

Classification and Tracking of Traffic Scene Objects with Hybrid Camera Systems

Ipek Baris, Yalin Bastanlar

Computer Vision Research Group, Department of Computer Engineering,
Izmir Institute of Technology, Izmir, Turkey
{ipekbaris, yalinbastanlar}@iyte.edu.tr

Abstract—In a hybrid camera system combining an omnidirectional and a Pan-Tilt-Zoom (PTZ) camera, the omnidirectional camera provides 360 degree horizontal field-of-view, whereas the PTZ camera provides high resolution at a certain direction. This results in a wide field-of-view and high resolution camera system. In this paper, we exploit this hybrid system for real-time object classification and tracking for traffic scenes. The omnidirectional camera detects the moving objects and performs an initial classification using shape-based features. Concurrently, the PTZ camera classifies the objects using high resolution frames and Histogram of Oriented Gradients (HOG) features. PTZ camera also performs high-resolution tracking for the objects classified as the target class by the omnidirectional camera. The object types we worked on are pedestrian, motorcycle, car and van. Extensive experiments were conducted to compare the classification accuracy of the hybrid system with single camera alternatives.

Keywords—*vehicle classification, vehicle detection, object detection, object tracking, hybrid camera system, omnidirectional camera, pan-tilt-zoom camera.*

I. INTRODUCTION

An omnidirectional camera provides 360° field of view in the horizontal plane with a low resolution per viewing angle. A PTZ (Pan-Tilt-Zoom) camera, however, provides high resolution at its viewing direction. A hybrid system combining the powerful aspects of both camera types can provide a wide angle and high resolution surveillance. This powerful combination is used in many application areas including robot navigation [1], 3D reconstruction [2] and surveillance. A typical surveillance task employing a hybrid camera system is to point the PTZ camera towards the position of the moving object detected by the omnidirectional camera [3-6]. In [3], this scheme is described for tracking general objects where no additional operation on the PTZ camera is performed. In [4], hybrid system is used for a two-level face detection method. Face detection algorithm initially runs on the panoramic image which is the unwarped omnidirectional image. Then, the PTZ camera is directed to the marked regions and it performs a second face detection. A similar study [5] goes one step further and tries to keep the detected face in the middle of PTZ camera frame during tracking and zooms when necessary.

In our study, we not only direct the PTZ camera towards the detected object but we also perform object classification with the PTZ camera in the aim of improving the initial classification. Among previous studies, the closest one to the proposed system is [6] where objects are classified as large or

small vehicles. When compared to that work, we employ a larger set of shape-based features and identify multiple classes (pedestrian and three vehicle types) instead of just size categorization. The reported classification accuracies in that work are 80.5% and 87.8% for omnidirectional and PTZ cameras respectively. This is below what we achieved with our system (97.5% and 99.2%). Furthermore, we actively use the zoom function of the PTZ camera to be capable of high-resolution recording.

Before giving the details of our method, let us briefly present more related work on vehicle classification. The most popular approach is to employ shape-based features which are computed using the blobs extracted by background subtraction. Since vehicles are moving objects and their shapes can be easily segmented, this approach was commonly used in previous studies. For instance, in [7], authors created a feature vector consisting of area, breadth, compactness, elongation, perimeter, length, axes of fitted ellipse, centroid and five image moments from the blob tracked with Kalman Filter. Then, vehicle classification is performed within Sedan, Semi, Truck+SUV+Van classes with weighted k-Nearest Neighbors (kNN). Kumar et al. [8] proposed to use a Bayesian network for classification with size, shape, velocity, and position of the vehicles. Buch et al. [9] conducted vehicle detection and classification using 3D models which are projected onto the image plane for obtaining a silhouette match measure. Regarding shape-based classification work with omnidirectional cameras, Khoshabeh et al. [6] used only the area of the vehicle foreground blob to label the vehicles as large or small. Karaimer et al. [10,26] extracted shape-based features, namely convexity, elongation, rectangularity and Hu moments, from the largest blobs and classified them as motorcycle, car or van with SVM and kNN.

Gradient-based features are also used for object classification. A seminal research of this approach was presented by Dalal and Triggs [11], who performed human detection using Histogram of Oriented Gradients (HOG) features extracted from sliding windows. Samples were classified with linear Support Vector Machines (SVM). Since then, HOG features have been used for vehicle classification as well. A recent example can be seen in [12]. An example of the HOG+SVM approach employed for omnidirectional cameras is given in [13], where authors modified HOG computation for the omnidirectional camera geometry and detected car, van and pedestrian. Two other studies with gradient-based features are computing HOG features on manually selected virtual perspective views generated from the omnidirectional camera

[14] and extracting HOG features from the bounding rectangles obtained by background subtraction [15]. Extracting HOG features for omnidirectional cameras is computationally heavier than perspective cameras. Aiming real-time processing, we used shape-based features for our omnidirectional camera and HOG features are computed only for PTZ camera frames.

The paper is organized as follows. In Section II, the proposed hybrid system is described. In Section III, the design and the results of the experiments are presented. In Section IV, the conclusions are given.

II. THE PROPOSED HYBRID SYSTEM

In the proposed system, the omnidirectional camera is responsible of detection and initial classification (Section II.A). The PTZ camera is responsible of a second classification (Section II.B) which leads to improved hybrid classification. PTZ camera is also capable of high-resolution tracking for the objects detected by the omnidirectional camera (Section II.C).

A. Object Detection and Classification with the Omnidirectional Camera

The job performed by the omnidirectional camera is illustrated in Fig. 1. The first stage is extracting the blobs (silhouettes) belonging to the moving objects in the scene. For this task, we employed Adaptive Background Learning (ABL) algorithm which is one of the fastest background subtraction algorithms [16]. Next, noise-like particles around the silhouettes are cleaned by morphological operations. Also, blobs smaller than an area threshold are discarded to eliminate irrelevant objects such as cats and dogs.

The following step consists of tracking the blobs with Kalman Filter [17] and association between the blobs in current frame and previously detected blobs by using Hungarian Algorithm [18,19]. 2D position (object centroid) and velocity are predicted with Kalman Filter. Hungarian Algorithm finds detection-track pairs with minimum cost which is calculated as the Euclidean distance between the centroid of the detection and the associated track. Concurrently, the system checks whether the object is about to pass through the classification region. The classification region is defined as the angle range $[-30^\circ, +30^\circ]$ where 0° is assigned to the direction that camera is perpendicular to the road (Fig. 2a). If the object enters this region, system signals the PTZ camera to start detection and classification.

Classification (final step) is also performed in this region. Following the extraction of shape-based features, the detected blob is classified using kNN classifier. As long as the object is inside the predefined angle range, each blob belonging to that object is classified and the assigned class is voted. Once it passes the range, the most voted class is selected as the object's final class. A classification example is given in Fig. 2. The shape-based features we used are area, elongation and height-width ratio of the silhouettes. Area is able to separate small vehicles from the large ones. Elongation is calculated by (1) [20] where S and L are the smallest and largest edges of the minimum rotated rectangle covering the object. It discriminates motorcycles and pedestrians from the other classes. Especially motorcycle is the least elongated object type.

$$\text{Elongation} = 1 - S/L \quad (1)$$

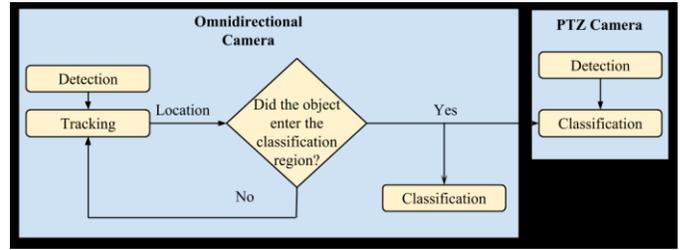


Fig. 1: Steps of the classification task, performed by the hybrid system.

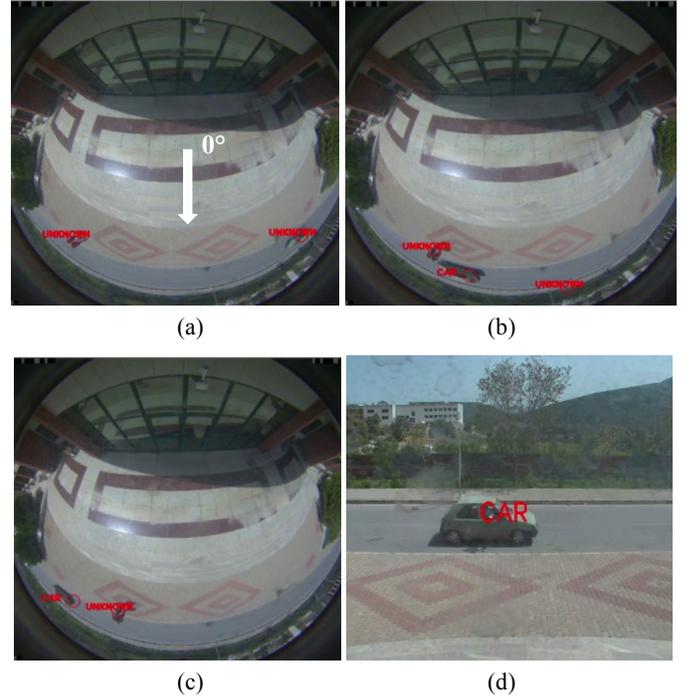


Fig. 2: Example of hybrid system classification. (a) The objects are not in the angle range of the classification region yet, they are labeled as 'UNKNOWN'. (b) Classification of the car is done since it passed through the angle range. (c) The tracking of the classified object is continued with its new label 'CAR'. (d) The result of PTZ camera classification. Due to the delay in the PTZ camera, this frame was captured after the object was classified by the omnidirectional camera. The classification result of omnidirectional camera can be used for the PTZ camera classification while choosing the SVM classifier.

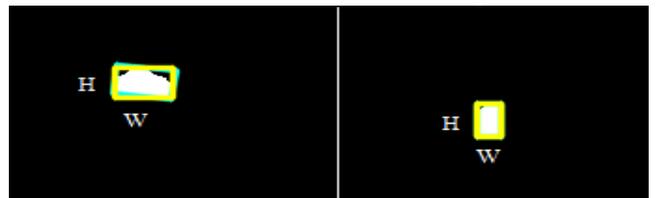


Fig. 3: Silhouette examples of a motorcycle (left) and a pedestrian (right) whose elongation values are close to each other (0.42 and 0.41, respectively) but height-width ratios are quite different (0.65 for motorcycle, 1.63 for pedestrian). H and W denote the height and width of the non-rotated bounding rectangle as given in (2).

Height-width ratio, R , is calculated by using (2) where H and W are the height and width of the minimum bounding rectangle, respectively.

$$R = H/W \quad (2)$$

There are cases that motorcycle and pedestrian have close elongation values (Fig. 3 shows an example), where height-width ratio distinguishes between motorcycle and pedestrian.

Fig. 4 shows the features of the annotated silhouettes of all samples in our training dataset for shape-based classification. For better visualization, pairs of normalized features are plotted in 2D. As can be observed, elongations of car and van samples and areas of pedestrian and motorcycle samples are similar, but using all three features enables us to separate each class from the others.

B. Object Detection and Classification with the PTZ Camera

Triggered by the omnidirectional camera (when the object enters the classification region), PTZ camera performs a second classification using HOG features. PTZ camera uses the side-views of the objects for extracting HOG features since the training is performed with images seeing the vehicles from one side. To precisely locate the bounding rectangle of the object in the PTZ camera frame, moving object is detected with background subtraction in the PTZ camera as well. Improved Mixture of Gaussians (MOG2) background subtraction algorithm [21] is used in this stage of our system. We have applied ABL and MOG2 for the same objects in the PTZ camera and observed that MOG2 is better for noise elimination and resultant silhouettes cover the entire object. Examples obtained with both background subtraction algorithms are given in Fig. 5.

Stages of vehicle classification on the PTZ camera are shown in Fig. 6. First stage extracts the silhouette which is the largest blob in PTZ frame. Second stage finds the minimum bounding rectangle and calculates the height-width ratio. Then, an SVM decider uses height-width ratio and chooses the classifier to which the object is sent (low height-width ratio indicates that it's a car or a van). After the classifier is selected, the rectangle is enlarged to fit into the defined aspect ratio (1:1 for pedestrian/motorcycle, 1:2.5 for car/van) and the region is rescaled to the working resolution (120x120 or 120x300 pixels). Finally, the object is classified by using SVM with HOG features computed on the region. Two SVM classifiers are trained. Pedestrian/motorcycle SVM is trained with 1764 HOG features from 120x120 (1:1 ratio) pixels detection window and it discriminates pedestrians from motorcycles. Car/van SVM is trained with 4788 HOG features from 120x300 (1:2.5 ratio) window to discriminate cars from vans.

The classification described above is independent of the omnidirectional camera and as we will see in Section III.A its performance is not good since height-width ratio of the bounding rectangle may choose wrong type of SVM classifier. Thus, we propose a hybrid classification scheme, where the final class detected with the omnidirectional camera is used to choose the SVM classifier. In this way, we merge the shape-based classification result with HOG-based classification and this improves the performance as will be discussed in Section III.B. A visual example is given in Fig. 2d.

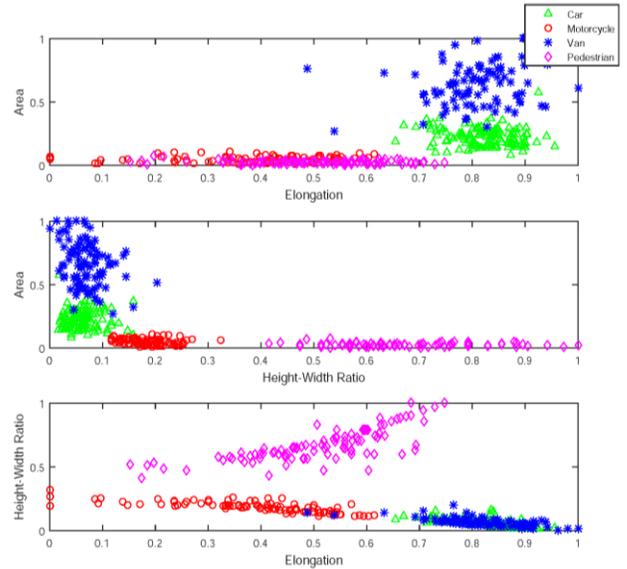


Fig. 4: Normalized shape based features of samples in our dataset



Fig. 5: The silhouettes at the left column are extracted by ABL, the ones at the right are extracted by MOG2 background subtraction algorithm [21].

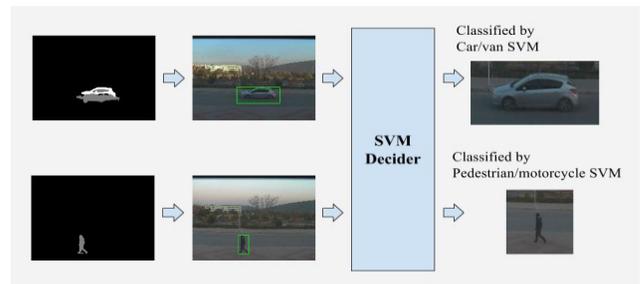


Fig. 6: Operations done with PTZ camera. Minimum bounding rectangle enclosing the silhouette is obtained and the SVM decider chooses the classifier (car/van or pedestrian/motorcycle). To make this choice SVM decider uses either the height-width ratio of the bounding rectangle (PTZ-only classification) or the final class detected with omnidirectional camera (hybrid classification). The rectangle is enlarged and the window is cropped (aspect ratio is 1:1 for motorcycle/pedestrian, 1:2.5 for van/car). The classification of the object is performed based on HOG features.

C. High-resolution Object Tracking

PTZ camera is able to track only one object at a time, whereas several moving objects may exist in the scene. Here, it is reasonable to assume that there is a target class (predefined by the user) and the PTZ camera is directed only to the objects that are classified as the target class by the omnidirectional camera. When an object is classified as the target class, to direct the PTZ camera, pan and tilt values are calculated using the location of the object and zoom value is determined by the area of the object in the omnidirectional image. Meanwhile, the omnidirectional camera continues detection and classification tasks for the whole scene. In this way, wide-angle surveillance and high-resolution recording for the desired objects are performed simultaneously.

Even though the position estimate of Kalman Filter is able to track a detected vehicle on the omnidirectional camera, it is not sufficient to catch the vehicle due to the delays occurred during steering the PTZ camera. To solve this problem, the estimated position is modified using (3) where (X_{PTZ}, Y_{PTZ}) is the modified position, (X_K, Y_K) is the centroid predicted by Kalman Filter, $(\Delta X_K, \Delta Y_K)$ is the centroid displacement between consecutive frames and β is a variable which enables us to direct the PTZ camera ahead in the predicted course. β is defined when the object is in $[-120^\circ, +120^\circ]$. It gets its maximum value at 0° (cf. Fig. 2a) where the displacement is maximum for a fixed speed, and it linearly decreases from 0° to the boundaries.

$$X_{PTZ} = X_K + \Delta X_K \cdot (1 + \beta), \quad Y_{PTZ} = Y_K + \Delta Y_K \cdot (1 + \beta) \quad (3)$$

To direct the PTZ camera to the object of interest detected in the omnidirectional camera, the geometric relation between the cameras should be extracted. One method is performing external calibration of the hybrid system. This can be done either by a complete external calibration procedure using calibration patterns [22] or a simplified procedure where some assumptions about the camera setup are made [23]. Another major approach, called as spatial mapping, is based on data collection and data fitting (interpolation) [24]. In [25], only the tilt angle of the PTZ camera is estimated by spatial mapping. We estimate both pan and tilt angles by spatial mapping.

In spatial mapping method, it is assumed that the pan and tilt angles of the PTZ camera are highly correlated with the corresponding pixel coordinates in the omnidirectional image. We put a grid pattern on the omnidirectional image plane shown in Fig. 7 and for each point on the grid; we collected corresponding pan and tilt angles of the PTZ camera, i.e. the angles when the PTZ camera is directed to that grid point. A look-up table is constructed, for the locations between grid points, pan/tilt values are estimated with bilinear interpolation.

To obtain high resolution images of an object, we need to zoom in when the object covers a small part of the PTZ camera frame. We relate the size of the silhouette in the omnidirectional image to the zoom value of PTZ camera by collecting samples of an object and moving it closer and far away from the camera. In a sense, we perform spatial mapping for the zoom value. As a result, red circles in Fig. 8 are obtained. To satisfy those points, we derived (4) which is depicted in Fig. 8 as a line.

$$zoom = \begin{cases} maxzoom & \text{if } x < area_threshold \\ \max\left(1, maxzoom \times \left(1 - \frac{x - area_threshold}{1000 - area_threshold}\right)\right) & \text{if } x \geq area_threshold \end{cases} \quad (4)$$

In this equation, maximum value used for the zoom parameter of the PTZ camera is denoted by $maxzoom$, which corresponds to 24x zoom in our case. If the area of the bounding rectangle of the object is smaller than $area_threshold$ (in our case it is 200 pixels), then we do not zoom in more and use the $maxzoom$ value. If the area is larger, then $zoom$ is computed. The computed zoom value cannot be smaller than 1 since 1x is the widest viewing angle.

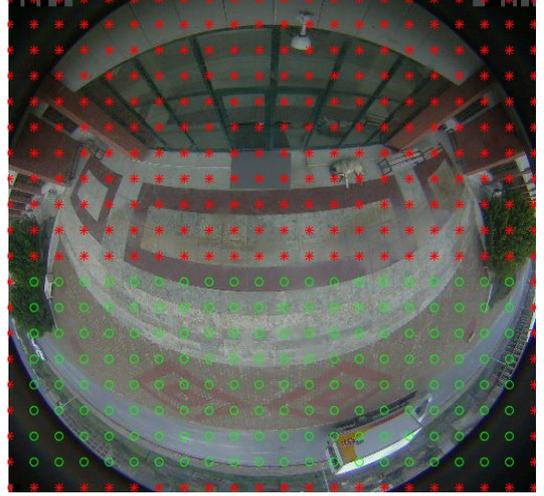


Fig. 7: Grid points of the omnidirectional camera used for spatial mapping. Red dots are not seen from PTZ camera or not in the region of our interest. Constructed look-up table contains the pan/tilt pairs of the green circles.

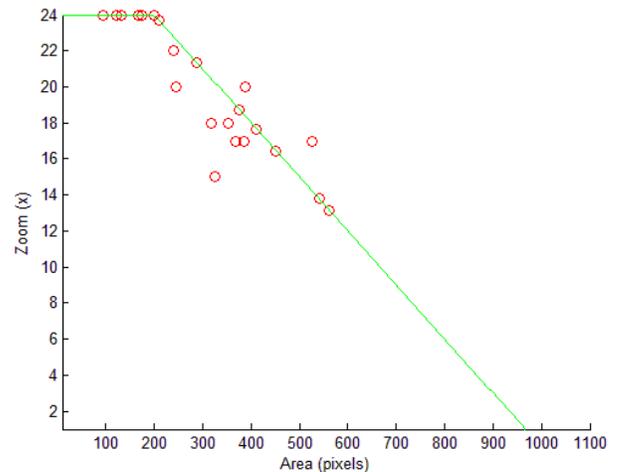


Fig. 8: Calculation of the zoom parameter of the PTZ camera. X axis is the area of the silhouette's bounding rectangle (in pixels) in the omnidirectional camera, y axis is the zoom value of the PTZ camera (from 1x to 24x zoom). In our camera, 1x refers to 3.5mm focal length and 24x refers to 84 mm focal length. Red circles were obtained by experiment; green lines were obtained by (4).

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

We used an Oncam Grandeye 360° with fisheye lens as the omnidirectional camera and Samsung SNP-3500 as the PTZ

camera. The cameras were mounted at the front side of a building within the university campus. The resolutions of the omnidirectional and PTZ cameras were set as 528x480 and 1024x768 pixels, respectively. Higher resolutions prevented us from working real-time. We also employed OpenCv 3.0 Library with CUDA to achieve real-time performance. The experimental results are summarized under three subsections which are classification with the omnidirectional camera, classification with the hybrid system and high-resolution object tracking.

A. Classification with the Omnidirectional Camera

We collected 96 motorcycle, 125 car, 100 van and 102 pedestrian samples as the training set for kNN classification. Another set was constructed for testing (94 vans, 113 cars, 71 motorcycles and 83 pedestrians). An important property of this dataset is that the same vehicles (and pedestrians) are also captured with the PTZ camera. In this way, we are able to test the classification accuracy of the hybrid system for the same samples. After the features are extracted, samples in the test set were classified with kNN (cf. Section II.A). Table 1 shows the confusion matrix, per class accuracies and the overall accuracy (97.5%) when $k = 5$.

B. Classification with the Hybrid System

As detailed in Section II.B, PTZ camera classifies the objects using HOG features and SVM classifier. For the training set samples, we manually labeled the bounding boxes of objects in the PTZ camera frame. Samples of cars and vans were scaled to 120x300 pixels (1:2.5 ratio), others were scaled to 120x120 pixels (1:1 ratio) and then their HOG features were computed. The dataset was augmented 12 times by shifting the bounding boxes left, right, up, and down and by zooming in and out. This provides a more robust training, since in the test phase the bounding boxes after background subtraction may not be at the ideal position and scale within the PTZ camera frame.

Table 2 gives the performance of the PTZ camera classification when the SVM classifier choice (car/van SVM or pedestrian/motorcycle SVM) is made by using the height-width ratio of the bounding box detected in the PTZ camera. We examined the training set samples and determined 0.65 as the height-width ratio threshold which yields good performance in separation of motorcycles/pedestrians from cars/vans. In the table, we see that approximately 91% of samples were classified correctly. This is lower than the omnidirectional camera classification accuracy given in Table 1. The majority of the misclassified samples are the ones sent to the wrong SVM classifier. Therefore, the choice of SVM classifier made by height-width ratio is not good enough.

Next, we present the results of the hybrid classification where the SVM classifier is selected using the classification result of the omnidirectional camera. In other words, if the kNN classification result of an object is a car or a van, HOG features of that object in the PTZ camera are evaluated by car/van SVM classifier. Table 3 presents the results where the overall accuracy is increased to 99% and samples choosing wrong SVM classifier are completely eliminated. This indicates that the performance of the hybrid system is higher than the omnidirectional-only or PTZ-only approaches.

TABLE I: CONFUSION MATRIX OF THE EXPERIMENT WITH TEST DATA AND ACCURACY OF CLASSIFIER ON OMNIDIRECTIONAL CAMERA.

		# samples	Predicted Class				Accuracy
			Van	Car	Motorcycle	Pedestrian	
Real Class	Van	94	90	4	0	0	95.74%
	Car	113	1	112	0	0	99.12%
	Motorcycle	71	0	0	71	0	100%
	Pedestrian	83	0	0	4	79	95.18%
Classifier Accuracy							97.51%

TABLE II: ACCURACY OF THE PTZ CAMERA CLASSIFICATION WHEN SVM IS SELECTED ACCORDING TO THE HEIGHT-WIDTH RATIO OF OBJECT BOUNDING RECTANGLE ON THE PTZ CAMERA FRAME.

Class	# samples	# samples choosing wrong SVM	PTZ Classifier		Accuracy
			True Classification	False Classification	
Van	94	9	83	2	88.30%
Car	113	7	106	0	93.81%
Motorcycle	71	13	57	1	80.28%
Pedestrian	83	0	83	0	100%
Classifier Accuracy					91.14%

TABLE III: ACCURACY OF THE HYBRID CLASSIFICATION WHEN SVM IS SELECTED ACCORDING TO THE CLASSIFICATION RESULT FROM THE OMNIDIRECTIONAL CAMERA.

Class	# samples	# samples choosing wrong SVM	PTZ Classifier		Accuracy
			True Classification	False Classification	
Van	94	0	92	2	97.87%
Car	113	0	113	0	100%
Motorcycle	71	0	70	1	98.59%
Pedestrian	83	0	83	0	100%
Classifier Accuracy					99.17%

C. Object Tracking with the Hybrid System

We tested the tracking module described in Section II.C by experiments. Fig. 9a shows a case when there is only one moving object in the scene, which was previously classified as ‘pedestrian’ and that object is going to be tracked until it leaves the scene. There may be more objects in the scene, one of them belonging to the target class. An example is given in Fig. 9b. In this case, the PTZ camera tracks the target object as other objects enter or leave the scene (Fig. 9c). If another target object is detected in the scene, PTZ camera keeps with the first one as long as its label in the omnidirectional camera stays valid. Object stays within the video frame and fills the frame properly. Our tracking experience revealed that while the PTZ camera is adjusting zoom, a delay occurs which could cause missing the fast objects such as cars. Since the PTZ camera is close to the road, fast vehicles pass through the scene in a few seconds and they can be tracked if we disable zoom-in functionality.

IV. CONCLUSIONS

We performed real-time classification and high-resolution recording with our hybrid camera system. The classification module consists of a step performed by the omnidirectional camera then another classification step by the PTZ camera. We observed that hybrid classification has higher performance than the single camera (omnidirectional-only or PTZ-only) approaches. In the tracking module, corresponding

pan/tilt/zoom values of the target object are computed and sent to the PTZ camera. In this way, while wide-angle detection and classification continues in the omnidirectional camera, high-resolution recording of objects is performed by the PTZ camera.

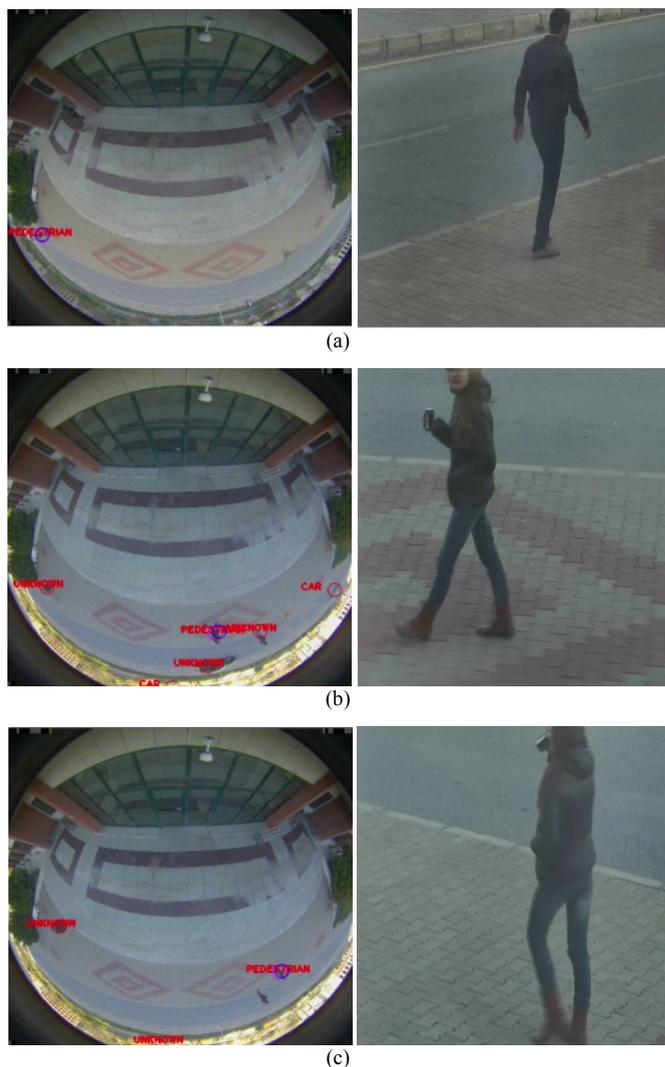


Fig. 9: Pedestrian tracking with the hybrid camera system. (a) There is only one moving object in the scene, which was previously classified as 'pedestrian' and being tracked. High-resolution capturing is being performed with the PTZ camera. (b) Pedestrian tracking when the omnidirectional camera detects multiple objects. (c) PTZ camera stays with the detected pedestrian in a later frame of the same scene.

ACKNOWLEDGMENT

This work is supported by the TUBITAK Project No: 113E107.

REFERENCES

[1] G. Adorni, M. Mordonini, S. Cagnoni and A. Sgorbissa, "Omnidirectional stereo systems for robot navigation". In *Computer Vision and Pattern Recognition Workshop*, 2003.
 [2] Y. Bastanlar, A. Temizel, Y. Yardimci and P. Sturm, "Multi-view structure-from-motion for hybrid camera scenarios", *Image and Vision Computing*, vol. 30(8), 2012, pp.557-572.

[3] J. Hu, S. Hu and Z. Sun, "A real time dual-camera surveillance system based on tracking-learning-detection algorithm". In: *Chinese Control and Decision Conference*, 2013, pp. 886-891.
 [4] S. Ghidoni, A. Pretto and E. Menegatti, "Cooperative tracking of moving objects and face detection with a dual camera sensor". *IEEE Int. Conference on Robotics and Automation*, 2010, pp. 2568-2573.
 [5] R. Khemmar, J.Y. Ertaud and X. Savatier, "Face Detection & Recognition based on Fusion of Omnidirectional & PTZ Vision Sensors and Heterogeneous Database". *International Journal of Computer Applications*, vol. 61, 2013, pp.35-44.
 [6] R. Khoshabeh, T. Gandhi and M. M. Trivedi, "Multi-camera based traffic flow characterization & classification". In: *IEEE Intelligent Transportation Systems Conference*, 2007, pp. 259-264.
 [7] B. Morris and M. Trivedi, "Robust classification and tracking of vehicles in traffic video streams". In: *IEEE Intelligent Transportation Systems Conference*, 2006, pp. 1078-1083.
 [8] P. Kumar, S. Ranganath, H. Weimin and K. Sengupta, "Framework for real-time behavior interpretation from traffic video". *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, 2005, pp.43-53.
 [9] N. Buch, J. Orwell and S.A. Velastin, "Detection and classification of vehicles for urban traffic scenes". In: *IET International Conference on Visual Information Engineering*, 2008, pp.182-187.
 [10] H.C. Karaimer and Y. Bastanlar, "Detection and classification of vehicles from omnidirectional videos using temporal average of silhouettes". In: *Computer Vision Theory and Applications*, 2015.
 [11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886-893.
 [12] S. Battiato, G.M. Farinella, A. Furnari, G. Puglisi, "An integrated system for vehicle tracking and classification". *Expert Systems with Applications*, vol. 42, 2015, pp.7263-7275.
 [13] I. Cinaroglu and Y. Bastanlar, "A direct approach for object detection with catadioptric omnidirectional cameras". *Signal, Image and Video Processing*, vol.10, 2016, pp.413-420.
 [14] T. Gandhi and M. Trivedi, "Video based surround vehicle detection, classification and logging from moving platforms: Issues and approaches", *IEEE Intelligent Vehicles Symposium*, 2007.
 [15] H.C. Karaimer, I. Cinaroglu and Y. Bastanlar, "Combining shape-based and gradient-based classifiers for vehicle classification". In: *IEEE Intelligent Transportation Systems Conference*, 2015, pp. 800-805.
 [16] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos". *Computer Vision and Image Understanding*, vol. 122, 2014, pp.4-21.
 [17] G. Welch and G. Bishop, *An introduction to the kalman filter*. University of North Carolina, Department of Computer Science Technical Report TR 95-041, 1995.
 [18] H.W. Kuhn, "The Hungarian method for the assignment problem". *Naval research logistics quarterly*, vol. 2, 1955, pp.83-97.
 [19] J. Munkres, "Algorithms for the assignment and transportation problems". *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, 1957; pp.32-38.
 [20] M. Yang, K. Kpalma and J. Ronsin, "A survey of shape feature extraction techniques". In: Yin PY, editor. *Pattern Recognition Techniques, Technology and Applications*. IN-TECH, 2008. pp. 43-90.
 [21] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction". In: *IEEE International Conference on Pattern Recognition*, 2004, pp.28-31.
 [22] B. He, Z. Chen and Y. Li, "Calibration method for a central catadioptric-perspective camera system". *Journal of the Optical Society of America*, vol. 29, 2012, pp.2514-2524.
 [23] Y. Bastanlar, "A simplified two-view geometry based external calibration method for omnidirectional and PTZ camera pairs". *Pattern Recognition Letters*, vol. 71, 2016, pp.1-7.
 [24] X. Zhou, R. T. Collins, T. Kanade and P. Metes, "A Master-Slave System to Acquire Biometric Imagery of Humans at Distance". In: *ACM International Workshop on Video Surveillance*, 2003, p.113-120.
 [25] M. Tarhan and E. Altug, "A catadioptric and pan-tilt-zoom camera pair object tracking system for UAV". *Journal of Intelligent & Robotic Systems*, vol. 61, 2011, pp.119-134.
 [26] H.C. Karaimer, I. Baris and Y. Bastanlar, "Detection and classification of vehicles from omnidirectional videos using multiple silhouettes". *Pattern Analysis and Applications*, vol. 20(3), 2017, pp. 893-905.