

CLASSIFICATION AND TRACKING OF VEHICLES WITH HYBRID CAMERA SYSTEMS

**A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
MASTER OF SCIENCE
in Computer Engineering**

**by
İpek BARIŞ**

**June 2016
İZMİR**

We approve the thesis of **İpek BARIŞ**

Examining Committee Members:

Assoc. Prof. Dr. Muhammed Gökhan CİNSDİKİCİ
International Computer Institute, Ege University

Assist. Prof. Dr. Yalın BAŞTANLAR
Department of Computer Engineering, İzmir Institute of Technology

Assist. Prof. Dr. Mustafa ÖZUYSAL
Department of Computer Engineering, İzmir Institute of Technology

02 June 2016

Assist. Prof. Dr. Yalın BAŞTANLAR
Supervisor, Department of Computer Engineering
İzmir Institute of Technology

Assoc. Prof. Dr. Yusuf Murat ERTEN
Head of the Department of
Computer Engineering

Prof. Dr. Bilge KARAÇALI
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

First and foremost, I would like to express my deep gratitude to my supervisor Asst. Prof. Dr. Yalın Bařtanlar, for his excellent guidance, caring, patience and sharing his knowledge with me during development of this thesis work. I also thank Asst. Prof. Dr. Mustafa Özuysal for his advises in the system implementation and Assoc. Prof. Dr. Muhammed Gökhan Cinsdikici for evaluating this work.

I would also like to thank my lab-mates for their friendship and creating such a good atmosphere in the lab.

The thesis work was a part of the project 113E107 (“Classification of Objects in Traffic Scenes using Omnidirectional and PTZ Cameras”) supported by The Scientific and Technical Research Council of Turkey (TUBITAK).

ABSTRACT

CLASSIFICATION AND TRACKING OF VEHICLES WITH HYBRID CAMERA SYSTEMS

The integrated usage of several vision systems is especially important for surveillance applications. In case of a hybrid system combining an omnidirectional and a PTZ (pan-tilt-zoom) camera, the omnidirectional camera provides 360° horizontal FOV (Field of View) with a low resolution per viewing angle whereas the PTZ camera provides high resolution at a certain direction. In this thesis work, we introduce a hybrid system combining the powerful aspects of both camera types and aims a wide angle high resolution surveillance for traffic scenes. The hybrid system provides real-time object classification and high resolution tracking. The omnidirectional camera detects the moving objects and then it performs an initial classification by using shape-based features. Concurrently, the PTZ camera classifies the objects in detail by using HOG (Histogram of Oriented Gradients)+SVM (Support Vector Machine) pair. The object types we worked on are pedestrian, motorcycle, car and van. In the experiments, we compared the classification accuracy of omnidirectional camera, PTZ camera and hybrid system. Aiming high resolution tracking, the PTZ camera tracks the objects belonging to the user defined class and detected by using the omnidirectional camera.

ÖZET

HİBRİT KAMERA SİSTEMLERİ İLE TAŞIT SINIFLANDIRMA VE TAKİBİ

Görme sistemlerinin entegre kullanımı, özellikle gözetleme uygulamaları için önemlidir. Bir tümyönlü kamera, bir de PTZ (Pan-Tilt-Zum) kamerayı birleştiren bir hibrit sistemi ele alırsak, tümyönlü kamera yatay ekseninde 360° düşük çözünürlüklü görüş açısı sağlarken PTZ kamera belirli yönde yüksek çözünürlük sağlar. Bu tez çalışmasında, her iki kamera türünün güçlü yönlerini birleştiren ve geniş açılı yüksek çözünürlüklü gözetleme amaçlayan bir hibrit sistem tanıttık. Bu hibrit kamera sistemi, gerçek zamanlı sınıflandırma ve takip sağlıyor. Hibrit sistem sınıflandırması için, tümyönlü kamera hareketli nesnelere tespit eder, şekil tabanlı öznitelikler kullanarak ilk kademe sınıflandırma yapar. Eş zamanda, PTZ kamera HOG (Yönlü Gradyan Histogramı)+SVM (Destek Yöney Makinesi) çifti ile ikinci kez sınıflandırma yapar. Çalıştığımız nesne türleri yaya, motosiklet, araba ve dolmuş. Tümyönlü kamera, PTZ kamera ve hibrit sistem ile yapılan sınıflandırmaların başarılarını karşılaştırdık. Yüksek çözünürlüklü takip modülünde ise, PTZ kamera, kullanıcı tanımlı sınıfa ait ve tümyönlü kamera tarafından tespit edilen nesneyi sürekli çerçeve içerisinde tutar.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF SYMBOLS	x
LIST OF ABBREVIATIONS	xi
CHAPTER 1. INTRODUCTION	1
1.1. Motivation	1
1.2. Related Works	2
1.2.1. Object Detection and Classification with Standard FOV Cam- eras	2
1.2.2. Object Detection and Classification with Omnidirectional Cam- eras	3
1.2.3. Detection and Tracking with Hybrid Systems	4
1.3. Contributions	5
1.4. Organization of Thesis	6
CHAPTER 2. OBJECT DETECTION, TRACKING AND CLASSIFICATION WITH THE OMNIDIRECTIONAL CAMERA	7
2.1. Background Subtraction	8
2.2. Kalman Filter and Hungarian Algorithm.....	9
2.3. Extraction of Shape Based Features	12
2.4. kNN Classification	13
CHAPTER 3. OBJECT DETECTION AND CLASSIFICATION WITH THE PTZ CAMERA	17
3.1. Background Subtraction	17
3.2. SVM Classification with HOG Features	18

CHAPTER 4. HIGH RESOLUTION TRACKING WITH HYBRID CAMERA SYSTEM	22
4.1. Pan and Tilt Angle Calculation	22
4.2. Zoom Calculation	23
4.3. Object Tracking	24
CHAPTER 5. EXPERIMENTAL RESULTS	28
5.1. Classification with Omnidirectional Camera	28
5.2. Classification with Hybrid Camera System	28
5.3. High Resolution Object Tracking	31
CHAPTER 6. CONCLUSION	36
REFERENCES	37

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1 Photos of the motorcycle were taken with an omnidirectional camera and a PTZ camera	1
2.1 System diagram of proposed hybrid classification	7
2.2 Examples of the silhouettes based on ABL background subtraction along with morphological operations	8
2.3 Silhouette examples of a motorcycle and a pedestrian whose elongation values are close to each other but height-width ratios are quite different	14
2.4 2D normalized shape-based features of samples in our dataset	15
2.5 Example of hybrid system classification	16
3.1 Operations done with PTZ camera	18
3.2 Examples of the silhouettes based on MOG2 background subtraction along with morphological operations	19
3.3 Comparison between ABL and MOG2 background algorithms	20
4.1 Grid points of the omnidirectional camera used in look-up table	23
4.2 Samples from the look-up table	24
4.3 Bilinear Interpolation	25
4.4 Graph of area vs. zoom	26
4.5 System diagram of proposed cooperative object tracking	27
4.6 Block diagram for modifying position to calculate pan and tilt angles	27
5.1 Test samples for the SVMs	32
5.2 Misclassified examples of motorcycle and van.	32
5.3 Example of pedestrian tracking	33
5.4 Example of van tracking	34
5.5 Example of car tracking	35

LIST OF TABLES

<u>Table</u>	<u>Page</u>
5.1 Accuracy of omnidirectional camera classification using shape based features .	29
5.2 Accuracy of PTZ camera classification when SVM is selected based on height-width ratio threshold	30
5.3 Accuracy of PTZ camera classification when SVM is selected based on kNN classifier using height-width ratio as feature	30
5.4 Accuracy of hybrid system classification	31



LIST OF SYMBOLS

B	background model
I	input frame
a	background learning rate
A	state transition matrix
dt	elapsed time
\vec{s}	state vector
\vec{m}	measurement vector
Q	process noise covariance matrix
n_a	process noise magnitude
R	measurement noise covariance matrix
n_x	measurement noise in x direction
n_y	measurement noise in y direction
C	input control matrix
u	input control signal
E	elongation
S	short edge of the minimum rotated bounding rectangle
L	long edge of the minimum rotated bounding rectangle
W	width of the minimum contour rectangle
H	height of the minimum contour rectangle
X_{PTZ}	X coordinate of modified object centroid to compute pan-tilt angles
Y_{PTZ}	Y coordinate of modified object centroid to compute pan-tilt angles
X_{kalman}	X coordinate of object centroid computed by Kalman Filter
Y_{kalman}	Y coordinate of object centroid computed by Kalman Filter
β	variable providing to direct PTZ camera ahead in the predicted course

LIST OF ABBREVIATIONS

FOV	Field of View
PTZ	Pan-Tilt-Zoom
kNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
HOG	Histogram of Oriented Gradients
SVM	Support Vector Machine
ABL	Adaptive Background Learning
MOG2	Improved Mixture of Gaussians
GPU	Graphics Processing Unit

CHAPTER 1

INTRODUCTION

1.1. Motivation

An omnidirectional camera is a stationary camera providing 360° FOV (Field of View) in the horizontal plane with a low resolution per viewing angle. A PTZ (Pan-Tilt-Zoom) camera, on the other hand, provides high resolution at narrow angle and its viewing direction and zoom value can be remotely controlled. Two sample images of a scene with a motorcycle, captured by an omnidirectional and PTZ camera are shown in Figures 1.1a and 1.1b, respectively.

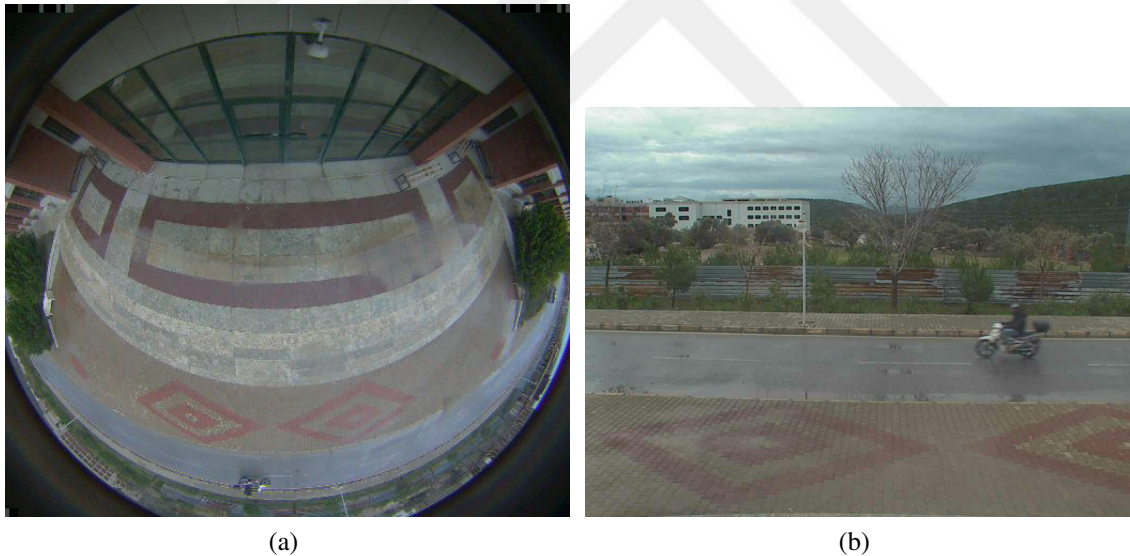


Figure 1.1. Photos of the motorcycle were taken with an omnidirectional camera (1.1a) and PTZ camera (1.1b)

A hybrid system combining the powerful aspects of both camera types aims a wide angle and high resolution surveillance. In the rest, the term ‘hybrid system’ will refer to a camera system including an omnidirectional and a PTZ camera. This powerful combination can be used in many application areas including robot navigation [1], 3D

reconstruction [3] and surveillance. Pointing the PTZ camera to a moving object detected by the omnidirectional camera is a typical surveillance task employing a hybrid camera system [22],[36], [17],[14].

A hybrid system for traffic scenes can provide the following benefits:

- Consider a scenario where the detection and tracking of a certain vehicle type is required. In such a scenario, the omnidirectional camera can make a rough classification of the vehicles in the scene. Then, the PTZ camera oriented by the omnidirectional camera will be able to track only the objects that are potentially in the target class. First elimination mentioned is very important for the cases where there are different types of objects moving and PTZ camera cannot handle tracking all objects in the scene.
- PTZ cameras can obtain high resolution images of the tracked object. The images can be used for detailed classification or recorded only for surveillance. At the same time the omnidirectional camera can continue with its monitoring and classification tasks. In this way, monitoring traffic and acquiring high resolution images of target vehicle types can be done in parallel.
- Besides, an omnidirectional-PTZ camera pair can perform the tasks which otherwise would require many standard FOV cameras.

1.2. Related Works

This section reviews some past research regarding shape-based and gradient-based classification as divided into two subsections for which are standard FOV and omnidirectional cameras. Besides, the literature review on detection and tracking with hybrid systems is also presented in another subsection.

1.2.1. Object Detection and Classification with Standard FOV

Cameras

Shape-based features are computed by using the blobs extracted based on a background subtraction algorithm such as adaptive background algorithm. As an example of

shape based features from PTZ or standard cameras, a feature containing length, bounding box dimensions is used to separate car and truck [15]. Another example, Morris et al. [27] derives ten features that are area, bounding box (width and height), convex area, ellipse, extent, solidity, perimeter from the blob tracked with Kalman Filter whose prediction states used for data association. Then including tracking information, the classification of the vehicle is done within Sedan, Semi, Truck+SUV+Van with weighted kNN (k-Nearest Neighbors) after having applied LDA (Linear Discriminant Analysis) to remove redundancy. Kumar et al. [24] proposed a framework that the classification of the vehicles is done based on the size, shape, velocity, and position of the vehicle using a Bayesian network. Buch et al. [5] conducted vehicle detection and classification using 3D models based on the manufactured dimensions of vehicles projected onto the image plane for obtaining a silhouette match measure.

Instead of using shape based features, it is possible to extract image-based features for object detection. For instance, all pixels of the image of tracked vehicle were used as features after the image had been resized [27]. Another example for this purpose is in human detection [10] which object appearance and shape are characterized as HOG (Histogram of Oriented Gradients) features extracted with the sliding window approach, and new samples were classified with linear SVM (Support Vector Machine). Considering its overwhelming time complexity, it is necessary to obtain a region of interest based on background subtraction on video for HOG features extraction as we applied on this thesis work.

1.2.2. Object Detection and Classification with Omnidirectional Cameras

Shape based and gradient based features were used in omnidirectional cameras as well. Khoshabeh et al. [22] classified the vehicles as large (truck, bus, etc.) or small (car, motorcycle, etc.) by only the covered area in the image.

Karaimer et al. [20] extracted shape based measurements which are convexity, elongation, rectangularity and Hu moments, from average silhouettes of the largest blobs in predetermined range. Firstly, convexity filtered out the poor detections that might not belong to vehicle classes. Then, if the object was passed from the filter, it was classified as motorcycle, car or van by using SVM with the other features.

As an example of a classification with SVM using HOG features, Cinaroglu et al. detected car, van and pedestrian with HOG computation adopted mathematically to omnidirectional camera [8], [9].

Other examples are usage of HOG features computed from recorded videos [13], [21]. For instance, Gandhi et al. [13] extracted HOG features as labeled the bottom center of the vehicles manually by using their virtual perspective views from omnidirectional camera. This method, however, is not applicable for real-time system. Karaimer et al. [21] extracted HOG features onto bounding rectangles obtained based on background subtraction.

1.2.3. Detection and Tracking with Hybrid Systems

To cooperatively use a hybrid system such as ours, the geometric relation between the omnidirectional camera and the PTZ camera should be extracted. Only after that, it is possible to direct the PTZ camera to the object of interest detected in the omnidirectional camera. There are several major approaches to solve this geometric problem.

The first one is performing a complete external calibration of the hybrid system without restricting the rotation and translation between the camera pair. This allows all degrees-of-freedom but usually it is not practical and highly time consuming. For instance, a large pattern on the floor is required for the method in [7] and [16]. In [11], the rotation and translation between the cameras are extracted via 3D Euclidean reconstruction of scene points following projective reconstruction by factorization which is computationally expensive due to using non-linear minimization techniques.

A second group of studies make assumptions about the camera setup to be able to use some geometric constraints for a practical external calibration. One of the most common assumptions is that the optical axes of the two cameras coincide [25] (i.e. one is on top of the other). In this approach, intrinsic parameters of the omnidirectional cameras are extracted beforehand. Then, the distance between an object and the center of the omnidirectional image is used to compute tilt (vertical) angle of the PTZ camera. Pan angle is found by defining the reference zero position of PTZ camera in the omnidirectional image. In some other studies [32],[2], less restrictive assumptions were made where the cameras can move almost anywhere but their optical axes should be perpendicular to the ground. Solving geometric equations by using only two scene points it was possible to

extract rotation and translation parameters.

Another major group do not estimate any external parameters of the hybrid system, but directly estimate the relation between the pixels of the omnidirectional camera frame and the pan/tilt angles of the PTZ camera. I.e. What should be the PTZ camera pan/tilt values for a given pixel in the omnidirectional image? This is called spatial mapping and based on data collection and fitting (interpolation). The method called homography calibration in [6] falls into this category. In [30] and [33] where hybrid surveillance and tracking systems were proposed, tilt angle of the PTZ camera is estimated by interpolation of several points, whereas pan angle is computed by using the zero reference point. In [18], a similar approach is used for the presented hybrid face detection and tracking system.

1.3. Contributions

We have implemented an intelligent hybrid system providing real-time object classification and tracking. The object types we have worked on are pedestrian, motorcycle, car and van. The hybrid system consists of two modules, one for classification and the other for tracking.

In the first module, the omnidirectional camera performs rough classification based on shape-based features of the detected object. In parallel, PTZ camera classifies the object in detail by using its gradient-based features. For the classification in PTZ camera, we have designed pedestrian-motorcycle SVM with 1764 HOG features and car-van SVM with 4788 HOG features, which provides less computation and minimum storage. To select which of the mentioned SVMs is employed for a detected object, we implemented two different approaches. The first one is using height/width ratio of the blob detected by the PTZ camera, the other one is using the result of the omnidirectional camera classification.

The procedures done in the omnidirectional camera part of the first module are applied to the second module as well. The user defines a class of the object to be tracked. The omnidirectional camera eliminates the objects which do not belong to the defined class, and determines the candidate object for tracking. Then, the PTZ camera is controlled by defining pan/tilt/zoom values of the object detected with omnidirectional camera in order to get high resolution frames. We tested our system for detecting a target class

object in the omnidirectional camera and steering the PTZ camera for high-resolution tracking.

1.4. Organization of Thesis

The organization of this thesis is as below:

Chapter 2: This chapter gives the information on procedures done with the omnidirectional camera. The algorithms related to the background subtraction, data association and tracking are given. Then, how we extract the shape-based features and perform the classification is explained.

Chapter 3: This chapter describes the tasks on the PTZ camera. Background Subtraction with PTZ camera is described. Then, the extraction of HOG features and classification with SVM are explained.

Chapter 4: In this chapter, we explain how the omnidirectional and PTZ cameras are used cooperatively. Spatial mapping in between and calculation of pan, tilt, and zoom parameters are described.

Chapter 5: The implementation of the environment for the experiments is briefly explained. The results of the experiments on classification and tracking are evaluated.

Chapter 6: Future works and conclusions are given.

2.1. Background Subtraction

Background subtraction is a process to segment moving objects from the background in an image sequence. The simplest approach is the subtraction of the current frame from the previous frame. Despite its simplicity, it is not often applicable in real environments [29]. On the other hand, there are robust and basic background subtraction algorithms, also adaptive to change in environment, in the literature. The evaluation of the background subtraction algorithms from basic to complex is presented by [31]. ABL (Adaptive Background Learning) which can be considered one of the fast algorithms in this evaluation has been employed for silhouette extraction in omnidirectional video. In this algorithm, simply, a background model is updated with a learning rate for each frame and then foreground objects are detected by the subtraction between the input frame and updated background model. After silhouette extraction, a series of morphological operations are applied to clean the silhouettes. The silhouette examples for each class are given in Figure 2.2.



Figure 2.2. Examples of the silhouettes based on ABL background subtraction along with morphological operations. The images were cropped from the omnidirectional images for better visualization. The silhouettes at top-left: van, top-right: pedestrian, down-left: motorcycle, down-right: car.

The extraction of the silhouette with ABL algorithm involves the following steps:

1. Initially, if the background model is empty, the input frame is copied as a background model.
2. The foreground image is obtained by computing the absolute difference between input image and background model.

3. The background model is updated with the formulation given in Equation 2.1, where B is the background model, I is the input frame and a is the background updating factor.
4. The foreground image is converted to binary image by applying threshold.

The steps 2-4 are repeated for incoming frames.

$$B(x, y, t) = (1 - a)B(x, y, t - 1) + aI(x, y, t) \quad (2.1)$$

Morphological operations, closing followed by opening with disks, are applied to the foreground image to remove small holes and small objects. After morphological operation, the contour area is computed and the object is assigned as a candidate for tracking. Applying an area threshold, which is the below of the minimum area value of the training dataset, the candidate object higher than the threshold is assigned as detected object. Then, the centroid and the contour of the detected object are used for tracking and classification in the following phases.

2.2. Kalman Filter and Hungarian Algorithm

Kalman Filter estimates the state of a dynamic system by using a form of feedback control. The filter is very powerful for refining the observations if the system is in noisy environment. Kalman Filter equations fall into two groups: time update equations and measurement update equations. Time update equations compute the predicted states and the predicted state is updated from the observation with the measurement update equations [34].

Kalman filter design depends on the system. It is characterized with a state transition matrix, control input matrix, state variables and measurements. If the noise is taken into account, process noise covariance matrix, measurement noise covariance matrix are defined as well. Prediction and measurement phase are same in all of the systems. The following three core phases were taken for Kalman Filter process:

- *Preliminary:* Initially, the system model is built as a 4x4 state transition matrix with state variables which are position, velocity. State transition matrix is denoted by A where dt variable is elapsed time between the states:

$$A = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

\vec{s} is a vector containing state variables:

$$\vec{s} = [x, y, V_x, V_y]$$

Measurements are fed with the centroids of the silhouettes and are defined as entries of \vec{m} vector:

$$\vec{m} = [x, y]$$

A 4x4 process noise covariance matrix, Q , is defined assuming acceleration is process noise, n_a :

$$Q = \begin{bmatrix} n_a \times \frac{dt^4}{4} & 0 & n_a \times \frac{dt^3}{2} & 0 \\ 0 & n_a \times \frac{dt^4}{4} & 0 & n_a \times \frac{dt^3}{2} \\ n_a \times \frac{dt^3}{2} & 0 & n_a \times dt^2 & 0 \\ 0 & n_a \times \frac{dt^3}{2} & 0 & n_a \times dt^2 \end{bmatrix}$$

Measurement noise and process noise are independent from each other. Therefore, 2x2 measurement noise covariance matrix is defined as R where n_x and n_y are measurement noise in x and y directions:

$$R = \begin{bmatrix} n_x & 0 \\ 0 & n_y \end{bmatrix}$$

- *Prediction:* This phase predicts the k_{th} state of the signal s and prior error covariance matrix P^- by using the equations 2.2 and 2.3. C and u are input control matrix and input control signal, respectively. There is no actuator such as motor in the system, hence the control input signal is zero and C is neglected.

$$\vec{s}_k = A\vec{s}_{k-1} + C\vec{u}_k \quad (2.2)$$

$$P_k^- = AP_{k-1}A^T + Q \quad (2.3)$$

- *Updating:* After prediction phase, the estimations are corrected with updating phase. In the updating phase, the first step is computation of the Kalman gain, K_k in Equation 2.4. The estimation of the s_k is updated with measurement m_k as Equation 2.5. In the final step, the posterior error covariance is calculated with Equation 2.6.

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad (2.4)$$

$$\vec{s}_k = \vec{s}_k^- + K_k (\vec{m}_k - H \vec{s}_k^-) \quad (2.5)$$

$$P_k = (I - K_k H) P_k^- \quad (2.6)$$

The prediction and updating phases are recursively repeated. The source codes for Hungarian Algorithm and Kalman Filter¹ have adopted to the system.

The omnidirectional camera is located at the region where several objects can appear concurrently at the scene. Assigning the largest blob as the silhouette belonging to the moving object for the classification [20] is inapplicable for this case. The object should be tracked and data association should be performed if a new object appears or multiple objects exist at the scene for preventing data confusion. Hungarian Algorithm and Kalman Filter together provide a solution for multiple target problems. One example based on Hungarian Algorithm and Kalman Filter using an omnidirectional camera is given in [12], which aims simultaneous localization and mapping.

Hungarian Algorithm (also referred to as Kuhn-Munkres algorithm) is a global method solving assignment and transportation problems in polynomial time [23],[28]. The implementation of the algorithm for the system is based on the following steps:

1. A cost matrix is created which has the number of rows as the length of the detection vector and the number of columns as the length of the tracks vector. Each cell of the matrix is initialized with the Euclidean distance between the centers of each detection and each track.

¹<https://github.com/Smorodov/Multitarget-tracker> (accessed 15 April 2015).

2. For each row of the matrix, the smallest element is subtracted from the each entry in its row.
3. The elements which are zero in the resulting matrix are marked as a starred zero if there is no other starred zero in their rows or columns.
4. The columns which contain starred zero are covered. If the total of covered columns is equal to minimum count between the number of the detections and tracks, it skips to the last step.
5. A noncovered zero is found and marked as a primed zero. Unless there isn't starred zero in its row, it skips to step 6. Otherwise, the column containing starred zero is uncovered and the row is covered. This step is repeated until uncovered zeros remain. At last, the smallest uncovered value is saved.
6. Until column of the primed zero does not contain any starred zero, each starred zero is unstarred, primed ones are starred. After all primes are erased and every line is uncovered, it returns to step 4.
7. The smallest uncovered value found in Step 4 is added to all entries of each covered row, and it is subtracted from the entries of each uncovered column. Then, it skips to step 5.
8. Lastly the identities of track and pedestrian, designated with the position of the starred zeros, are matched and stored in an assignment vector. For each assignment, if the cost is above predetermined distance, it is removed from the vector.

2.3. Extraction of Shape Based Features

The objects within a class show similar patterns. These patterns are characterized as the features. The most important process for implementing an automated system for object classification is the feature selection. The features should provide the system grouping new data with high cohesion and loose coupling. As the represent of patterns in vehicles, the shape based features such as rectangularity, elongation etc. can fulfill the requirement for vehicle classification with omnidirectional cameras [20]. The shape

based features were employed for the classification via the omnidirectional camera in our system as well.

To select features from a pool, which primarily contains area, elongation, rectangularity, convexity and height-width ratio, their potentials of separation were analyzed. Rectangularity and convexity do not have enough variation to discriminate between vehicle classes. Therefore, area, elongation and height-width ratio were included in the feature vector used for classification. The calculation of them is based on contours (silhouettes).

The description of each feature is as follows:

1. *Area*: It is a one dimensional feature giving the silhouette area. It can separate small vehicles from the large ones [22].
2. *Elongation*: It is calculated by using the Equation 2.7 [35]. In the equation: E is elongation, S is the short and L is the long edge of the minimum rotated bounding rectangle. It discriminates motorcycles and pedestrians from the other classes.

$$E = 1 - (S/L) \quad (2.7)$$

3. *Height-width Ratio*: Height-width ratio, $Ratio$, is calculated by using Equation 2.8 where H , W are height and width of the minimum contour rectangle, respectively. Since the ratio of the pedestrian larger than 1 and the other ones are smaller than 1, it can be discriminative parameter to separate the pedestrians from the others. There are cases that motorcycle and pedestrian have elongation values which does not differ from each other. In such cases, height-width ratio distinguishes between motorcycle and pedestrian (an example is given in Figure 2.3).

$$Ratio = H/W \quad (2.8)$$

To illustrate the combination of the features, 2D normalized features are plotted in Figure 2.4. As can be observed, if the elongations of the samples are approximately same, area can separate van-car from pedestrian-motorcycle, and height-width ratio separates pedestrians from other classes.

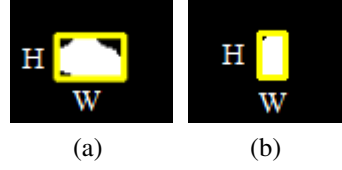


Figure 2.3. Silhouette examples of a motorcycle (left) and a pedestrian (right) whose elongation values are close to each other (0.42 and 0.41, respectively) but height-width ratios are quite different (0.65 for motorcycle, 1.63 for pedestrian). H and W denote the height and width of the non-rotated bounding rectangle (Equation 2.8).

2.4. kNN Classification

Classification of the vehicles is the final stage in this part of the system. While the angle of the object is in the range of predetermined angles (in the system, this range is set as $[-30^\circ, 30^\circ]$ assigning 0° is the angle that the camera is the closest to the road), firstly features of the object is being extracted and then it is being classified using the features. Each time the object is classified by, the assigned class is voted. Once it passes the range, the most voted class is assigned as the object's final class. An example of this procedure given in 2.5.

kNN classification is the approach employed to the system. It is a non-parametric algorithm that stores all labeled training dataset and decides the class of the object by looking k nearest ones of the dataset with the distance measurement. This approach has been applied by using features extracted from either single silhouettes (e.g. [27],[26]) or multiple silhouettes [19].

Before the implementation of the classifier, the samples in training dataset were labeled and their features were normalized by using rescaling method given in Equation 2.9 where x is a value of the feature, x' is its normalized form and the denominator is the difference between the maximum value and minimum value of the corresponding features in the training dataset.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.9)$$

The features of the test data were also rescaled. If the value of the feature is upper 1, it is assigned to 1. Similarly, if the value of the feature is below 0, it is assigned to 0. The objects are classified by using the normalized features.

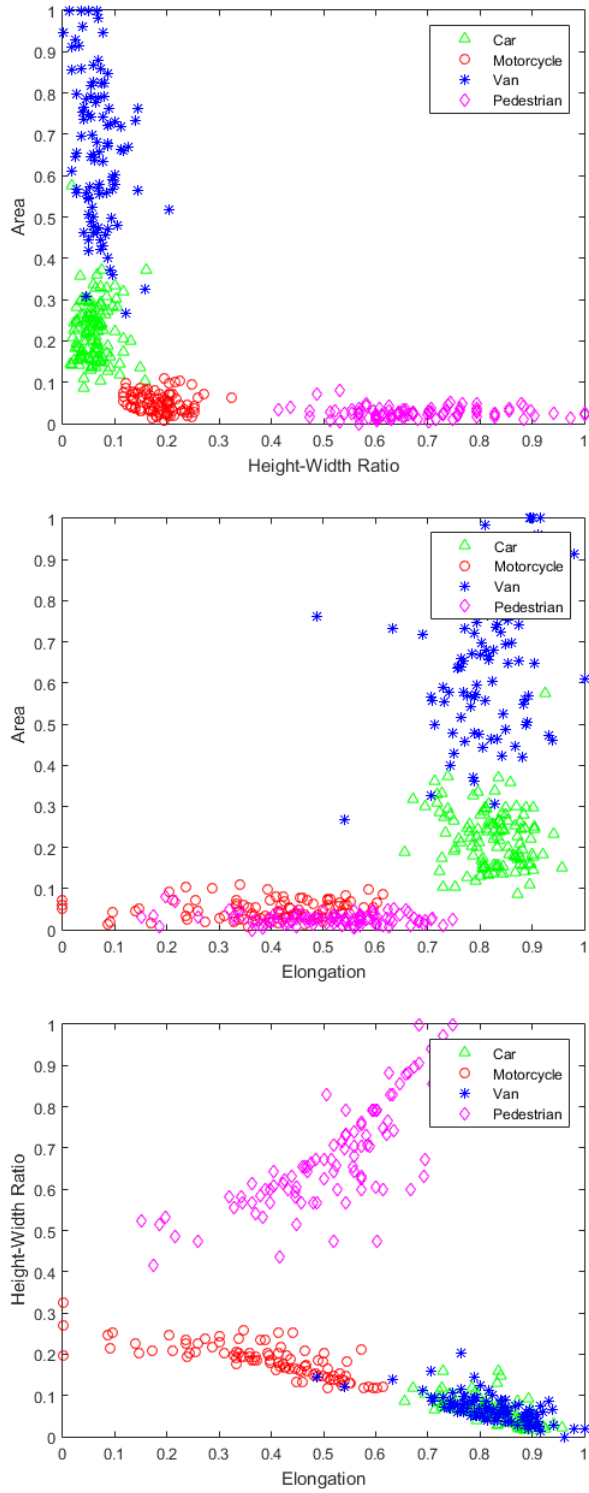


Figure 2.4. 2D normalized shape-based features of samples in our dataset.

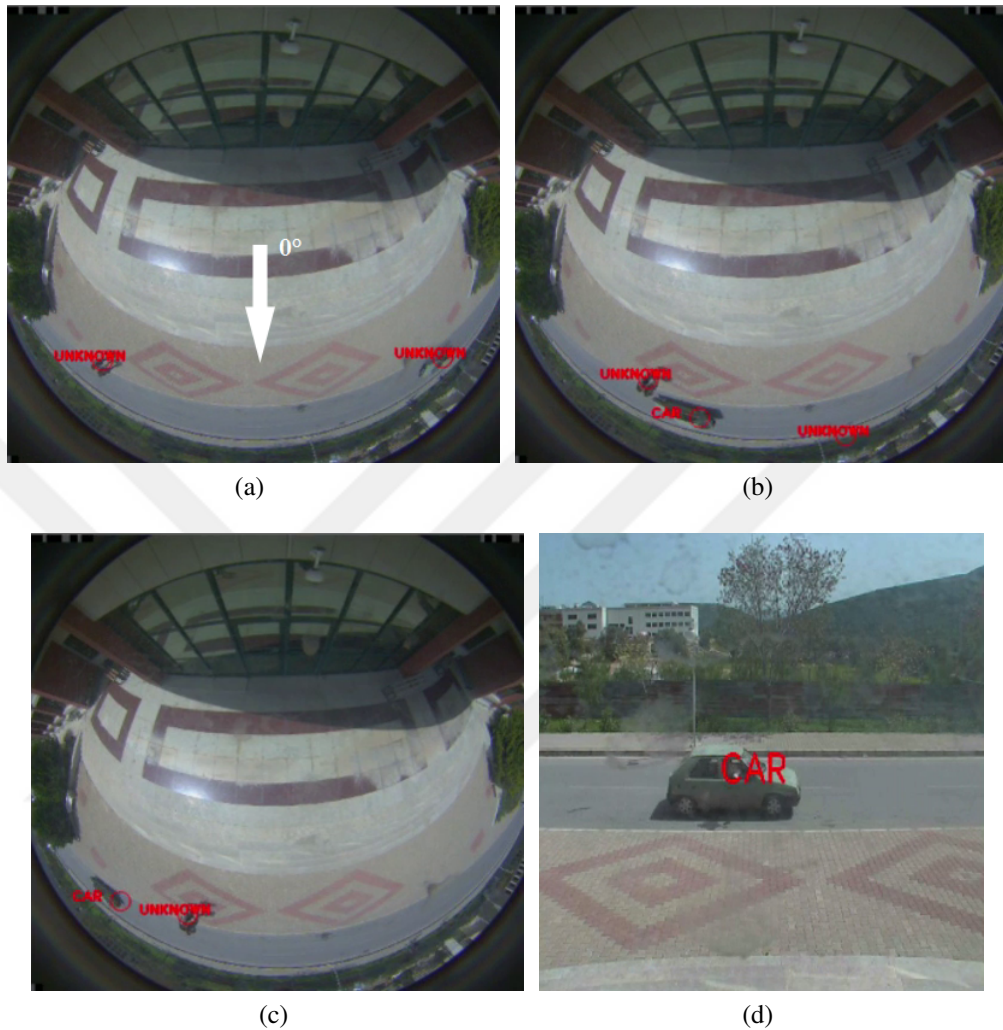


Figure 2.5. Example of hybrid system classification. (a) The objects are not in the target angle range yet, so not classified, they are labeled as ‘UNKNOWN’. (b) Classification of the car is done since it passed through the angle range. (c) The tracking of the classified object is continued with its new label ‘CAR’. (d) The result of the 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.

CHAPTER 3

OBJECT DETECTION AND CLASSIFICATION WITH THE PTZ CAMERA

This chapter presents the methodology we employed for the PTZ camera of the proposed hybrid system. PTZ camera, controlled by the omnidirectional camera, is only sufficient enough to keep the moving objects in the frame of PTZ camera. Exact location of the object cannot be estimated since the IP cameras are not synchronized due to the delays that occur while acquiring frame from them and control of the PTZ camera. Therefore, moving object is detected based on background subtraction with the PTZ camera as well. The object detection and classification starts when the PTZ camera is positioned at the desired location where side view of the vehicle is seen, and the PTZ camera is not moved while detection and classification are being done.

Stages of the classification after background subtraction and morphological operation on PTZ camera are shown in Figure 3.1. 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. In third stage, an SVM decider chooses the classifier to which the object is sent by using height-width ratio (low height-width ratio indicates it is a car or a van). After the classifier is selected, the rectangle is enlarged and the region on the enlarged rectangle rescaled to fit into the defined height-width ratio (1:1 for pedestrian/motorcycle, 1:2.5 for car/van). As a final step, the object is classified by using SVM with HOG features computed on the region.

The details of background subtraction and classification performing on the PTZ camera are given in subsections 3.1 and 3.2, respectively.

3.1. Background Subtraction

HOG features are generally used for object detection in an image via sliding window approach. Since this is a time consuming process, in a real-time system a region which encloses the object should be automatically defined by a background subtraction

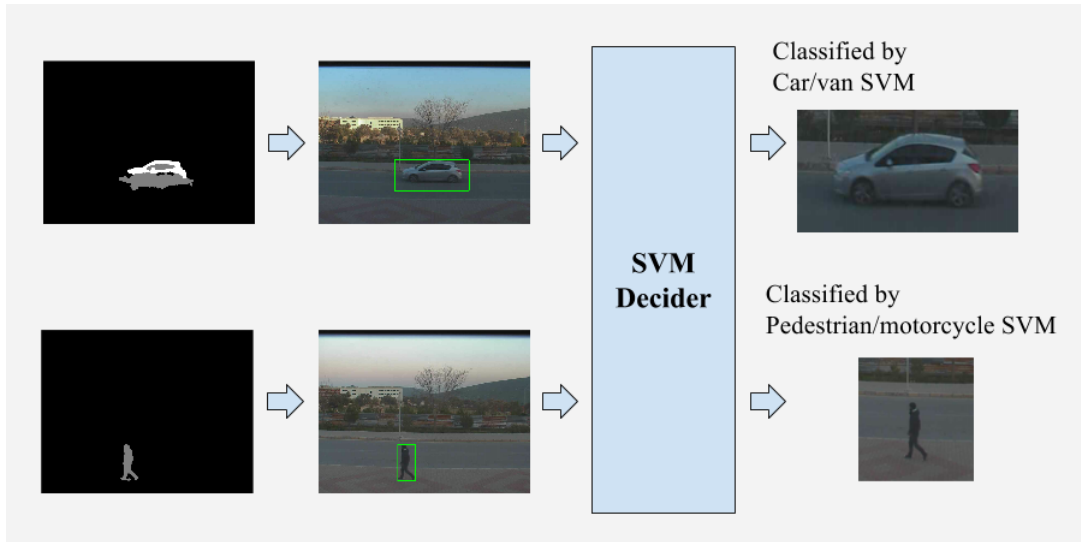


Figure 3.1. Operations done with PTZ camera. Minimum bounding rectangle enclosing the silhouette is obtained and the SVM classifier is determined using height-width ratio. The rectangle is enlarged and the window is cropped (the ratio of the window is 1:1 for motorcycle and pedestrian, 1:2.5 for van and car). The classification of the object is performed based on HOG features.

algorithm. In this step, MOG2 (Improved Mixture of Gaussians) background subtraction algorithm is deployed to the system [37]. We observed that MOG2 is better than ABL in noise and shadow robustness for PTZ frames and it generates better windows enclosing the silhouettes. A pair of examples obtained with both background subtraction algorithms are given in Figure 3.2.

In MOG2 algorithm, a background model is estimated as parametric Gaussian mixture probability density from a training dataset. Any pixel which does not meet the model is assumed foreground pixel, others are deemed to be background pixel. MOG2 selects needed Gaussian components per pixel automatically and for each new sample, the training dataset is updated periodically by online clustering algorithm that if the object remains stable for a long time, it is added to the training set discarding old ones and hence the background model is reestimated. We computed average processing time of MOG2 algorithm processed in GPU (Graphics Processing Unit) by repeating 1000 times, and found 6 msec as the average processing time which is appropriate for real-time process capability of the hybrid system.

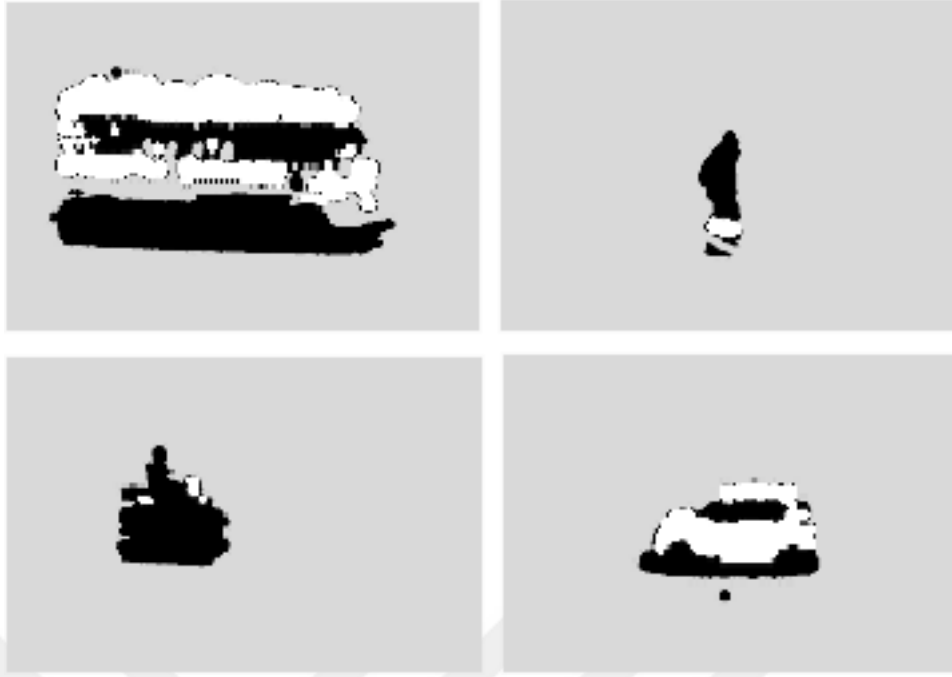


Figure 3.2. Examples of the silhouettes based on MOG2 background subtraction along with morphological operations. The images were cropped from the PTZ images for better visualization. The silhouettes at top-left: van, top-right: pedestrian, down-left: motorcycle, down-right: car.

3.2. SVM Classification with HOG Features

HOG is a feature descriptor based on oriented histograms, proposed for pedestrian detection and classification [10]. It provides information on object shape and appearance. SVM classifier is a supervised learning method used for binary classification. Moving object seen from the PTZ camera is classified at coarse level by using HOG+SVM pair in the hybrid system. We trained two different SVM classifiers for this purpose. Pedestrian/motorcycle SVM was trained with 1764 HOG feature values from 120x120 (1:1 ratio) pixels detection window. Car/van SVM was trained with 4788 HOG feature values from 120x300 (1:2.5 ratio) pixels detection window. Pedestrian/motorcycle SVM is trained as pedestrian images are positive samples, motorcycles are negative samples. In the same manner, car/van SVM is trained as car images are positive samples, vans are negative samples.

Following steps describe how HOG descriptors are computed [10]:

1. As a first step, gradient values are computed with a 1D centered derivative mask

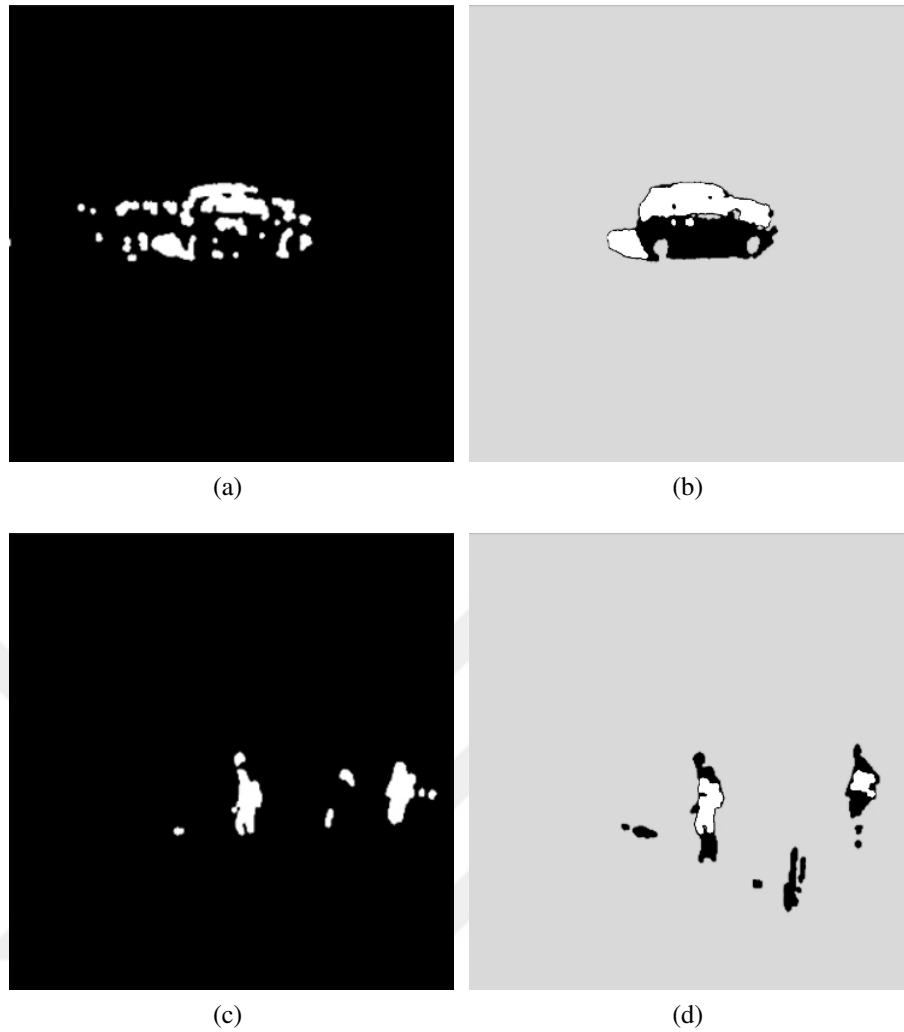


Figure 3.3. Comparison between ABL and MOG2 background algorithms. The silhouettes at the left column are extracted with ABL (a-c), the ones at the right column (b-d) are the MOG2 background subtraction algorithm [37].

defined as $[-1, 0, 1]$ applying to image in vertical and horizontal directions. Gradient values are unsigned. This step extracts the information of the object about its contour, silhouette and texture.

2. In the second step, the detection window is divided into cells each of which is 15×15 pixels with a rectangle shape. Gradient values computed in the first step within a cell are placed into a 9-bin histogram. Since the gradient values are unsigned, histogram channels range from 0° to 180° . Each pixel within a cell has a vote weighted with magnitude of the corresponding gradient value. These votes define the contribution of the gradient to the histogram.

3. Third step groups the adjacent cells and normalizes the histogram. Group of cells are described as blocks. 30x30 pixels blocks are defined for that purpose. As describing 15x15 pixels block stride, it provides the blocks 50% overlapping. The set of normalized histograms represent a HOG descriptor.

Apart from using height-width ratio of the silhouette detected with PTZ camera, the object's class detected with omnidirectional camera can be used to choose the SVM classifier, result of which is given in Section 2.1. To do this, we should ensure that the same object instance is classified in both cameras. PTZ camera uses the side-views of the objects since the training is performed with images seeing the vehicles from one side. Position of the vehicle in the omnidirectional camera is used to guarantee that the vehicle is in the FOV of the PTZ camera. This can be considered as the hybrid system classification since both cameras are involved. A visual example of such classification given in Figure 2.5d.

CHAPTER 4

HIGH RESOLUTION TRACKING WITH HYBRID CAMERA SYSTEM

The cameras are located in place where more than one objects can exist in the scene. PTZ camera is able to track only one object when such an event occurs. Therefore, we propose a cooperative object tracking system that PTZ camera tracks only an object which belongs to the target class (specified by the user) and detected by the omnidirectional camera. This chapter describes the computation of pan, tilt and zoom values to be sent to the PTZ camera and the procedures of the cooperative object tracking system.

4.1. Pan and Tilt Angle Calculation

The positions of the cameras were not aligned. To achieve a relation between pan/tilt angles and the coordinates of the omnidirectional camera, we simply built up look-up table. For the look-up table, we put a grid pattern on the omnidirectional image plane shown in Figure 4.1 and for each grid point on the grid; we collected corresponding pan and tilt angles of the PTZ camera. Two samples from the look-up table are shown in Figure 4.2.

For the intermediate locations (between the grid points), pan/tilt values are estimated with bilinear interpolation given in Equation 4.1 where the points are shown in Figure 4.3. In Equation 4.1, $(x,y) = P$ is the location of the object, to be interpolated and its neighborhood are denoted by $Q_{ij} = (x_{ij}, y_{ij})$. Function f finds corresponding pan or tilt angle of given coordinates. Other coordinates of the object where are out of or on boundary of the grid world are assigned as outliers and any command is sent to PTZ camera.

$$\begin{aligned} f(x, y) = 1/(x_2 - x_1)(y_2 - y_1) [& f(Q_{11})(x_2 - x)(y_2 - y) + \\ & f(Q_{21})(x - x_1)(y_2 - y) + \\ & f(Q_{12})(x_2 - x)(y - y_1) + \\ & f(Q_{22})(x - x_1)(y - y_1)] \end{aligned} \quad (4.1)$$

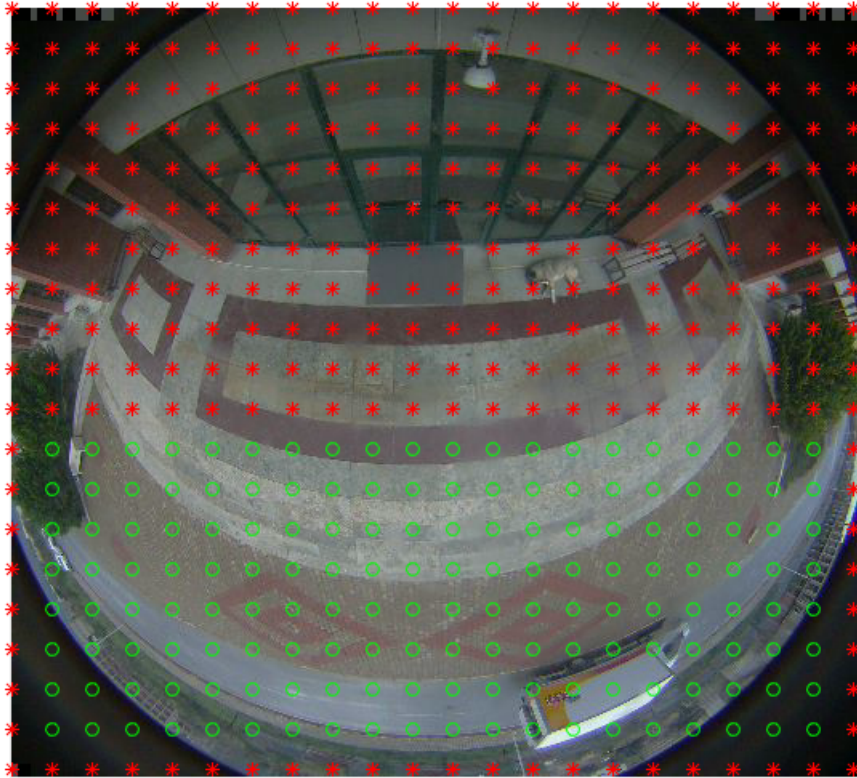


Figure 4.1. Grid points of the omnidirectional camera used in look-up table. Red dots are not seen from PTZ camera or not the region of our interest, so they were neglected. Constructed look-up table contains the pan/tilt pairs of the green circles.

4.2. Zoom Calculation

To obtain high resolution images of the object, we need to zoom in when the object covers a small part of the PTZ camera frame. We relate the area of the minimum bounding rectangle of the object in the omnidirectional image to the zoom value of PTZ camera by collecting samples of an object moving it close and far away from the camera. In a sense, we perform spatial mapping for the zoom value. As a result, red circles in Figure 4.4 are obtained. To satisfy those points, we derived Equation 4.2 which is depicted in Figure 4.4.



(a)



(b)

Figure 4.2. Samples from the look-up table (a-b). Images on the left side are from the omnidirectional camera, the rest is from the PTZ camera. Red circles on the omnidirectional camera correspond to grid coordinates, As for the ones on PTZ camera, they are centers of the corresponding points.

$$z(x) = \begin{cases} max_{zoom} & \text{if } x < threshold_{area} \\ \left(1 - \frac{x - threshold_{area}}{1000 - threshold_{area}}\right) \times max_{zoom} & \text{if } x \geq threshold_{area} \end{cases} \quad (4.2)$$

In this equation, maximum value for zooming in object is denoted by max_{zoom} . If the area of the minimum bounding rectangle of the object is smaller than $threshold_{area}$, then we do not zoom in more and use the max_{zoom} value. If the area is larger, than the best zoom value is computed. As can be observed from Figure 4.4, $threshold_{area}$ is around 200 pixels, and if the area is more than 1000 pixels, no more zoom out is possible.

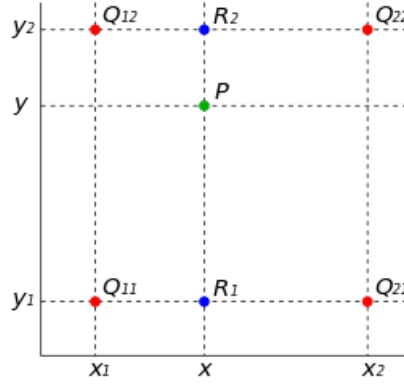


Figure 4.3. Bilinear Interpolation. Q_{ij} are grid points, and P is the point to be interpolated (The courtesy of wikipedia.org).

4.3. Object Tracking

The steps of the cooperative object tracking are as illustrated in Figure 4.5. First, object detection, tracking and classification are performed on omnidirectional camera. If the class of the object is the specified class defined by user and PTZ camera does not track another object at that time, pan and tilt values are calculated by using the location of the object and zoom is determined by the area of the object on omnidirectional camera.

Even though the position estimate of Kalman Filter is able to track detected vehicle on the omnidirectional camera, it is not sufficient to catch the vehicles due to the delays occurred during steering the PTZ camera. To solve this problem, the position is modified by using Equation 4.3 and then pan-tilt values, to be sent to the PTZ camera, are calculated by using this modified position. The block diagram of these procedures is shown in Figure 4.6. In Equation 4.3, (X_{PTZ}, Y_{PTZ}) denotes the modified position to be used for calculating pan and tilt angles, (X_{kalman}, Y_{kalman}) is the prediction of the centroid computed by Kalman Filter, ΔX_{kalman} , ΔY_{kalman} are the centroid displacements between consecutive frames and β is the variable which enables us to direct the PTZ camera ahead in the predicted course. β is defined at the range between -120° and 120° and it gets its maximum value at 0° (cf. Figure 2.5a) where change in displacement is maximum for a fixed speed, and it linearly decreases from 0° to the boundaries (-120° and 120°).

$$X_{PTZ} = X_{kalman} + \Delta X_{kalman} \cdot (1 + \beta), \quad Y_{PTZ} = Y_{kalman} + \Delta Y_{kalman} \cdot (1 + \beta) \quad (4.3)$$

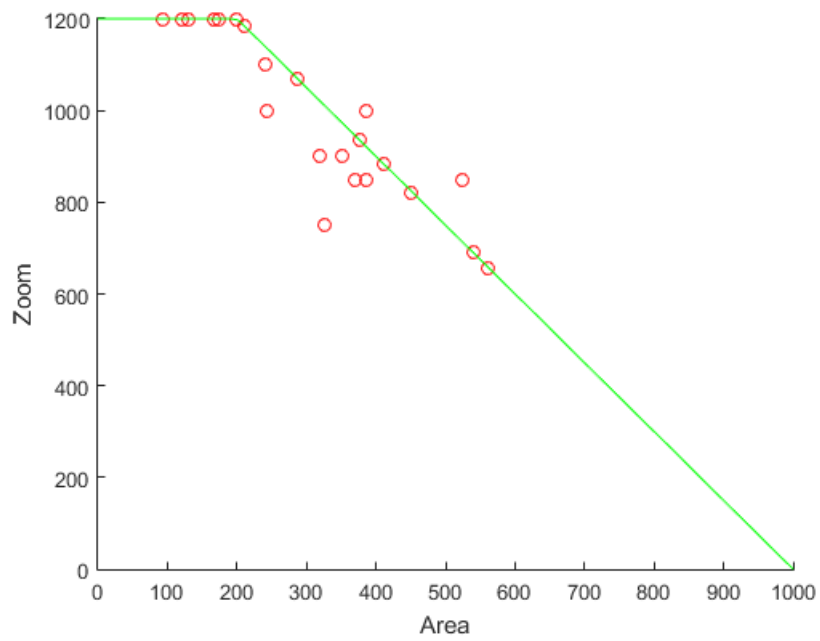


Figure 4.4. Graph of area vs. zoom. X axis is the area of bounding rectangle (in pixel), y axis is the zoom value of the PTZ camera (1200 refers to 24x zoom)) Red circles were obtained by the experiment; green lines are obtained by the Equation 4.2 which we used for the zoom calculation.

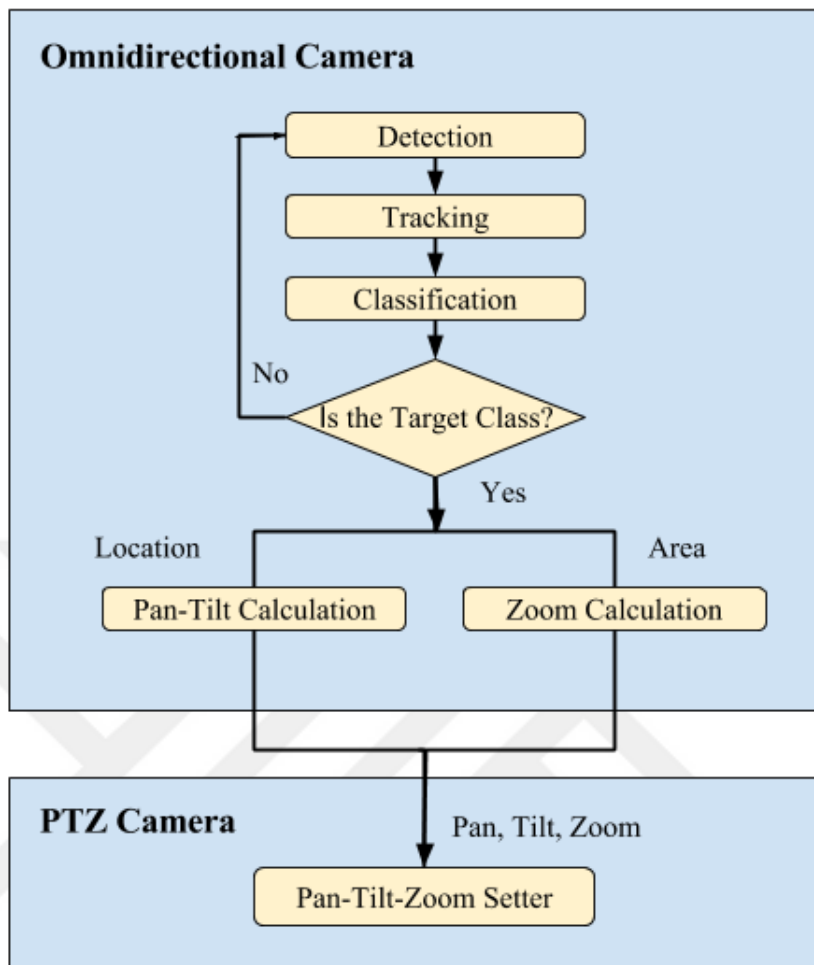


Figure 4.5. System diagram of proposed cooperative object tracking

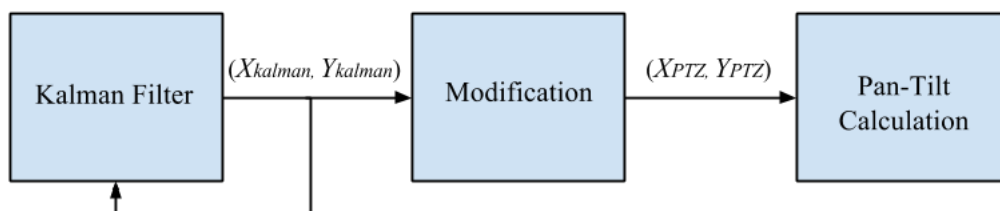


Figure 4.6. Block diagram for modifying position to calculate pan and tilt angles

CHAPTER 5

EXPERIMENTAL RESULTS

Oncam Grandeye 360° with fisheye lens as an omnidirectional camera, and Samsung SNP-3500 as a PTZ camera were used for the experiments. The cameras were mounted at the front side of a building where there is comparatively more traffic circulation rather than other buildings, at Izmir Institute of Technology. The resolution of the frames acquired from omnidirectional camera are 528x480 pixels, ones sampled from PTZ camera are 1024x768 pixels. The system has been coded with C/C++ in Visual Studio 2013. OpenCv 3.0 Library with CUDA [4] was used as library for adaptation to real-time image processing. The experimental results are summarized under three subsections that are classification with omnidirectional camera, classification with hybrid system and tracking performance.

5.1. Classification with Omnidirectional Camera

96 motorcycle, 125 car, 100 van and 102 pedestrian samples of single silhouettes were collected as dataset for kNN classification and their shape-based features were extracted. We implemented kNN classifier (cf. 2.4) where $k = 5$, using this dataset.

Another set was constructed for testing (94 vans, 113 cars, 71 motorcycles and 83 pedestrians). An important property of this set is that the same vehicles (and pedestrians) were also captured with the PTZ camera (Some PTZ captures are shown in Figure 5.1). In this way, we were able to observe the classification accuracy of the hybrid system for the same samples. After the features had been extracted, samples in test set were classified with this kNN classifier. Table 5.1 shows the confusion matrix, per class accuracies and the overall accuracy (97.51%).

Table 5.1. Confusion matrix of the experiment with test data and accuracy of classifier on omnidirectional camera.

		Sample Number	Predicted Classes				Accuracy
			Van	Car	Motorcycle	Pedestrian	
Real Classes	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%
						Accuracy	97.51%

5.2. Classification with Hybrid Camera System

As described in more detail in Chapter 3, moving objects are classified using HOG features and SVM. 94 motorcycle, 126 car, 101 van and 104 pedestrian samples, each one of them is single frame, collected and their bounding boxes were labeled for the purpose of using as training set. Samples of cars and vans were scaled to 120x300 (1:2.5 ratio) pixels, the samples of other classes were scaled to 120x120 (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 be at the ideal position and scale within the PTZ camera frame.

Table 5.2 gives information on the accuracy of the PTZ classification when the SVM is selected in accordance with height-width ratio of the rectangle. For the height-width ratio, we analyzed the samples in the dataset, and determined 0.65 as the threshold which yields good performance in separation of motorcycles/pedestrians from cars/vans. In the table, it can be observed that approximately 91% of samples was classified correctly. This is lower than the omnidirectional camera classification accuracy given in Table 5.1. The majority of the misclassified samples are the ones sent to the wrong SVM classifier. We also tried kNN method to see whether the height-width ratio of the sample is closer to motorcycles/pedestrians or cars/vans, we obtained a similar performance shown in Table 5.3. Therefore, choice of SVM classifier made by height-width ratio is not good enough.

The experiment was repeated so that the SVM classifier is selected according to the final class of the object determined by the omnidirectional camera. In other words, if

Table 5.2. Accuracy of PTZ camera classification when SVM is selected according to the height-width ratio of bounding rectangle

Class	Sample Number	False SVM Number	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%
Accuracy					91.14%

Table 5.3. Accuracy of PTZ camera classification when SVM is selected based on kNN classifier using height-width ratio as feature

Class	Sample Number	False SVM Number	PTZ Classifier		Accuracy
			True Classification	False Classification	
Van	94	9	83	2	88.30%
Car	113	8	105	0	92.92%
Motorcycle	71	12	58	1	81.69%
Pedestrian	83	0	83	0	100%
Accuracy					91.14%

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. This can be considered as the hybrid camera system classification since both cameras are involved. Table 5.4 provides the results of this experiment. As is observed, the accuracy of the classification increased from approximately 91% to 99%. The system had biggest improvement in motorcycle class whose accuracy rose to 98.59%. False SVM choices were completely resolved. Misclassified examples of motorcycle and van are illustrated in Figure 5.2.

Table 5.4. Accuracy of PTZ camera classification when SVM is selected according to the classification result from omnidirectional camera

Class	Sample Number	False SVM Number	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%
Accuracy					99.17%

5.3. High Resolution Object Tracking

We tested the tracking module described in Chapter 4.3 by experiments. Figure 5.3a shows a case where there is only one object in the scene, which was previously classified as ‘pedestrian’ and that object is tracked until it leaves the scene. In the tracking scenario, it is reasonable to assume that there is a target class (e.g. pedestrian) and the PTZ camera is directed to the objects that are classified as the target class by the omnidirectional camera. There may be more than objects in the scene, one of them belonging to the target class. An example is given in Figure 5.3b. In this case, the PTZ camera tracks the target object as other objects enter or leave the scene (Figure 5.3c). Figure 5.4 and Figure 5.5 show the examples of the car and van tracking.

Our tracking experience revealed that there is a delay while PTZ is adjusting zoom and this delay could cause missing faster objects such as car, motorcycle. On the other hand, the system, described in 4.3, is capable of tracking pedestrians even there are several objects in the scene and PTZ camera could zoom in or out the object keeping desired sizes.

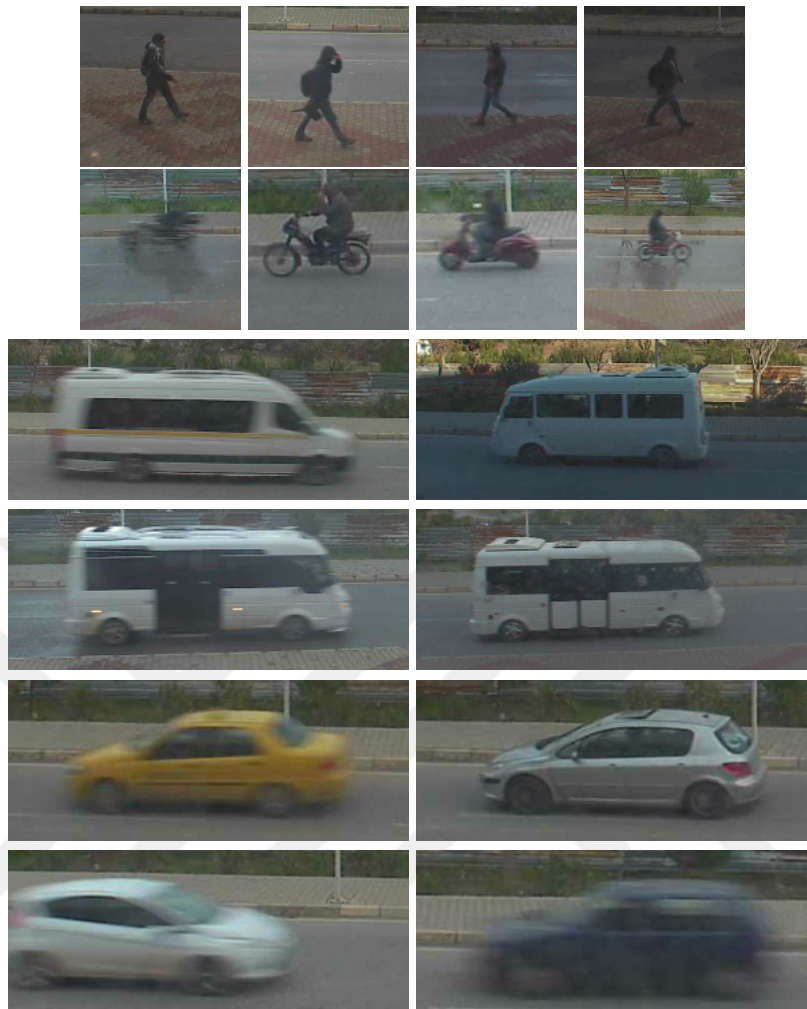


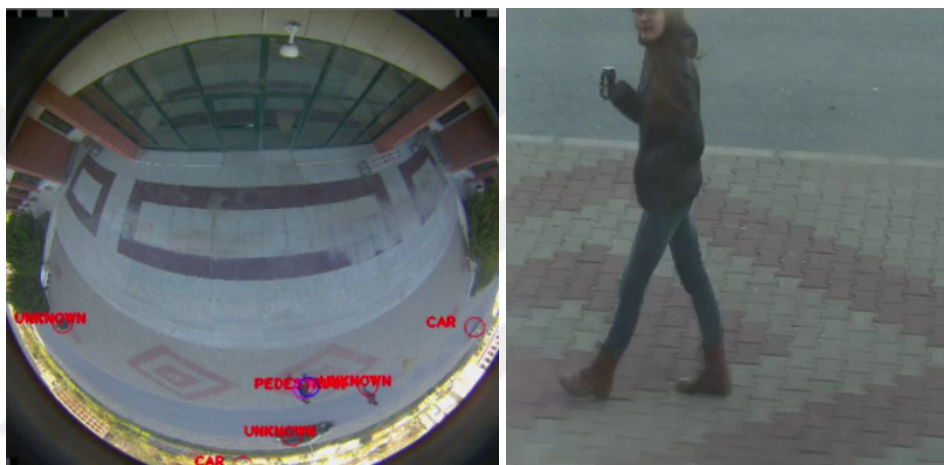
Figure 5.1. Test samples for the SVMs



Figure 5.2. Misclassified examples of motorcycle and van.



(a)



(b)



(c)

Figure 5.3. (a) : Pedestrian tracking when there is only one moving objects in the scene, which was previously classified as ‘pedestrian’ and tracked as it moves. (b,c) : Pedestrian tracking when the omnidirectional camera detects multiple objects in the scene. PTZ camera stays with the pedestrian in a later frame.

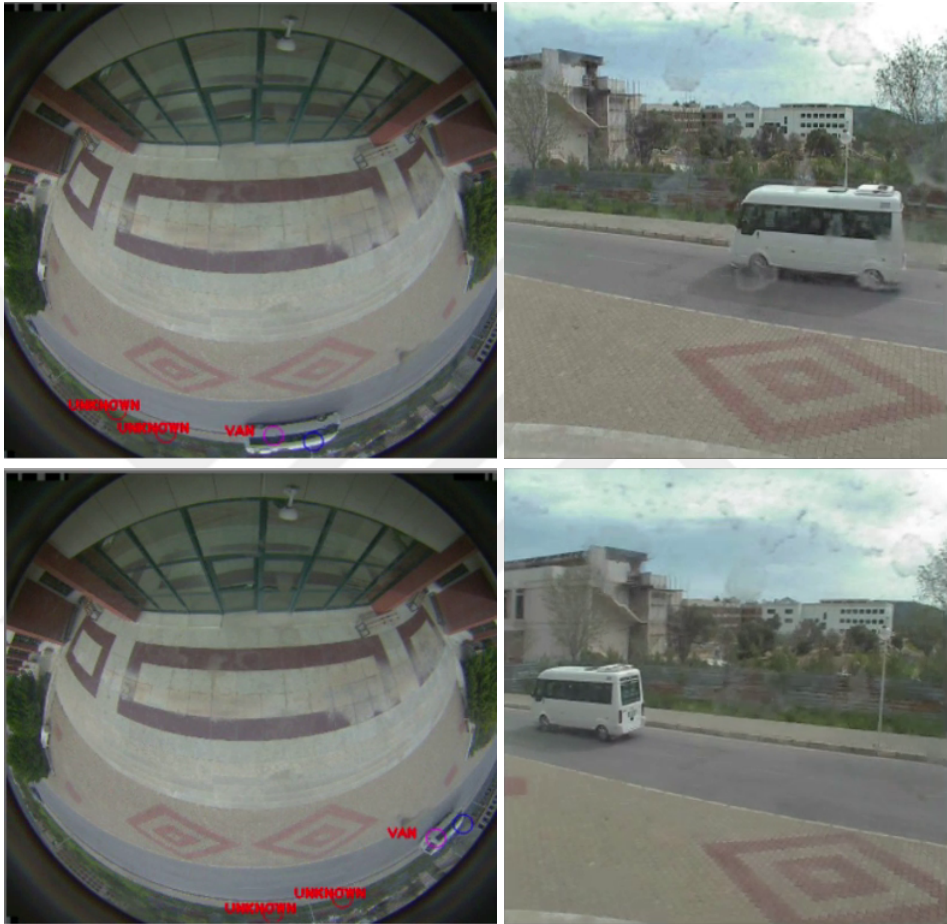


Figure 5.4. Example of van tracking. Purple circles indicate the location after updated with the formula given in Equation 4.3.



Figure 5.5. Example of car tracking. Purple circles indicate the location after updated with the formula given in Equation 4.3.

CHAPTER 6

CONCLUSION

We proposed a hybrid system containing an omnidirectional and a PTZ cameras. The hybrid system is capable of real-time co-operative classification and tracking providing wide angle and high resolution surveillance for traffic scenes.

For the co-operative classification, we employed a module for omnidirectional camera frames that provides multiple object detection and classification simultaneously. The omnidirectional camera classified the objects by using kNN classifier with shape-based features which are elongation, height-width ratio and area. Concurrently, the PTZ camera applied coarse level classification by using SVM classifier with HOG features. We implemented two different SVMs in order to have less computation and storage. One of the SVMs is pedestrian/motorcycle SVM, the other one is car/van SVM. We applied two different approaches for decision of which SVM classifying the object. The first approach was that height-width ratio of the silhouette extracted from PTZ frame decides the SVM. The other approach was that the decision based on the result of omnidirectional camera classification. The results showed that although the PTZ camera didn't classify the objects satisfactorily standalone, accuracies of the omnidirectional camera and the hybrid system classification were similar.

For the co-operative tracking, while the omnidirectional is classifying the objects, corresponding pan/tilt/zoom values of the target object whose class is user-defined are calculated and sent to the PTZ camera. The results indicated that the hybrid system showed good performance on pedestrian tracking. In this way, the PTZ camera can monitor the pedestrians, and take high resolution their images. Then, the images can be analyzed for tasks such as object recognition.

As a future work, we will detect the objects which are not included in current classification pool. For instance, moving objects such as dog can be observed the scenes, and they are classified mostly as motorcycle at present system. We will expand the system in order to classify that kind of objects as undefined type. Additionally, we will find a more effective solution for the problem of synchronization between the cameras.

REFERENCES

- [1] Adorni, G., M. Mordonini, S. Cagnoni, and A. Sgorbissa (2003). Omnidirectional stereo systems for robot navigation. In *Computer Vision and Pattern Recognition Workshop, Conference on*, Volume 7, pp. 79–79. IEEE.
- [2] Bastanlar, Y. (2016). A simplified two-view geometry based external calibration method for omnidirectional and PTZ camera pairs. *Pattern Recognition Letters* 71, 1–7.
- [3] Bastanlar, Y., A. Temizel, Y. Yardimci, and P. Sturm (2012). Multi-view structure-from-motion for hybrid camera scenarios. *Image and Vision Computing* 30(8), 557–572.
- [4] Bradski, G. and A. Kaehler (2008). *Learning OpenCV: Computer vision with the OpenCV library.* ” O’Reilly Media, Inc.”.
- [5] Buch, N., J. Orwell, and S. A. Velastin (2008). Detection and classification of vehicles for urban traffic scenes. In *Visual Information Engineering, 2008. VIE 2008. 5th International Conference on*, pp. 182–187. IET.
- [6] Chen, C.-H., Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi (2008). Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking. *Circuits and Systems for Video Technology, IEEE Transactions on* 18(8), 1052–1063.
- [7] Chen, X., J. Yang, and A. Waibel (2003). Calibration of a hybrid camera network. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 150–155. IEEE.
- [8] Cinaroglu, I. and Y. Bastanlar (2014). A direct approach for human detection with catadioptric omnidirectional cameras. In *Signal Processing and Communications Applications Conf. (SIU)*, pp. 2275–2279. IEEE.
- [9] Cinaroglu, I. and Y. Bastanlar (2016). A direct approach for object detection with catadioptric omnidirectional cameras. *Signal, Image and Video Processing* 10, 413–420.

- [10] Dalal, N. and B. Triggs (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, Volume 1, pp. 886–893. IEEE.
- [11] Deng, X., F. Wu, Y. Wu, F. Duan, L. Chang, and H. Wang (2012). Self-calibration of hybrid central catadioptric and perspective cameras. *Computer Vision and Image Understanding* 116(6), 715–729.
- [12] Gamallo, C., M. Mucientes, and C. V Regueiro (2013). A fastslam-based algorithm for omnidirectional cameras. *Journal of Physical Agents* 7(1), 12–21.
- [13] Gandhi, T. and M. Trivedi (2007). Video based surround vehicle detection, classification and logging from moving platforms: Issues and approaches. In *Intelligent Vehicles Symp.*, pp. 1067–1071. IEEE.
- [14] Ghidoni, S., A. Pretto, and E. Menegatti (2010). Cooperative tracking of moving objects and face detection with a dual camera sensor. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 2568–2573. IEEE.
- [15] Gupte, S., O. Masoud, R. F. Martin, and N. P. Papanikolopoulos (2002). Detection and classification of vehicles. *Intelligent Transportation Systems, IEEE Transactions on* 3(1), 37–47.
- [16] He, B., Z. Chen, and Y. Li (2012). Calibration method for a central catadioptric-perspective camera system. *JOSA A* 29(11), 2514–2524.
- [17] Hu, J., S. Hu, and Z. Sun (2013). A real time dual-camera surveillance system based on tracking-learning-detection algorithm. In *Control and Decision Conference (CCDC), 2013 25th Chinese*, pp. 886–891. IEEE.
- [18] Iraqui, H., Y. Dupuis, R. Bouteau, J.-Y. Ertaud, and X. Savatier (2010). Fusion of omnidirectional and PTZ cameras for face detection and tracking. In *Emerging Security Technologies (EST), 2010 International Conference on*, pp. 18–23. IEEE.
- [19] Karaimer, H. C. (2015). Shape based detection and classification of vehicles using

- omnidirectional videos. Master's thesis, Izmir Institute of Technology, Izmir, Turkey.
- [20] Karaimer, H. C. and Y. Bastanlar (2015). Detection and classification of vehicles from omnidirectional videos using temporal average of silhouettes. In *International Conference on Computer Vision Theory and Applications*.
- [21] Karaimer, H. C., I. Cinaroglu, and Y. Bastanlar (2015). Combining shape-based and gradient-based classifiers for vehicle classification. In *Intelligent Transportation Systems Conference (ITSC)*, pp. 800–805. IEEE.
- [22] Khoshabeh, R., T. Gandhi, and M. Trivedi (2007). Multi-camera based traffic flow characterization & classification. In *Intelligent Transportation Systems Conference*, pp. 259–264. IEEE.
- [23] Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval research logistics quarterly* 2(1-2), 83–97.
- [24] Kumar, P., S. Ranganath, H. Weimin, and K. Sengupta (2005). Framework for real-time behavior interpretation from traffic video. *Intelligent Transportation Systems, IEEE Transactions on* 6(1), 43–53.
- [25] Liu, Y., H. Shi, S. Lai, C. Zuo, and M. Zhang (2014). A spatial calibration method for master-slave surveillance system. *Optik-International Journal for Light and Electron Optics* 125(11), 2479–2483.
- [26] Morris, B. and M. Trivedi (2006a). Improved vehicle classification in long traffic video by cooperating tracker and classifier modules. In *Video and Signal Based Surveillance, IEEE International Conference on*, pp. 9–9. IEEE.
- [27] Morris, B. and M. Trivedi (2006b). Robust classification and tracking of vehicles in traffic video streams. In *Intelligent Transportation Systems Conf. (ITSC)*, pp. 1078–1083. IEEE.
- [28] Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics* 5(1), 32–38.

- [29] Piccardi, M. (2004). Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, Volume 4, pp. 3099–3104. IEEE.
- [30] Scotti, G., L. Marcenaro, C. Coelho, F. Selvaggi, and C. Regazzoni (2005). Dual camera intelligent sensor for high definition 360 degrees surveillance. In *Vision, Image and Signal Processing, IEE Proceedings*, Volume 152, pp. 250–257. IET.
- [31] Sobel, A. and A. Vacavant (2014). A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding* 122, 4–21.
- [32] Tan, S., Q. Xia, A. Basu, J. Lou, and M. Zhang (2014). A two-point spatial mapping method for hybrid vision systems. *Journal of Modern Optics* 61(11), 910–922.
- [33] Tarhan, M. and E. Altuğ (2011). A catadioptric and pan-tilt-zoom camera pair object tracking system for uavs. *Journal of Intelligent & Robotic Systems* 61(1-4), 119–134.
- [34] Welch, G. and G. Bishop (1995). An introduction to the kalman filter. university of north carolina, department of computer science. Technical report, TR 95-041.
- [35] Yang, M., K. Kpalma, and J. Ronsin (2008). A survey of shape feature extraction techniques. *Pattern recognition*, 43–90.
- [36] Yao, Y., B. Abidi, and M. Abidi (2006). Fusion of omnidirectional and PTZ cameras for accurate cooperative tracking. In *Video and Signal Based Surveillance, IEEE International Conference on*, pp. 46–46. IEEE.
- [37] Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, Volume 2, pp. 28–31. IEEE.