STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LX**, Number 2, 2015

OBSTACLE RECOGNITION IN TRAFFIC BY ADAPTING THE HOG DESCRIPTOR AND LEARNING IN LAYERS

ROXANA MOCAN AND LAURA DIOŞAN

Abstract. Despite many years of research, obstacle recognition is still a difficult, but very important task. We present a multi-class approach, that extracts from images the Histogram of Oriented Gradients (HOG) based on aspect ratio of Region of Interest (ROI) and use them in a multi-class classification problem. For the learning phase we propose an original approach based on decision trees. Numerical experiments are performed on a benchmark dataset consisting of animal, pedestrian, car and sign (labeled) images captured in outdoor urban environments and indicate that the proposed model is able to improve the performance of the recognition process.

1. Introduction

Nowadays it is necessary to increase the speed to keep up with traffic and that triggers a higher risk for accidents. This risk is also increased by the high number of road users. Statistics show clearly that the number of accidents and casualties (drivers and pedestrians) has got alarming levels (it is necessary to reduce the number of traffic events). The surveillance of pedestrians, cars, motorcycles and animals in traffic is important for increasing safety.

In this field researchers had made great improvements, classifying each category individually, but using a multiclass classification algorithm for most frequent objects in traffic scene can be useful and faster. In many machine learning methods, a binary classifier is easy to construct, while, in most applications, a multiclass classifier is needed.

A classification task with more than two classes where objects belonging to the same class may vary from each other in views or shapes, has an increased difficulty. Usually this type of problem decomposes trivially into a set of unlinked binary problems.

Received by the editors: June 16, 2015.

2010 *Mathematics Subject Classification.* 68T05, 91E45.

1998 *CR Categories and Descriptors.* I.2.6 [**Artificial Intelligence**]: Learning – *Induction*; I.2.6 [**Artificial Intelligence**]: Learning – *Concept learning*.

Key words and phrases. Multiclass classification, HOG, Decision Trees, Boosting.

The most used algorithm is OAA (One Against All). For this method k classifiers are needed (k is the number of classes). In this method you construct the ith classifier using ith class as positive and the rest of the classes as negative, $i \leq$ number of classes. To predict the class that a test sample belongs to, each classifier has to be verified and to take a vote of confidence. The vote of confidence takes the majority answer from each classifier. Having to train all samples for each classifier involves higher computational time and also means that the number of samples will not be balanced because if x is the number of samples from each class, each classifier will get x positive images and $x * (k-1)$ negative images.

The proposed approach wants to improve the results by equilibrating the positive and negative data and reduce training and testing time by making only $k-1$ classifications and k/2 predictions without taking the vote of confidence.

The outline of the paper is as follows. After briefly reviewing related work in Section 2, we present the background of our system in Section 3. Numerical experiments are presented in Section 4, while the conclusions and further work are highlighted in Section 5.

2. Related work

There are great approaches in this field, researchers used learning for binary classification or for multiclass classification and robust feature extractor. In some papers [3] they found that histogram of oriented Gradients (HOG) can be used for pedestrian detection. The utility of boosting [5] was proved by making the weak classifiers to generalize well and improving prediction results by creating a strong multiclass classifier for urban traffic objects. Creating a hierarchical clustering to create a decision tree by merging classes repeatedly [6] and finding most similar classes using Euclidian distance. Support Vector Machine (SVM) have been successfully employed for a variety of different multiclass classification and regression tasks [2]. Recently in [9], [1] methods that use simple appearance based features which also take advantage of the temporal information, are discussed. Chen et al. introduced the e Differences of Histograms of Oriented Gradients (DHoG) feature based on the changes introduced in the HOG descriptor due to rigid and non rigid motion of vehicles and pedestrians.

3. Proposed approach

The purpose of this work is to correctly classify different type of objects from the road assistance field using a mono-camera. To reach this goal we propose a system which has two components, one for image processing and the other for object recognition. The novelty is from the learning part, since

in the image processing part we used HOG descriptors (the literature proving their effectiveness [3]).

3.1. **Processing the images.** Histogram of Oriented gradients is a robust descriptor that can be described as the distribution of the intensity gradients or edge direction. Each image is divided in regions and each region is divided in four cells. The gradients and orientations from a cell are computed and results a histogram with several bins. Usually the cell has a fixed number of pixels, but the number can vary. Concatenating all four histograms will result a vector. To have the entire HOG vector, only concatenate all regions from image.

In most approaches images are resized, to get a fixed size of HOG vector. For example we can resize all images to 128 x 128 pixels and each region get the fixed number of pixels (16 pixels per region), and get 64 regions from each image. From 64 regions * 36 (region size) = 2304 feature vector. But we can get the same size for feature vector without resizing. For example we have an image with a pedestrian and the image size is 64 x 128 pixels, we only set the cell size at 8 columns and 16 rows. The differences between these 2 methods for HOG computation are showed at Numerical Experiments section.

3.2. **Learning in layers.** The proposed method for learning is using the binary decision tree principles. In a decision tree, each node splits the instance space in two sub-spaces according to a rule. In this case, the rule is if an instance is from positive class or negative class.

The proposed method classifies all data in layers. Each layer is attached to a classifier and the decision rule is in which layer to go if the prediction is positive or negative.

The first layer will classify all k classes, but the second and third layer will classify only $k/2$ classes ($k/2$ classes will be classified in second layer and the rest will be classified in third layer). If the number of samples from each class are equal then the positive and negative numbers of samples are equilibrate. Also using this method we don't need any confidence vote that also takes time.

3.3. **Performance measurements.** The performance measurements that could be taken into account in the case of a classification problem are:

- the true positive rate (TPR): number of positive samples that are predicted well / total number of positive samples,
- the false positive rate (FPR): number of negative samples that are predicted as positive / total number of negative samples
- the precision: the percent of relevant classification among the proposed ones;

- F-measure: the weighted harmonic mean of precision and recall. Grater F-measure (the maximal value being 1) signifies a correct and complete classification;
- the accuracy (Acc): (number of positive samples that are predicted well + number of negative samples that are predicted well) / total number of samples).

4. Numerical experiments

4.1. **Data sets.** For this paper we used four classes: animal, pedestrian, car, signs. The training and testing samples are regions of interest from traffic scene images. The images are gathered from internet: Caltech image data base [4], Inria Person Dataset [3] and LISA dataset [7], [8].

The total number of images is 12037 where 4000 for animal, 4000 for pedestrians, 2547 for cars and 1490 for signs.

The training set has 3500 images with animals, 3500 with pedestrians, 2447 with cars and 1390 with signs. The testing set has 500 images with animals, 500 with pedestrians, 100 with cars and 100 with signs. The samples are regions of interest from cropped images from traffic scene images.

4.2. **Image processing.** Image processing algorithms are implemented using OpenCv libraries.

All images are grey and for feature extraction we used HOG algorithm. From each image we extracted 2304 features (64 blocks * 4 * 9 bins per block). Because the regions of interest from images had different aspect ratio (pedestrian ≈ 0.319, animal ≈ 1.347, signs ≈ 1.1508 and car ≈ 2.5) it was necessary to take different sizes of cells to get the same size of HOG vector. We made that just for simply concatenate class matrixes but it showed that computing HOG in this way made one class samples discriminative from other classes.

We also computed HOG with images that are resized to 128 x 128 pixels and the differences are showed in Tables 1 and 3.

4.3. **Learning and testing.** Machine Learning algorithms are implemented using OpenCv libraries.

For learning we used boosting algorithm with 200 binary decision trees as weak classifiers, with max depth = 2.

As training samples we took for first layer two classes and labeled them as positive, and the other two as negative and trained them using boosting. We choose which class goes to positive and negative by observation the data and the images. This layer uses 10837 images. For second layer we took one class from first layer positive and set as positive for second layer too, and the other class from first layer positive as negative for second layer and trained

them with boosting too. It can be seen that the second layer uses only 5947 images ($\approx 1/2$ from all images).

For third layer we took one class from negative samples from first layer and divided in two (positive and negative), and then train with boosting, the same as first and second layer. For this layer are used 4890 images. For more classes the numbers of layers are increasing and number of samples decreasing.

For testing the first layer we took all the test images from all classes and compute the prediction vector. If for an image the prediction from first layer says that is positive, it goes directly to the second layer and compute the second layer prediction, in case the prediction says that is negative that sample goes to third layer and compute the prediction that says from which class is it.

We trained the same samples using OAA creating four models. For us this means, for training, 3500 samples positive and 7337 negative for animal and pedestrian classes, 2447 samples positive and 8390 negative for cars class and 1390 positive and 9447 negative for signs class. Each model results from training using boosting with the same parameters as learning in layers.

4.4. **Experiment 1.** For the first step we proposed to compare our learning approach to OAA for images that are resized. The results (TPR, FPR, Precision, Recall, F-measure for each class, respectively and Global Accuracy and its confidence interval (for a probability of 95%) for all classes) are represented in Table 1. In addition, in Table 2 the training and testing time involving in this experiment are presented.

		Animal	Pedestrain	Car	Road sign
LiL	TPR	0.66	0.73	0.71	0.99
	FPR	0.18	0.16	0.04	0.03
	Precision	0.72	0.76	0.61	0.75
	Recall	0.66	0.73	0.71	0.99
	F-measure	0.69	0.74	0.66	0.99
	Global Acc		0.776±0.023		
OAA	TPR	0.7	0.75	0.72	0.98
	FPR	0.19	0.21	0.04	0
	Precision	0.72	0.71	0.62	1
	Recall	0.7	0.75	0.72	0.98
	F-measure	0.71	0.73	0.66	0.98
	Global Acc		0.714±0.025		

TABLE 1. Comparison of performances obtained by learning in layers and OAA for resized images

The numerical results from Tables 1 and 2 indicate:

- a better performance (Global accuracy) of the proposed approach in comparison with OAA;
- a better TP rate for car and sign classes, but weaker for animal and pedestrian;
- a better FP rate for animal, pedestrian and sign, but weaker for car;
- the training time for proposed approach is improved;
- the testing time is also better.

	LiL		OAA	
	Training	Testing	Training	Testing
All classes	3069.41	0.012	34544.53	4.535

TABLE 2. Comparison of LiL and OAA running time (seconds) for resized images

4.5. **Experiment 2.** For the second step we proposed also to see if the representation alternative where we changed HOG representation instead of changing image size could imrpove the classification performances. The results (TPR, FPR, Precision, Recall, F-measure for each class, respectively and Global Accuracy and its confidence interval, for all classes) are represented in Table 3. Again, we give the training and testing time (see Table 4).

		Animal	Pedestrain	Car	Road sign
LiL	TPR	0.94	0.99	0.82	1
	FPR	0.025	0.011	0.018	0.001
	Precision	0.96	0.98	0.8	0.99
	Recall	0.94	0.99	0.82	1
	F-measure	0.95	0.98	0.81	0.99
	Global Acc		0.94±0.013		
OAA	TPR	0.95	0.98	0.56	0.94
	FPR	0.074	0.001	0.067	0.006
	Precision	0.9	0.99	0.43	0.93
	Recall	0.95	0.98	0.56	0.94
	F-measure	0.92	0.98	0.48	0.93
	Global Acc		0.861±0.019		

TABLE 3. Comparison of learning in layers and OAA rates for unmodified images

Results from Tables 3 and 4 indicate that:

	LiL		OAA	
	Training	Testing	Training	Testing
All classes	2588.5	0.007	2716	4.34

TABLE 4. Comparison of learning in layers and OAA running time (seconds) for unmodified images

- the proposed approach has better general performance than OAA; TP rate is increased for pedestrian, car and sign classes and weaker for animal class;
- FP rate is higher for animal, car and sign classes, but worst for pedestrians; training time for proposed approach is improved; testing time is also higher.

Differences between the proposed method and One Against All can be summarized as follows.

In the training stage:

- Time for training is shorter because in OAA we use for each classifier 10837 samples and in the proposed method we train only in first layer 10837 samples, in second and third the number of training samples is 5947 and 4890.
- In OAA we need 4 classifiers and in the proposed method we need only three.

In the testing stage:

- in OAA we have compare with four models, and in this method only with two.
- we need to take the confidence vote, in the proposed method we don't.
- the disadvantage for this method is the propagation of the error. If a sample is miscalculated in first layer, can't be recovered in the next layers.

5. Conclusions

Differences between the proposed method and One Against All can be summarized as follows.

An important problem was investigated in this paper: obstacle recognition in images. Each image was represented by using HOG descriptors, whose parameters were adapted to the image size. The extracted features were incorporated finally into a classifier based on decision trees in order to construct a decision model that can be utilized in order to label un-seen images. The learning stage is developed on layers, in order to help the classification model.

We have studied how the process of image resizing vs. adapting the HOG's parameters influence the results of multi-class classification. The results obtained by using the adapted descriptor indicate a better performance of the decision model. Furthermore, the special learning approach improve the classification performances (in comparison to the classical OAA model).

As further work we plan to validate our approach by using more data for this classes or other classes, model validation (cross-validation) and to adapt other image descriptors and also to investigate other types of learning strategies.

References

[1] L. B. Chen, R. S. Feris, Y. Zhai, L. M. Brown, and A. Hampapur. An integrated system for moving object classification in surveillance videos. In *Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on*, pages 52–59, 2008.

[2] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20:273, 1995.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Conference on*, pages I: 886–893, 2005.

[4] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 304–311, 2009.

[5] I. Giosan, A.D. Costea, and S. Nedevschi. Urban traffic dense-stereo obstacle classification using boosting over visual codebook features. In *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, pages 111–116, 2013.

[6] X. Huaitie, S. Fasheng, and L. Yongsheng. Support Vector Machine algorithm based on kernel hierarchical clustering for multiclass classification. In *Electrical and Control Engineering (ICECE), 2010 International Conference on*, pages 2201–2204, 2010.

[7] Andreas Mogelmose, Mohan M. Trivedi, and Thomas B. Moeslund. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1484–1497, 2012.

[8] Sayanan Sivaraman and Mohan M. Trivedi. A general active-learning framework for on-road vehicle recognition and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):267–276, 2010.

[9] L. Zhang, S. Li, X. Yuan, and S. Xiang. Real-time object classification in video surveillance based on appearance learning. In *Computer Vision and Pattern Recognition, 2007. CVPR 2007. IEEE Conference on*, pages 1–8, 2007.

Department of Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania

E-mail address: roxanamocan@ymail.com, lauras@cs.ubbcluj.ro