, Missouri USA, 2006.
 [3] U.D. Nadimpalli, R.R. Price, S.G. Hall, and P. Bomma, "A comparison of image processing techniques for bird recognition," *Biotechnology Progress*, vol. 22(1), pp. 9-13, 2006.
 [4] Bird Identification, <http://www.all-birds.com/Identify.htm>, accessed by 18 Jan. 2011.
 [5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR 2005*, volume 1, pages 886-893, 2005.
 [6] M. Villamizar, J. Scandaliaris, A. Sanfeliu and J. Andrade-Cetto, Combining color-based invariant gradient detector with

HoG descriptors for robust image detection in scenes under cast shadows, *IEEE International Conference on Robotics and Automation*, pp. 1997-2002, Kobe, 2009.

[7] T. Ojala, M. Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51-59, 1998.
 [8] T. Ahonen, A. Hadid, and M. Pietikinen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2037-2041, 2006.
 [9] C. Zeng and H. Ma, "Robust Head-Shoulder Detection by PCA-Based Multilevel HOG-LBP Detector for People Counting," In *ICPR 2010*, pp. 2069-2072, 2010.
 [10] X. Wang, T. X. Han and S. Yan, "An HOG-LBP Human Detector with Partial Occlusion Handling," In *ICCV 2009*, pp. 32-39, Kyoto, 2009.
 [11] Y. Freund and R. E. Schapire. A decision-theoretic generalization of online learning and an application to boosting. *J. Comp. & Sys. Sci.*, 55(1): 119-139, 1997.
 [12] P. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 511-518, Kauai, Hawaii, 2001.
 [13] Q. Zhu, S. Avidan, M.C. Yeh, and K.T. Cheng. "Fast human detection using a cascade of histograms of oriented gradients", *CVPR 2006*, pp. 1491-1498, 2006.
 [14] Corinna Cortes and V. Vapnik, "Support-Vector Networks", *Machine Learning*, 20, 1995.
 [15] P. Dickinson, C. Qing, S. Lawson and R. Freeman, Automated visual monitoring of nesting seabirds. In: *Workshop on visual observation and analysis of animal and insect behaviour*, Istanbul, August 2010.
 [16] P. Dickinson, A. Hunter and K. Appiah, Segmenting video foreground using a multi-class MRF, *International Conference of Pattern Recognition (ICPR)*, pp. 1848-1851, Istanbul, August 2010.