2015 IEEE 18th International Conference on Intelligent Transportation Systems

Combining shape-based and gradient-based classifiers for vehicle classification

Hakki Can Karaimer, Ibrahim Cinaroglu, Yalin Bastanlar
Computer Vision Research Group, Dept. of Computer Engineering,
Izmir Institute of Technology
Izmir, Turkey
{cankaraimer, ibrahimcinaroglu, yalinbastanlar}@iyte.edu.tr

***Abstract*—In this paper, we present our work on vehicle classification with omnidirectional cameras. In particular, we investigate whether the combined use of shape-based and gradient-based classifiers outperforms the individual classifiers or not. For shape-based classification, we extract features from the silhouettes in the omnidirectional video frames, which are obtained after background subtraction. Classification is performed with kNN (k Nearest Neighbors) method, which has been commonly used in shape-based vehicle classification studies in the past. For gradient-based classification, we employ HOG (Histogram of Oriented Gradients) features. Instead of searching a whole video frame, we extract the features in the region located by the foreground silhouette. We use SVM (Support Vector Machines) as the classifier since HOG+SVM is a commonly used pair in visual object detection. The vehicle types that we worked on are motorcycle, car and van (minibus). In experiments, we first analyze the performances of shape-based and HOG-based classifiers separately. Then, we analyze the performance of the combined classifier where the two classifiers are fused at decision level. Results show that the combined classifier is superior to the individual classifiers.**

Keywords—vehicle classification, shape-based classification, gradient-based classification, histogram of oriented gradients, combined classifier, omnidirectional cameras.

I. INTRODUCTION

Object detection and classification are important tasks for many research and application areas including transportation management, traffic safety and intelligent vehicles. Quite a variety of approaches have been proposed for object detection. A major group in these studies uses the sliding window approach in which the detection task is performed via a moving and gradually growing search window. Features based on gradients, colors, etc. can be used for classification. A significant performance improvement was obtained with this approach by employing HOG (Histogram of Oriented Gradients) features [1]. Later on, this technique was enhanced with part based models [2]. In [3] and [4], HOG features are used for vehicle detection as well. Using Haar-like features is another option for the sliding window approach. Example studies include face detection [5] and car detection [6].

Another major approach for object detection and classification is using shape-based features. Since vehicles are moving objects and their shapes can be segmented from the background image, this approach was commonly used for vehicles in the past. For instance, in [7], authors created a feature vector consisting of area, breadth, compactness, elongation, perimeter, convex hull perimeter, length, axes of fitted ellipse, centroid and five image moments of the foreground blobs. Classification is performed by weighted kNN (k Nearest Neighbor).

These two major approaches are compared in [8]. HOG or Haar-like features are named as image based features and the features of the shape-based approach are called image measurement based features. It was stated that using simple measurements extracted from the shapes is computationally cheaper. Extracting image based features for each position of sliding window requires a considerable amount of time. To avoid the computational load for image based features approach, in the study of [3], windows encompassing the moving vehicles are located manually. We are also able to examine the performances of the mentioned two approaches on standard images. The accuracy of the HOG-based method in [3] is lower than the accuracy of shape-based classification in a previous work by the same authors [7].

In this paper, we are not in defense of one these two main classification approaches. We investigate if the cooperative use of the shape-based and HOG-based classifier increases the performance of classification. Moreover, we perform this analysis for omnidirectional camera videos. Omnidirectional cameras provide 360 degree horizontal field of view in a single image (vertical field of view varies). If a convex mirror is placed in front of a conventional camera for this purpose, then the imaging system is called a catadioptric omnidirectional camera. An example image is given in Fig. 1. This enlarged view is an important advantage in many application areas such as virtual tours [9], robot navigation [10] and 3D reconstruction [11]. So far omnidirectional cameras have not been widely used for vehicle detection partly due to the fact that the objects are warped in omnidirectional images and techniques developed for standard cameras should be modified.

For shape-based classification, we apply kNN with the features extracted from silhouettes obtained after background subtraction. The extracted features are convexity, elongation, and rectangularity. The vehicle types that we worked on are motorcycle, car and van (minibus). Vehicle classification with kNN was used many times before (e.g. [7][8][12][13]) and can be considered as the benchmark method of shape-based vehicle classification.

978-1-4673-6596-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ITSC.2015.135

IEEE computer society

For the gradient-based classifier, we chose HOG+SVM classification since it is a commonly used method in visual object detection. Since computing the HOG features for sliding windows requires considerable time for the whole image, we only extract features in the region located by the foreground silhouette obtained by the background subtraction.

The performances of these two classification approaches are computed separately and compared with the case where they are combined together to make a decision. While combining methods we use a decision-level fusion rather than combining the features and training a model. We obtain scores from each classifier separately and merge them. In our experiments, we observed that the combined method performs better than the individual classifiers.

Our omnidirectional video dataset, together with annotations and binary frames after background subtraction, can be downloaded from our website (http://cvrg.iyte.edu.tr/). The organization of the paper is as follows. In Section II, we give more related work on vehicle detection and classification. Also a summary of the vehicle detection with omnidirectional cameras is presented. In Section III, we introduce the details of silhouette extraction method. Shape-based, HOG-based and combined classification methods are described in Sections IV, V and VI respectively. Experiment results are presented in Section VII. Finally, we give conclusions in Section VIII.

II. Related Work

In one of the earliest studies on vehicle classification with shape-based features [14], following the background subtraction step, location, length, width and velocity of foreground fragments are used to classify vehicles into two categories; cars and non-cars. In another study, [15], authors use position and velocity in 2D, the major and minor axes of the ellipse modeling the target and the aspect ratio of the ellipse as features in a Bayesian Network. In a 3D vehicle detection and classification study which is based on shape based features, [16], they use the overlap of the object silhouette with region of interest mask which corresponds to the region occupied by the projection of the 3D object model on the image plane. In a ship classification study, researchers use MPEG-7 region-based shape descriptor which applies angular radial transform to a shape represented by a binary image. They classify ships into six types with kNN [12].

Instead of standard video frames, some researchers employed time-spatial images, which are formed by using a virtual detection line in a video sequence. In this manner, authors in [13] construct a feature vector obtained from the foreground mask. Employed features are width, area, compactness, length-width ratio, major and minor axis ratio of fitted ellipse, solidity and rectangularity. The samples are classified by kNN.

Regarding HOG-based classification with omnidirectional cameras, in the study of [3], HOG features are computed on virtual perspective views generated from omnidirectional images. In [17], as a part of the proposed human tracking algorithm, HOG features are computed on the omnidirectional image. However, a rectangular rotating and sliding window is used with no mathematical modification for the omnidirectional camera. In [4], authors modify HOG computation for omnidirectional camera geometry and show that the performance of using modified HOG on omnidirectional images is superior to the regular HOG on omnidirectional and panoramic images.

Regarding the shape based classification studies with omnidirectional cameras, [18] uses only the area of the blobs and classifies them into two classes; small and large vehicles. In one of our previous studies with omnidirectional cameras [19], we detect each vehicle type separately using shape features extracted from the foreground silhouettes.

III. Silhouette Extraction in Omnidirectional Videos

The silhouettes are obtained after a background subtraction step and a morphological operation step. For background subtraction, the algorithm proposed by [20] is used, which was one of the best performing algorithms in the review of [21]. The final binary mask is obtained by an opening operation with a disk, after which the largest blob is assigned as the silhouette belong to the moving object.

Features are extracted from the frame where the silhouette that belongs to the vehicle is closest to 0° (the direction in which the camera is closest to the road). Then the silhouette is rotated with respect to the center of omnidirectional image so that the center of the silhouette is at the level of the image center. This operation is described in Figure 1.

In a previous study, we showed that using multiple frames of a video has a better classification performance than using a single silhouette in a single frame [19]. Among a few approaches of using multiple silhouettes, 'average silhouette' was the best performing one. Therefore in this paper, we use the average silhouette method for shape-based classification.

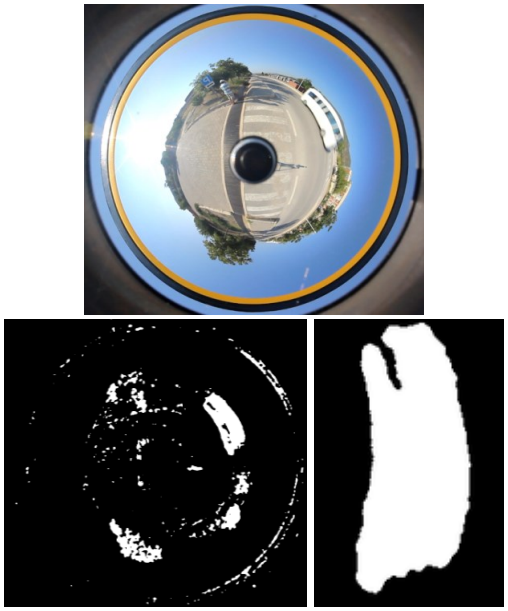

Fig. 1: Top: An example omnidirectional video frame containing a van while passing the road. Bottom-left: The same frame after background subtraction. Largest blob is assumed to belong to the moving object. Bottom-right: Rotated blob after morphological operations.

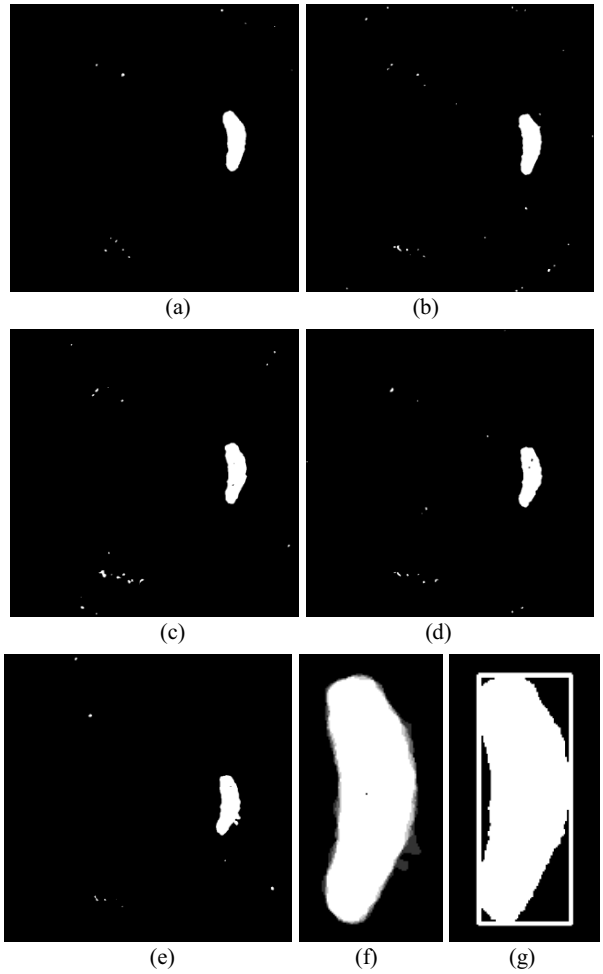

Fig. 2: Example binary images from a video containing a car instance. The centroid of the foreground object is at (a) +4.6° (b) +2.7° (c) 0.5° (d) -1.7° (e) -3.8° (a negative value shows the centroid is below image center). (f) Resultant 'average silhouette' obtained by the largest blobs in the binary images. (g) Thresholded silhouette and the minimum bounding rectangle.

To obtain an 'average silhouette' we define an angle range. The frames (silhouettes) are used only if they are in this range. Fig. 2 shows the silhouettes of a sample video where the angle range is [+5°,-5°]. Silhouettes are added to each other so that the center of gravity of each blob coincides with others. The cumulative image is divided by the number of frames which results in 'average silhouette' (Fig. 2f). We then apply an intensity threshold to convert average silhouette to a binary image and also to eliminate less significant parts which were supported by a lower number of frames. Thus we can work with more common part rather than taking into account every detail around all silhouettes (Fig. 2g). The threshold we select here eliminates the lowest 25% of grayscale levels.

IV. Shape-Based Classification

A. Feature Extraction

The features extracted from the foreground silhouette, average silhouette in our case, are introduced in the following.

1) The convexity is defined as

$$Convexity = \frac{O_{Convexhull}}{O} \quad (1)$$

where $O_{Convexhull}$ is the perimeter of the convex hull and O is the perimeter of the original contour [22].

2) Elongation is computed as follows

$$Elongation = 1 - W/L \quad (2)$$

where W is the short and L is the long edge of the minimum bounding rectangle (Fig. 2g) which is the smallest rectangle that contains every point in the shape [22].

3) Rectangularity measures how much a shape fills its minimum bounding rectangle [22]:

$$Rectangularity = A_S \,/\, A_L \quad (3)$$

where A_S represents the area of a shape and A_L represents the area of the bounding rectangle.

We did not use silhouette area itself as a feature since we use a portable image acquisition platform and we want our method to be independent of the distance between the camera and the objects. We also computed a few other features such as solidity and ellipse axes ratio but they did not improve the performance and we exclude them from the experiments we present in this paper. Also, having three features enables us to plot 3D feature space where we can observe the discriminative power of each feature individually (cf. Section IV.B).

B. K Nearest Neighbor Classification

Our shape-based classifier uses kNN on the features described in the previous subsection. In kNN classification, a new sample's label is determined according to the majority of the nearest *K* neighbors' labels in the training set.

Fig. 3a shows the features of the annotated (training) silhouettes of all samples in 3D where dimensions are rectangularity, elongation and convexity. Actual labels are indicated with different colors. We can see that looking at *K* nearest neighbors of a new sample can help us to determine the sample's label. Top-view of Fig. 3a is shown in Fig. 3b, where x and y axes refer to rectangularity and elongation respectively. It can be observed that elongation plays a dominant role to discriminate motorcycles from others. Fig. 3c shows the 2D space with dimensions convexity and rectangularity. Rectangularity is not adequate to discriminate cars from vans. With the help of convexity and elongation, car/van classification becomes more accurate.

V. HOG-Based Classification

Alternative to shape-based classification, one can employ HOG+SVM classification. Normally, HOG-based methods require sliding window approach as explained in Section I; however it requires huge amount of computation time which makes the algorithm useless in real-time classification tasks. To bring a practical solution to this problem, we make use of the silhouettes extracted in the background subtraction step (Section II). The classification can be performed for the region defined by the extracted silhouette. Here, classification of an instance refers to obtaining SVM scores for each object type and selecting the class with the highest score.

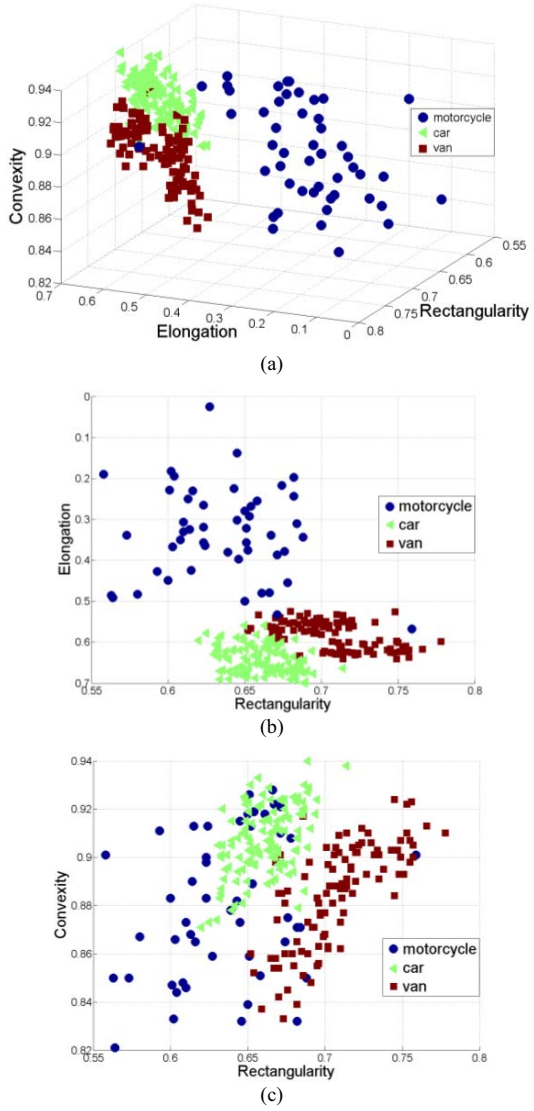

Fig.3: Extracted features of the annotated silhouettes. (a) All three dimensions. (b) First two dimensions. (c) Last two dimensions.

To generate SVM scores on omnidirectional images we use object models trained with standard (perspective) images of the object class. In this study we worked on car, van and motorcycle classes. We trained our car model using UIUC [23] and Darmstadt [24] sets together totaling 602 car side views. The size of the model is 40x100 pixels (height x width). The model trained for van detection is 40x100 as well. For this object type, we constructed a database of 104 images containing vans viewed from either side. The model for motorcycle is 128x96 and trained with 48 images of our own.

Contrary to [17], we do not take rectangular windows in the omnidirectional image but we modify the image gradients. Actually, we follow the approach that we proposed in [4]. The operations that we perform can be divided into two steps:

1. Modification of gradient magnitudes using Riemannian metric. With this metric, we can convert the gradient magnitudes in the omnidirectional image to the ones that would occur if our image was not warped. In particular, norm of the gradient reads

$$|\nabla_{S^2} I|^2 = \frac{(4+x^2+y^2)^2}{16} |\nabla_{\mathbb{R}^2} I|^2 \quad (1)$$

for our omnidirectional camera which is a catadioptric camera with a paraboloidal mirror. $\mathbb{R}^2$ represents the image domain and S^2 represents the unwarped sphere domain. We see that the modified gradients are just the scaled versions of the gradients in the image. So the gradients are multiplied according to their positions (x, y) before further process.

2. Conversion of gradient orientations to form an omnidirectional window. In addition to the gradient magnitudes, gradients orientations should be computed as if a perspective camera is looking from the same viewpoint. Fig. 4a shows a half of a synthetic omnidirectional image (400x400 pixels) where the walls of a room are covered with rectangular black and white tiles. Conventional HOG result of the marked region (128x196 pixels) in this image is given in Fig. 4b where gradient orientations are in accordance with the image. However, since these are vertical and horizontal edges in real world, we need to obtain vertical and horizontal gradients. Fig. 4d shows converted gradients for the region marked in Fig. 4c, which is an example of the modified HOG computation. For conversion of gradient orientations and magnitudes, we employ bilinear interpolation with backward mapping.

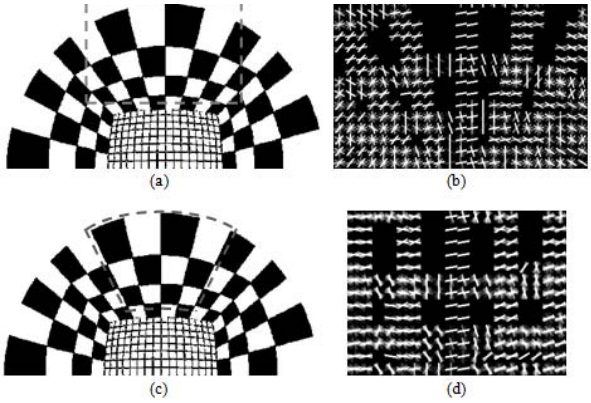

Fig. 4: Description of how the gradients are modified for an omnidirectional sliding window. Result in (b) is the regular HOG computed for the region marked with dashed lines in (a). Modified HOG computation gives the result in (d) for the region marked in (c). Vertical and horizontal edges in real world produce vertical and horizontal gradients in the modified version.

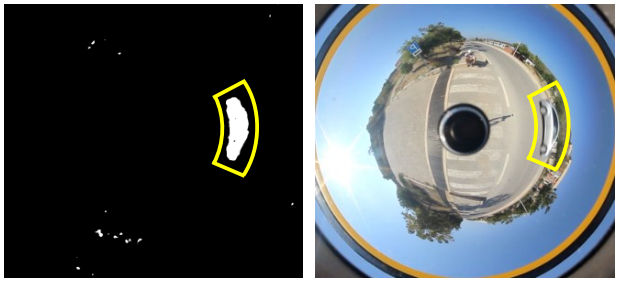

Fig 5: a) Binary image of a video frame when object is closest to 0°. A modified HOG computation window (yellow colored doughnut slice) is placed around the foreground silhouette. b) RGB image of the same frame. The modified HOG computation window is superimposed on the image.

The process described above is done for each sample in our dataset. Centroid of the silhouette in the omnidirectional image frame is obtained. A modified HOG window is placed at this centroid. While doing this a margin is left around the silhouette in accordance with the images in the training set. An example is shown in Fig. 5a. Same HOG window is placed on the video frame (Fig. 5b). This is the region from which the HOG features are computed. If we used an unmodified HOG window, then the region would be a rectangular one.

As mentioned before, HOG-based classification of an instance refers to obtaining SVM scores for each object type and selecting the class with the highest score. For the sample shown in Fig. 5, SVM scores obtained for car, van and motorcycle models are 1.28, 0.72 and -0.33 respectively. With these scores, classifier chooses the car class which is correct for this sample. The results for the complete dataset are presented in Section VII.

VI. Combined Classification

For each sample in the test set, we convert the obtained scores for each classifier to probabilities of belonging to the classes. Let $P(w_k|\mathbf{x}_{shape}^{(i)})$ denote the probability of i^{th} sample belonging to class k with shape-based features using kNN. For each sample, the sum of probabilities belonging to any of the three classes is equal to 1:

$$P\left(w_{car}\middle|\mathbf{x}_{shape}^{(i)}\right)+P\left(w_{van}\middle|\mathbf{x}_{shape}^{(i)}\right)+P\left(w_{motor}\middle|\mathbf{x}_{shape}^{(i)}\right)=1 \quad (5)$$

We compute the probabilities also for HOG+SVM classification. Then, we apply decision-level fusion of these two classifiers. Different combination schemes can be employed. According to the comprehensive study of Kittler *et al*. [25], two fundamental rules are the sum rule and the product rule. Sum rule is very intuitive and corresponds to choosing the class that maximizes the sum of probabilities:

$$\max_k \; P\left(w_k|\mathbf{x}_{shape}^{(i)}\right)+P\left(w_k|\mathbf{x}_{HOG}^{(i)}\right) \quad (6)$$

In this way, we guarantee that the selected class has a good support from both classifiers. Equation (6) assumed an equal weighting for shape-based and HOG-based classification, however different weights can also be applied:

$$\max_k \; \alpha\cdot P\left(w_k|\mathbf{x}_{shape}^{(i)}\right)+(1-\alpha)\cdot P\left(w_k|\mathbf{x}_{HOG}^{(i)}\right) \quad (7)$$

The product rule assumes that the classifiers are conditionally independent:

$$\max_k \; P\left(w_k|\mathbf{x}_{shape}^{(i)}\right)\cdot P\left(w_k|\mathbf{x}_{HOG}^{(i)}\right) \quad (8)$$

The results of these three schemes are presented in Experiments section. For weighted sum rule (7), α that gives the best performance on the training set is chosen.

VII. Experiments

Using a Canon 600D SLR camera and a mirror apparatus (www.gopano.com) we obtained a catadioptric omnidirectional camera. We constructed a video dataset of 124 cars, 104 vans and 48 motorcycles totaling 276 vehicle instances. The video resolution is 810x1440 pixels. Dataset is divided into training and test sets equally, i.e. the training set contains 50% percent of the total dataset. To obtain a more generalized result, we repeated the experiment by shifting the training and test sets and present the average result in the tables below.

Since we aim a fast method for classification, total processing time which includes background subtraction, feature extraction and classification should be limited with a few seconds. That can be considered as the time before another vehicle comes and becomes the largest silhouette (moving object) in the scene. The two classification approaches are implemented taking this aspect into account. For the shape-based classifier we use multiple silhouettes but we limit the angle range otherwise computation time exceeds bearable limits (we give time analysis in the next paragraph). Similarly for the HOG-based classifier, instead of sliding windows we fit a modified HOG computation window to the location of the foreground silhouette as explained in Section V.

A. Shape-based Classification Results

Table 1 summarizes the shape-based classification results where K is taken as 15 for kNN classification. Results are shown separately when the angle range for averaging silhouettes is set to ± 3 and ± 5 degrees. The results with $\pm 5°$ is slightly better than the results with $\pm 3°$. Although further increasing the angle range has the potential of improving the accuracy, this results in an unacceptable amount of computation time. With $\pm 3°$ angle range, we process an average of three frames per video sample. For 1 megapixel video frame, which is close to ones we are using, the background subtraction takes a little less than 2 seconds with the method we are using [20]. Silhouette averaging for three frames and classification of the average silhouette takes around 0.5 seconds. Thus, total processing for three frames takes approximately 6 seconds. This time is obtained with a desktop PC (Intel Pentium Core i7 3.4 GHz processor) where the codes are written in C/C++. Since spending more time is not reasonable, in the combined classification experiments given in Section VII.C, we take the angle range as $\pm 3°$ for the shape-based classifier.

Processing time may be shortened with using faster background subtraction algorithms. For instance, Adaptive Background Learning takes a quarter of the time spent by the method we use [20]. Also, the resolution of the video can be decreased for faster processing. In this way, number of silhouettes used in 'silhouette averaging' method can be increased. However, in our experiments we have seen that the silhouettes extracted with low quality background subtraction and lower resolution are very poor that the accuracy loss is much higher than the improvement which can be gained with more silhouettes.

Table 1: Shape-based classification accuracies for averaging silhouettes in ± 3 and ± 5 degrees. Numbers are average of two experiments where the test and training sets are interchanged. K=15 in kNN classification. Overall accuracy is obtained by weighting each class accuracy with no. of samples in that class.

Actual class:	car	van	motorcycle	overall
Accuracy ($\pm 5°$):	93.2%	90.0%	87.5%	91.0%
Accuracy ($\pm 3°$):	91.5%	86.3%	86.9%	88.7%

Table 2: HOG-based classification accuracies. Numbers are average of two experiments where the test and training sets are interchanged. Overall accuracy is obtained by weighting each class accuracy with no. of samples in that class.

Actual class:	car	van	motorcycle	overall
Accuracy:	93.0%	88.3%	89.2%	90.6%

Table 3: Combined classification results. Numbers are average accuracies of two experiments where the test and training sets are switched. Overall accuracy is obtained by weighting each class accuracy with no. of samples in that class.

Actual class:	car	van	motorcycle	overall
Sum rule:	93.2%	92.2%	95.7%	93.3%
Sum rule(α=0.3):	94.9%	98.0%	97.8%	96.5%
Product rule:	91.5%	94.2%	91.3%	92.5%

B. *HOG-based Classification Results*

Table 2 summarizes the HOG-based classification results where the class with the highest SVM score is chosen by the algorithm. We observe that the accuracies for the three classes are not much different from each other with an overall classification accuracy of 90.6%.

C. *Combined Classification Results*

Table 3 summarizes the accuracies obtained with combined classification schemes explained in Section VI. In overall, we can say that combining classifiers both with product rule and sum rule increases the performance since the overall accuracies are higher than the ones in Table 1 and Table 2. The improvement with the product rule is lower than the one obtained with the sum rule. This result is in accordance with [25], where the sum rule outperformed the product rule. The improvement with weighted sum rule is distinctive. Here, performance on the training samples is used to choose α.

Obtaining HOG scores takes about 0.1 sec. per class most of which is spent for converting gradients for the modified HOG window. For three vehicle classes, as in our experiments, three different scores are obtained which makes the total time 0.3 seconds. Therefore combined classification does not increase the time spent by shape-based classification much.

VIII. CONCLUSIONS

We proposed using shape-based and gradient-based techniques together for the classification of vehicles. For shape-based classification, we extracted features from the silhouettes in the omnidirectional video frames, which are obtained after background subtraction. Classification is performed with kNN (k Nearest Neighbors). For gradient-based classification, we employ HOG features and SVM classifier. We analyzed the performances of shape-based gradient-based and combined classifiers separately. Results show that the combined classifier is superior to the individual classifiers and the best combination is obtained with sum rule.

Shape-based vehicle classification algorithms are preferred due to their speed since the feature extraction step is computationally cheaper. In this work, we showed that once the foreground is extracted, HOG-based classification is also affordable and a classifier that takes into account both shape-based and HOG-based classification is more powerful.

ACKNOWLEDGMENT

This work is supported by TUBITAK (Project No 113E107).

REFERENCES

[1] N. Dalal, B. Triggs, “Histograms of Oriented Gradients for Human Detection”, Computer Vision and Pattern Recognition (CVPR), 2005.

[2] P. Felzenszwalb, D. McAllester, D. Ramanan, “A Discriminatively Trained, Multiscale, Deformable Part Model,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[3] T. Gandhi, M. Trivedi, “Video based surround vehicle detection, classification and logging from moving platforms: Issues and approaches”, IEEE Intelligent Vehicles Symposium, 2007.

[4] I. Cinaroglu, Y. Bastanlar, “A direct approach for object detection with catadioptric omnidirectional cameras”, Signal, Image and Video Processing, 2015, DOI: 10.1007/s11760-015-0768-2.

[5] H. Amine Iraqui, Y. Dupuis, R. Boutteau, J. Ertaud, X. Savatier, “Fusion of omnidirectional and ptz cameras for face detection and tracking,” Int. Conference on Emerging Security Technologies (EST), 2010.

[6] H.C. Karaimer, Y. Bastanlar, “Car detection with omnidirectional cameras using Haar-like features and cascaded boosting”, IEEE Conf. on Signal Processing and Communications Applications (SIU), 2014.

[7] B. Morris, M. Trivedi, “Improved vehicle classification in long video by cooperating tracker and classifier modules”, IEEE International Conference on Video and Signal Based Surveillance (AVSS), 2006.

[8] B. Morris, M. Trivedi, “Robust classification and tracking of vehicles in video streams”, IEEE Intelligent Transportation Sys. Conf. (ITSC),2006.

[9] Y. Bastanlar, “User behaviour in web-based interactive virtual tours”, Int. Conf. on Information Technology Interfaces (ITI), 2007.

[10] T. Goedeme, M. Nuttin, T. Tuytelaars, L. van Gool, “Omnidirectional Vision Based Topological Navigation”, International Journal of Computer Vision (IJCV), vol. 74(3), pp.219–236, 2007.

[11] Y. Bastanlar, A. Temizel, Y. Yardimci, P. Sturm, “Multi-view Structure-from-Motion for Hybrid Camera Scenarios”, Image and Vision Computing, vol. 30(8), pp.557–572, 2012.

[12] Q. Luo, T. Khoshgoftaar, A. Folleco, “Classification of ships in sur-veillance video”, Int. Conf. on Information Reuse and Integration, 2006.

[13] N. Mithun, N. Rashid, S. Rahman, “Detection and classification of vehicles from video using multiple time-spatial images”. IEEE Trans. on Intelligent Transportation Systems, 13, pp.1215–1225, 2012.

[14] S. Gupte, O. Masoud, R. Martin, N. Papanikolopoulos, “Detection and classification of vehicles”, IEEE Transactions on Intelligent Transportation Systems, 3, pp.37–47, 2002.

[15] P. Kumar, S. Ranganath, H. Weimin, K. Sengupta, “Framework for real-time behavior interpretation from video”. IEEE Transactions on Intelligent Transportation Systems, 6, pp.43–53, 2005.

[16] N. Buch, J. Orwell, S. Velastin, “Detection and classification of vehicles for urban scenes”, Int. Conf. on Visual Information Engineering, 2008.

[17] Y. Tang, Y. Li, T. Bai, X. Zhou, “Human Tracking in Thermal Catadioptric Omnidirectional Vision”, International Conference on Information and Automation (ICIA), 2011.

[18] R. Khoshabeh, T. Gandhi, M. Trivedi, “Multi-camera based flow characterization and classification”, Intell.Trans.Sys.Conf. (ITSC),2007.

[19] H.C. Karaimer, Y. Bastanlar, “Detection and Classification of Vehicles from Omnidirectional Videos using Temporal Average of Silhouettes”, Int. Conf. on Computer Vision Theory and Applications(VISAPP),2015.

[20] J. Yao, J. Odobez, “Multi-layer background subtraction based on color and texture”, Computer Vision and Pattern Recognition (CVPR), 2007.

[21] A. Sobral, A. Vacavant, “A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos”, Computer Vision and Image Understanding, 122, pp.4 – 21, 2014.

[22] M. Yang, K. Kpalma, J. Ronsin, “A survey of shape feature extraction techniques”, Pattern recognition, pp.43–90, 2008.

[23] S. Agarwal, D. Roth, “Learning a sparse representation for object detection”, European Conference on Computer Vision (ECCV), 2002.

[24] B. Leibe, A. Leonardis, B. Schiele, “Combined object categorization and segmentation with an implicit shape model”, Proc. of the Workshop on Statistical Learning in Computer Vision, 2004.

[25] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, “On Combining Classifiers”, IEEE Transactions on PAMI, 20(3), pp.226-239, 1998.