Hierarchical Recognition: Taxonomy of Animals

Seung-kook Yun
Computer Science and Artificial Intelligence Laboratory
MIT, Cambridge, MA, USA
yunsk@csail.mit.edu

***Abstract*— To recognize and animals from a visual image, we propose a hierarchical classification by a taxonomy tree. Trainings is done in bottom-up while recognition goes in top-down along the tree. An image of the species is autonomously assigned its view and trained sequentially at each level of the tree by joint boosting. We can efficiently find a high-level classifier based on low-level ones and this cascading classification allows a high rejection ratio of false alarms. The experimental results of ten sample species are shown to prove the efficiency of the algorithm**

I. Introduction

A goal of this project is to visually recognize a specific animal regardless of a view and its posture, and classify multi-species. Whereas research of a visual recognition was focusing on specific features of one object [1], a human is very good at visually classifying objects into categories even though pictures of them have dramatically different views and postures. Recently, machine learning techniques seem to open a way to categorized recognition in computer vision, and progress has been made about multi-view and multi-class object detection [2], [3]. Visual taxonomy of animals can be a good example of this research. There has been only a little work about taxonomy of sea-animals such as a fish [4] and a crab [5], and some insects [6], most of which are relatively rigid and have good texture to recognize. Normally, animals are highly deformable therefore they would be hardly well-recognized by traditional approaches.

In this project, we propose a hierarchical classification where an image of an animal is classified in multiple times while it goes down along a tree constructed from a hierarchy of animals and views. This cascading enables high rate of rejecting false alarms at cost of a little sacrifice of detection ratio, because an image with an interesting object will pass the classification with high probability while others will with low probability at each level. For training, joint boosting [3] is chosen as the classifiers at each level of the tree because it provides robust classification even with a small number of the training set. To speed up the trainings and get more efficiency in recognition, we shares features not only among horizontal classifiers (multi-view and multi-species) but also in vertical way, by training higher level classifiers based on only the features used in lower level ones. This bottom-up feature sharing and top-down cascading recognition gives a faster and robust way of training and detection.

Section III briefly introduces joint boosting as a way to classify multi-class objects. In section IV, we shows dataset and features, and construct multi-level classifiers along a taxonomy tree. We conclude this project in section V.

II. Related Works

One of the most popular area for categorized recognition is about human faces. Viola and Jones a propose boosting based face detection algorithm [2], which can tell not only faces themselves from other images but also whose faces they are. Oren and etc. also do a good work of recognizing faces and pedestrians by training the wavelet features [7]. They use support vector machine to distinguish the objects from background. Torralba, Murphy and Freeman present a modified algorithm of the boosting called joint boosting [3]. They want to share the features as many as possible among multi classes they intended to classify so that they pursue efficiency in training and detecting. Joint boosting is used to detect multi-views of a car and emotions from a face.

For visual detection of animals, a research team of Bonn university identifies bees by their patterns on the bodies [6]. They found the unique feature of each bee called *fingerprint* and used it to train a classifier. Han and Twefik decribed a crab recognition system based on eigen images of the crabs [5]. Two specific species of crabs are autonoumosly detected by a binary classifier.

III. Joint boosting

Boosting is one of the most popular tools for visual recognition [2], [3] because it enables fast recognition and is immune to over-fitting in some extent [8]. It finds a strong classifier by sequentially adding the best weak learner which reduces error in maximum. Therefore, the strong one has an additive form:

$$H(v,c) = \sum_{m=1}^{M} h_m(v,c)$$

where v is a feature vector, c is a labeled class, and h_m is a weak learner. Each weak learner make a simple yes or no decision (+1 or -1 in binary) based on a threshold.

$$h_m(v,c) = a\delta(v > \theta) + b$$

Boosting estimate optimal parameters a,b, and θ for each weak learner.

Although boosting is known as a binary classifier, Torralba, Murphy and Freeman proposed joint boosting [3] for multi-class recognition. The core of the joint boosting is that it tries to find weak learners shared among classes rather than train separate classifiers for each class, so that it provides faster training as well as requires smaller number of features. This joint boosting looks very promising and appropriate when we have a limited number of source images, because

it tries to find features as much shared as among the classes. They applied joint boosting to multi-view car detection and multi-class recognition and proved its efficiency. Details of the algorithm is given in [3]. We use joint boosting for multi-view and multi-class classification of the given species.

IV. Hierarchical Joint Boosting

In this project, we use top-down recognition and bottom-up feature selection. There are multi-level classifications: starting from a higher level classifier which tells differences among species, following down along the tree of taxonomy as shown in Fig 1, and classifying the image sequentially by a low-level classifier which deals with views of each animal. The classifiers will be introduced in Section IV-B and IV-C. The tree of taxonomy is constructed by bottom-up approach. We extract wavelet features from each image and assign a view by view clustering. Each animal is given 6 views, and the low-level classifier finds the corresponding views of a given image. The high level classifier re-uses the features of the low-level ones to enable faster training as climbing up along the tree. Details are shown in the following sections.

A. Database and Feature selection

We use images of ten species, seven of which are from Caltech256 database [9] while three of them are collected from the internet. For each image, we manually find the ground truth and label it with a rectangular box. The names of the species and corresponding number of the images are shown in Table I.

In order to capture global shapes as well as local ones from the images of the database, we use double density Haar wavelets. The Haar wavelets are known as a good candidate which finds differences in intensities between regions of an image [7]. Double density Haar wavelets are used to extract features from an image. We use three kinds of the filter size which are 32x32, 16x16, and 8x8, respectively. 3 types of 2-dimensional wavelets are implemented to represent the differences in vertical, horizontal, and diagonal as shown in Fig 2. To get a better resolution of the filtered image, an image is sampled by every half size of the filter. Before filtered by the wavelets, an image is normalized to 128x128 size and blurred so that we have a wider distribution of the features. Otherwise, sharp edges of the filtered image would not be shared among similar images because an animal is normally highly deformable.

An example of the features from the wavelets is shown in Fig 3. You can see that the images from 32x32 size wavelets looks like mass distribution of an eagle while 8x8 size appears like contour. We can get over 4000 features per image by the double density Haar wavelet. These features will be used for recognition of views and species.

B. Single species recognition: multi-view classification

Normally, animals are highly deformable and hard to be classified as one category. For example, pictures of dogs include faces, side views with standing and sitting, front views, and so on. Rather than manually sorting pictures into fixed views, we try to cluster images into the most relevant view autonomously. Based on these views, we use joint boosting for multi-view training.

TABLE I
SPECIES AND NUMBER OF THE IMAGES

name	number of images
butterfly	116
dolphin	113
goldfish	152
leopard	215
owl	123
scorpion	80
zebra	147
eagle	98
tiger	50
dog	60

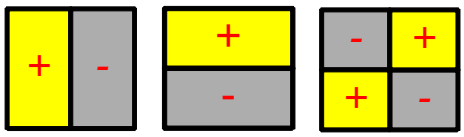


Fig. 2. 3 types of the Haar wavelet: (a)vertical (b)horizontal (c)corner

1) View clustering: We assume 6 views per species, considering front, two sides, rear, and face views. Of course, this may not be optimal and we do not address what is the appropriate number. To cluster images into 6 views, the wavelet features are used as basis. Since they have too many dimensions(over 4000), we reduce them by principle component analysis(PCA). Finally, PCA coefficients of all images of one species are clustered into 6 views by k-means. Fig 4 shows clustered 6 views of a dog.

2) Joint boosting to classify views: For each species, maximum 10 images are trained per view, for some views do not have enough images in them. As negative images, 50 images per the other species are collected in addition to 5000 background images from Caltech256. We found that every classifiers reaches 100% detection ratio for the training set within 150 rounds of the boosting. Fig 5 shows which features are selected by the classifier of a dog.

ROC curves of recognition for untrained images of all the species excluding a butterfly are shown in Fig 6. We omitted the result of a butterfly because it gave a broken result. All the other images which are not included in the training set are classified, and another 5000 background images are also tested to check the false alarms.

To examine relative qualities of the classifiers, we check false alarm rates at 80% detection rate. The results are tablized in Table II. The worst classifier is that of an eagle, and this makes a sense in that it has most dramatic change of postures according to the wingspan.

C. Multi species recognition: Boosting based on the low-level classifiers

So far, we focused on the recognition of a single species. Now considering these classifiers as child nodes of a tree

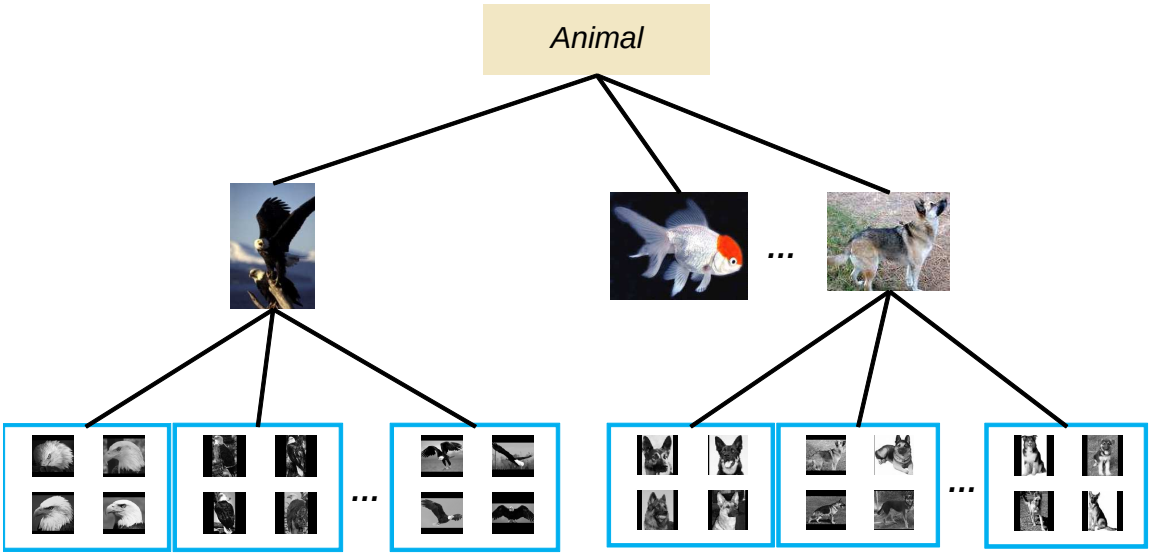


Fig. 1. taxonomy of selected animals

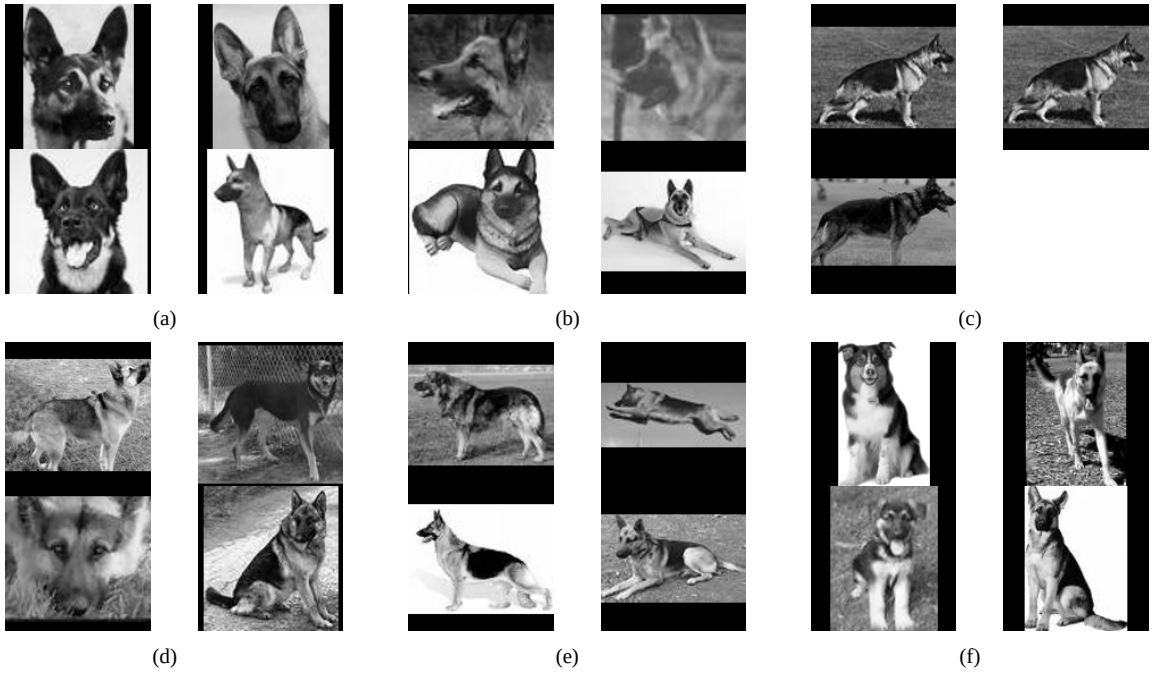


Fig. 4. clustered 6 views of a dog by PCA and k-means

TABLE II
FALSE ALARM RATES OF SPECIES AT 80% DETECTION

species	false alarm rate
dolphin	15%
goldfish	17%
leopard	20%
owl	17%
scorpion	15%
zebra	15%
eagle	30%
tiger	15%
dog	20%

called a taxonomy tree, we build a high level classifier as a parent node.

In this project, we assume only one parent node connected to all the species so that we have two levels of the classification. Note that the structure of a taxonomy tree can be arbitrary. For example, you can bound mammals, insects, fishes and birds into four groups and make a higher level node connected to them.

This tree structure enables faster recognition than we try every view classifier only if we can narrow down possible species within a small number of classifying features, and we will show it works out.

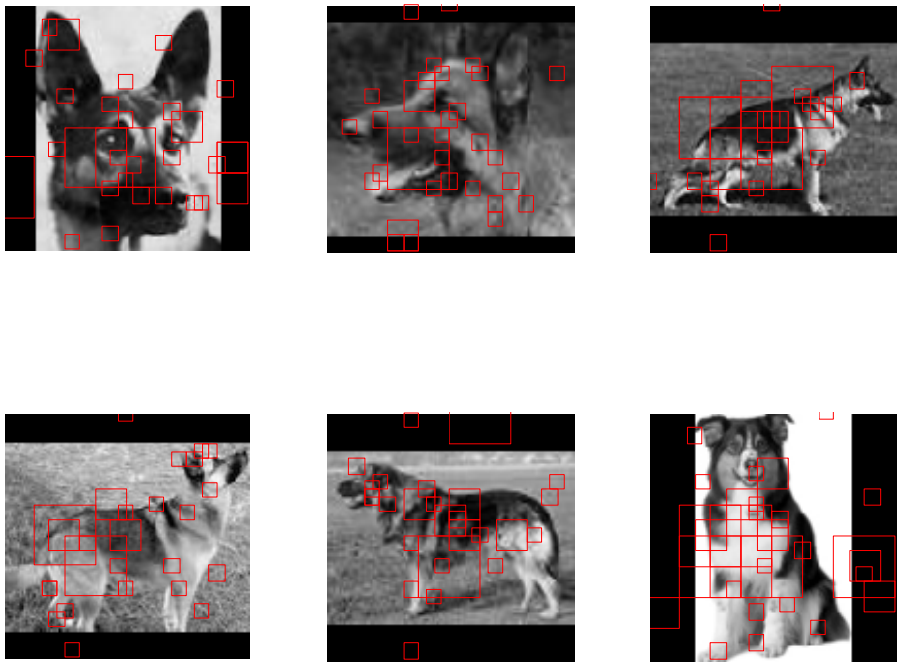


Fig. 5. Selected features by joint boosting classifier, size of the rectangle denotes that of the wavelet

We use joint boosting again to classify species rather than views. Every training images is assigned its label according to species regardless of its view. The wavelet features are still available, however exploiting all the features is not necessary, because we already have the low level classifiers. Rather than finding classifying features from all of them, we suggest to restrict a pool of the features by the ones selected by the low-level classifiers. This makes a sense in that they tried to find mostly shared features in one species when we implemented joint boosting. Fig 7 proves this suggestion by comparing the ROC curve of the training set from the features of the view classifiers with that from all the features. We trained the classifier by the same images used in view classifying. Both cases converge to perfect recognition after 300 rounds of the boosting, however in low number of rounds the former outperforms the latter, which means our method can give even better result as well as much faster.

Fig 8 shows the ROC curves for untrained images with the classifiers from 100, 200 and 500 rounds joint boosting. The curve of 500 rounds does not give a better result than that of 100 rounds. It may be from the small number of the images because only about 50 images per species are used in the training. The graph tells that we can recognize a species with 90% detection rate and 15% false alarm, by checking 100 features. Note that each boosting round finds one feature, and a corresponding threshhold and a weight.

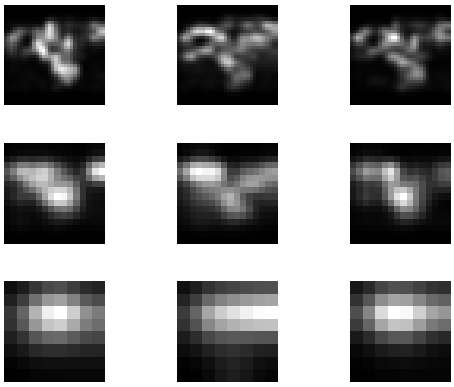
D. cascading joint boosting

Now we have two levels of the classifiers from joint boosting. In average, each levels has similar detection and false alarm rate. Note that over 10% false alarm rate is quite large becuase we will meet many more negative images than positive ones - a negative image means it does not include an interesing object - when we are searching an image for an animal. To reduce the false alarm rate and also to exploit the hierarchy, an image is classified twice from the high level (species) to the low level (views) as it moves down along the taxonomy tree. If it has object in it, double classifying will gives positive detection with about 70% probability. while a negative image will be mis-classified as a positive one in only 2% chance. Note that the resulting probability of detection is multiplication of each level's probability. We lose some positive detection rate, whereas we can reject much more false alarms. Furthermore, this is faster than trying every low-level classifier, since we need only 100 features at the high level and they also include some features of the corresponding low level features. Note that the high level one is made from the low-level ones.

V. Conclusion

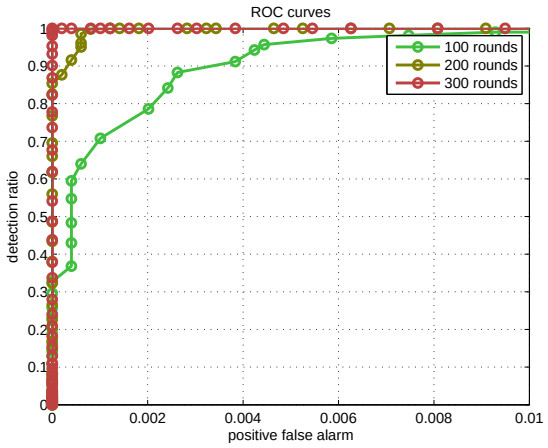
We propose a hierarchical recognition of animals based on cascading joint boosting. An image is classified at each level of the taxonomy tree as going down along it. Multi-level

(a)

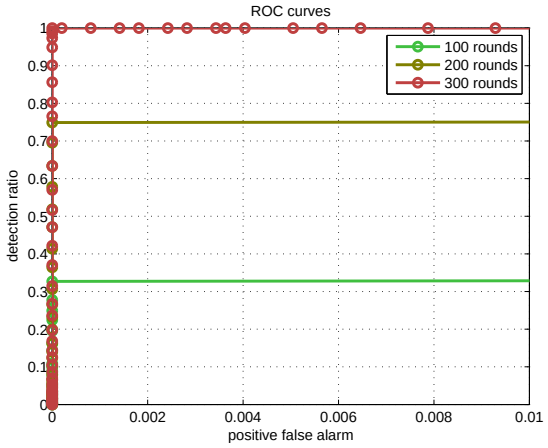


(b)

Fig. 3. (a) an image of an eagle (b) wavelet features: filtered by 8x8, 16x16, and 32x32 size wavelets in descending order



(a)



(b)

Fig. 7. ROC curves of (a) the classifier from the features of the low-level classifiers (b) the classifier from all the features

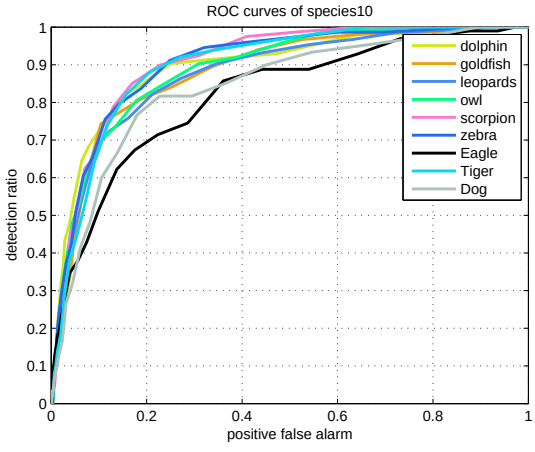


Fig. 6. ROC curves of the untrained images

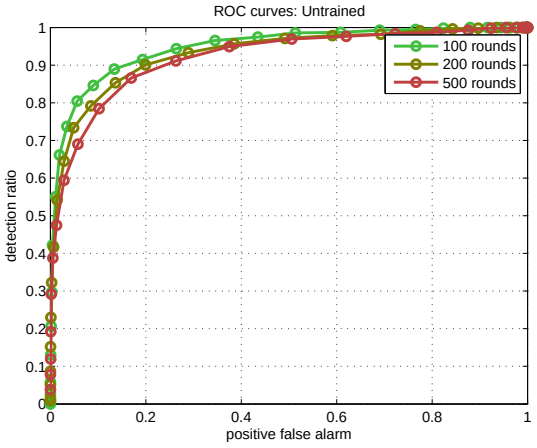


Fig. 8. ROC curves of the multi-species recognition

classification is proven useful to reject false alarms at cost of reduced positive detection rate. Trainings of the classifiers are processed efficiently in bottom-up, by using only small amount of the features that are included in the lower-level classifiers. Ten species of the animals are classified with 2 levels and 6 views, and the classifiers of each level gives 80% detection rate and 20% false alarm and 90% and 15%, respectively. There are many future work. Above of all, more images are required to learn better classifiers. We should think of better features for extremely deformable animals like an eagle. In addition, there may be a better way to cluster views.

VI. Acknowledgements

I am grateful to Prof. Freeman for his advice. Thank Dr. Torralba for kindly providing the source codes of the joint boosting.

References

[1] D. Lowe, "Object recognition from local scale invariant features," in *International Conference on Computer Vision*, Vancouver, Canada, July 2001, pp. 525–531.

[2] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, pp. 137–154, 2004.

[3] A. Torralba, K. Murphy, and W. Freeman, "Sharing visual features for multiclass and multiview object detection," *Technical Report, MIT*, 2004.

[4] Z. B., S. A., and K. I., "Sorting fish by computer vision," *Computers and electronics in agriculture*, pp. 175–187, 1999.

[5] K. J. Han and A. H. Tewfik, "Expert computer vision based crab recognition system." [Online]. Available: citeseer.ist.psu.edu/464372.html

[6] "Automatic identification of bees," Bonn University, Germany. [Online]. Available: www.informatik.uni-bonn.de/projects/ABIS

[7] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian detection using wavelet templates," in *In Proc. Computer Vision and Pattern Recognition*, 1997, pp. 193–199. [Online]. Available: citeseer.ist.psu.edu/oren97pedestrian.html

[8] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," 1998. [Online]. Available: citeseer.ist.psu.edu/friedman98additive.html

[9] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Institute of Technology, Tech. Rep. 7694, 2007. [Online]. Available: http://authors.library.caltech.edu/7694