Moving Object Classification in Surveillance Videos

Rashmi P. Bijwe
ME Student, Dept of CE, SIPNA's College of Engineering, Amravati (MS) INDIA
rashmi.bijwe100@gmail.com

Dr. Prashant R. Deshmukh
Head & Professor, Dept of CSE, SIPNA's College of Engineering, Amravati (MS) INDIA
pr_deshmukh@yahoo.com

ABSTRACT

Classifying moving objects to semantically meaningful categories is important for automatic visual surveillance. Any outdoor surveillance system must be able to track objects moving in its field of view, classify these objects. Here, a system is presented which is able to detect and classify people and vehicles outdoors in different weather conditions. The system is capable of correctly tracking multiple objects despite occlusions and object interactions. Object tracking is performed using background subtraction, followed by region correspondence. The proposed framework incorporates four video features, object size, object velocity, location and difference of histogram of oriented gradients (DHoG). This paper involves accurate results by relying on powerful discriminative features.

1 INTRODUCTION

Automatic detection and recognition of objects is of prime importance for security systems and video surveillance applications. Automated video surveillance addresses real time observation of people and vehicles within a busy environment. With the rapid development of video capture technology, video is becoming a cheap yet important media for information record. One important task in video surveillance is to classify moving objects into semantically meaningful categories.

Outdoor surveillance systems must be able to detect and track objects moving in its field of view, classify these objects and detect some of their activities. This paper discusses the issues that any good surveillance system needs to cope with. These include problems faced in detection of objects, lighting conditions, shadows, the different types of occlusions that occur in the scene and entries/exits of objects. Typical applications include constructing intelligent parking systems for different vehicles and systems of object retrieval from videos, and so on. However, this recognition task is difficult, due to the some aspects describe later.

However, object shapes in video may change drastically under different camera view angles. In addition, the detected shapes may be noised by shadow or other factors. Actually, shape-based approaches often require that the scene and camera view for test are very similar to those for training. Such assumptions are inadequate in real applications. Another important feature is based on object motion. They can be used to recognize humans and vehicles [14]. However, it is difficult to use motion to classify vehicles into more categories, such as car, truck, van, etc.

In addition to conventional features like object size and velocity, this paper use differences of histograms of oriented gradients (DHoG) to measure the amount of intraobject deformation from frame to frame. This is a very useful feature to differentiate vehicles from people in different camera views, scenes with shadows, and to better separate groups of people from vehicles.

The remainder of this paper is organized as follows: First the general problems faced by surveillance systems in outdoor environments are described in section 2. Section 3 presents the proposed approach to classify moving objects into categories of people and vehicles; application is presented in Section 4, and conclusions are drawn in Section 5.

2 GENERAL PROBLEMS AND CHALLENGES

Classifying objects in urban surveillance scenes is an important task that allows searches and alerts based on object type. This section addresses a simplified two-class object recognition problem: Given a moving object in the scene, our goal is to classify the object into either a person, groups of people or a vehicle. This is a very important problem in city surveillance, as many existing cameras are pointing to areas where the majority of moving objects are either humans or vehicles.

This surveillance approach is based upon extracting objects in the form of regions from the scene using a background subtraction method, tracking these objects using region correspondence, classifying these objects into people, groups of people and vehicles Important Problems in Realistic Scenarios Detection and tracking of objects in a static camera is a nontrivial task. A number of problems including change in illumination, shadows, occlusion etc arise in realistic environments, which need to be dealt with, by the surveillance systems.

Object Detection. The first problem for automated surveillance is the detection of interesting objects in visible range of the video camera. The objects can be persons, vehicles. Almost all outdoor surveillance systems employ some variant of background subtraction methods to extract objects from the scene. However the background subtraction methods can't deal with the following problems.

Quick changes in lighting conditions completely change the color characteristics of the background. Almost all real time background subtraction methods can't model quick and large illumination variations. Thus surveillance under partially cloudy days will fail.

Uninteresting moving objects. For example flags waving or winds blowing through trees for short burst of time. Reflection of moving objects from shiny or wet surfaces also causes problems.

Shadows. Background subtraction methods fail to distinguish between an object and its shadow. Shadows can be of two types 1) self shadow and 2) cast shadow. The self-shadow is the part of the object, which is not illuminated by direct light. The cast shadow is the area in the background projected by the object in the direction of light rays. In outdoor images cast-shadows are major problems in acquiring accurate silhouettes. Cast Shadows make accurate silhouette analysis impossible, that is separate objects can appear to be joined together due to shadows. Inaccurate silhouettes also cause problem during classification of objects. Note that any shadow detection and removal scheme should only remove cast shadows since removal of self-shadows will result in incomplete silhouettes.

In this paper, the above mentioned problems must be solved before robust object detection in real scenes is possible.

Tracking under Occlusion. The goal of tracking is to establish correspondence between objects across frames. Occlusion occurs when an object is not visible in an image because some other object/structure is blocking its view. Tracking objects under occlusion is difficult because accurate position and velocity of an occluded object can't be determined. Different cases of occlusion are described in the following

Inter-object occlusion occurs when one object blocks the view of other objects in the field of view of the camera. The background subtraction method gives a single region for occluding objects. If two initially non-occluding objects cause occlusion then this condition can be easily detected. However if objects enter the scene occluding each other then it is difficult to determine. If inter-object occlusion is occurring, The problem is to identify that the foreground region contains multiple objects and to determine the location of each object in the region. Since people usually move in groups, which results in frequent inter-object occlusion so detecting and resolving inter-object occlusion is important for surveillance applications.

Occlusion of objects due to thin scene structures like poles or trees causes an object to break into two regions. Thus more than one extracted region can belong to the same object in such a scenario. The problem is compounded if multiple objects are occluded simultaneously by such a structure.

Occlusion of objects due to large structures causes the objects to disappear completely for a certain amount of time, that is there is no foreground region representing such objects. For example, a person walks behind a building, or a person enters a car. A decision has to be made whether to wait for reappearance of the objects, or determine that the object has exited the scene.

Exits and Entries of Objects from the Scene Entry is defined as an object entering the field of view of the camera. Entries and exits are easy to detect if (exiting and entering) objects are separate in the camera view. However, detecting an entry and an exit of two (or more objects) at the same place and at the same time is difficult. If one person enters the scene at a certain position while another person leaves from the same position at the same time then this scenario needs to be distinguished from the situation in which person moves right near the exit and then start moving in the direction he came from.

This paper present an approach that attempts to solve some of the above mentioned problems. The classification problem is very challenging, this paper desire to satisfy the following requirements:

•Real-time processing and low memory consumption

•The system should work for arbitrary camera views

•Correct discrimination under different illumination conditions and shadow effects

•Able to distinguish similar objects (such as vehicles and groups of people)

This paper approach to address these issues consists of three elements: (1) an interactive interface to set regions of interest and correct for perspective distortions; (2) discriminative features, including a novel effective measurement based on differences of histograms of oriented gradients (DHOG); and (3) estimation and adaptation based on a probabilistic framework for feature fusion.

3 PEOPLE AND VEHICLE CLASSIFICATION

3.1 Moving Object Detection

Background subtraction (BGS) is a conventional and effective approach to detect moving objects in videos captured by fixed cameras. In urban environments, more sophisticated background modeling algorithms are required to handle snow, rain, shadows, reflections, quick lighting changes, day and night transitions, slow moving objects, and other difficult conditions that naturally arise in city surveillance scenarios. In order to address these issues, this paper use an adaptive background subtraction technique similar to the method proposed by Stauffer and Grimson (1999), which relies on a Gaussian mixture model for each pixel in the scene. In this way, multimodal backgrounds can be described by the pixel-wise mixture models. For instance, during snow conditions, a particular background pixel may be represented by two Gaussian encodings, for example, the mode “road” and another mode “snow,” respectively, so that only objects of interest are detected. In contrast to Stauffer and Grimson (1999), the method is improved to remove shadows and to enable the algorithm to work for quick lighting changes by combining the intensity and texture information of the foreground pixels.

In addition, the foreground regions are classified as moving objects, abandoned objects, or removed objects without using any tracking or motion information. This capability can avoid a common problem in background subtraction—fragmentation (one object is partitioned into multiple parts).

3.2 Object Tracking

After a number of moving objects are detected by background subtraction algorithm, this paper track them along the video sequence so that each object is given a unique identification number. This is a prerequisite for other higher-level modules, such as object and color classification, and several real-time alerts. As in crowded urban environments, tracking turns out to be a very difficult and computationally

(a) ROI for People (b) ROI for Vehicles (c) Size Models

Figure 1. Calibration tool interface. (a) ROI for people - entire camera view. (b) ROI for vehicles - drive ways. (c) Person size models specified by the user. In (c), user can create, move and resize a target size sample (the largest box in the figure).

expensive task, so it can be intentionally disabled at specific hours of the day when the activity level is high. In these cases, analytics that do not rely on tracking are enabled.

The first step of tracking method consists of simple association of foreground blobs across the temporal domain. In each frame, this paper creates a list of the foreground blobs found by the background subtraction algorithm. From preceding frames, this have a list of recent tracks of objects, together with their bounding boxes. Each foreground blob in the current frame is then matched against all existing object tracks based on the distance and areas of the corresponding bounding boxes, so that tracks can be updated, created, or deleted.

In a simple scene, with well-separated tracks, the association gives a one-to-one mapping between tracks and the foreground regions in the current frame, except when tracks are created or deleted. In more complex scenes, however, There may have situations where a single track is associated with more than one foreground blob (object split), or the opposite—several tracks can be associated with a single foreground region (object merge). In addition, when one object passes in front of another, partial or total occlusion takes place, with background subtraction detecting a single moving region. The tracking method should be able to segment this region, label each part appropriately, and label the detected objects when they separate.

3.3 Object Classification with Calibration

I Interactive Interface for Calibration

In many situations, objects of one class are more likely to appear in certain regions in the camera view. For instance, in the city street environment, people usually walk along the sidewalk, while on the other hand vehicles mainly run in the middle of the road. This is a strong cue in the object classification process.

In this system, tracked objects classified into two classes, people and vehicles, and users can specify the regions of interest (ROIs) of each object class in the camera view through calibration tool. In this interactive calibration tool, one or multiple ROIs for the target class can be created, modified, and deleted as needed. Screenshots of the interface are shown in Figure 1 a and b. Note that this information is just used as a prior for the location feature. The final class estimation is based on probabilistic fusion of multiple features, as we will describe later.

Another purpose of calibration tool is to correct for perspective distortions by specifying different sizes in different locations of the camera field of view. A continuous, interpolated size map is then created from the sample sizes specified by the user. Since this paper use view dependent features such as object size and velocity, among others, for classification, the use of this information in this system allows it to normalize these features and thus work for arbitrary camera viewpoints, while significantly improving the accuracy of classification. Figure 1c shows our interface.

Feature Extraction

Given the limited computational resources and the real-time requirement in practical video surveillance applications, the features used for object classification must be low cost and efficient for computation. In this framework, this paper uses four object features: object size, velocity direction, object location, and differences of histograms of oriented gradients (DHOG). The purpose of using *object size* is that size information is the most distinctive feature to distinguish single persons

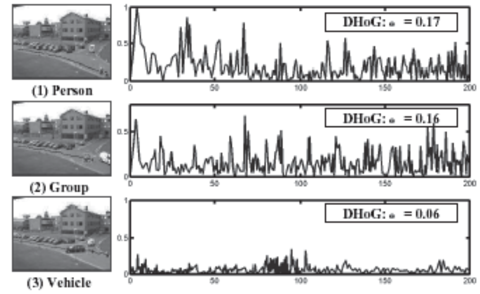

Figure 2. DHoG plots for three types of objects: Person,Group of People and Vehicle. Horizontal axis represents the track length of the samples, and vertical axis represents the DHoG values. Note that the overall DHoG values for Person Group of People are much higher than the one for Vehicle, and thus, it is a very discriminative feature to separate people from vehicles.

from vehicles since persons possess much smaller shapes than vehicles at the same location in the field of view. The size of an object at a specific frame is computed as the area of the corresponding foreground blob and then normalized by the size obtained from the interpolated size map at the same position.

In many scenarios, *velocity direction* of moving objects can be a distinctive feature. For instance, at a street intersection, pedestrians typically walk along the zebra crossings, which are perpendicular to vehicle movements. This paper equally discretizes the velocity direction measurement into 20 bins and normalize it in a similar way to the object size feature. There is an additional straightforward yet very informative feature: the *location* of the moving object. The location feature relates to the context of the environment, and its usage is applied through the settings of regions of interest specified by this calibration tool. This is a strong cue for identifying people in views such as roads and building entrances where vehicles seldom appear. Lastly, This paper has developed a novel view-independent feature—differences of histograms of oriented gradients (DHOG). Given an input video image with a foreground blob mask generated by the background subtraction module, the histogram of oriented gradients (HOG) is computed. It is well known that the HOG is robust to lighting condition changes in the scene. The HOG of the foreground object is computed at every frame in the track, and the DHOG is calculated in terms of the difference between HOGs obtained in consecutive frames in terms of histogram intersection. The DHOG models the intra-object deformation in the temporal domain. Thus, it is invariant to different camera views. In general, vehicles produce smaller DHOG than people since vehicles are more rigid when in motion. This feature is useful to distinguish large groups of people from vehicles, in which case they have similar shapes and sizes. Examples are shown in Figure 2 to demonstrate the effectiveness of using DHOG to distinguish vehicles from people (both single persons and groups of people). DHOG features have similar motivation as the recurrent motion image method (Javed and Shah 2002), but offer advantages as no precise object alignment is required.

4 PARKED CAR DETECTION

In this section, this paper describes a set of real-time alerts provided by this system, which are useful to detect potential threats in urban environments. Automatically detecting parked cars in urban scenarios can help security guards to quickly identify illegal parking, stopped vehicles in a freeway, or suspicious cars parked in front of security buildings at certain times. This paper describe a robust parked car detection system, which using various video cameras capturing images under a wide variety of conditions, including rain, day/night scenarios, etc. This approach is similar to the technique presented in Tian et al. (2008), but has novel aspects, which substantially reduce the rate of false alarms.

This paper use the background modeling algorithm described in Section 3 to detect foreground blobs in the scene. This paper also describes when these foreground blobs are static for a period of time, since the goal is to detect stopped vehicles. An important observation is that static foreground blobs may arise in the scene due to abandoned objects (e.g., parked cars) or due to removed objects (e.g., a car leaving a parking lot). As described in Tian et al. (2008), this paper provide a method to classify whether a static foreground region is abandoned or removed by exploiting context information about the static foreground regions. Incase the region is classified as abandoned object, then a set of user defined requirements is verified (e.g., minimum blob size and ROI) and a template-matching process is started to confirm that the object is parked for a user-defined period of time. If these conditions are satisfied, a parked car detection alert is triggered.

The method described above (Tian et al. 2008) is enhanced in two novel ways. First, use a naïve tracking algorithm (only across a user defined alert region of interest) to improve the classification of whether an object is abandoned or removed, for example, by looking at direction of movement and object speed as features. The second improvement addresses nighttime conditions where the template-matching process can fail due to low-contrast images. This matching scheme is designed to verify whether the vehicle is parked for the minimum time specified by the user. In many nighttime cases the car leaves the scene before this specified time, but the matcher still reports the car is there due to low contrast.

This approach substantially reduces the number of false alarms in challenging conditions, involving high volumes of data and weather changes, while keeping high detection rates.).

5 CONCLUSIONS

This paper has presented an integrated framework to classify moving objects into categories of People and Vehicles. The proposed framework incorporates four video features, object size, object velocity, location and difference of histogram of oriented gradients (DHoG). These features are easy to implement and computationally inexpensive. This system can handle urban environments efficiently. The framework being developed can automatically switch the system to low or high-activity modes. In future, the classification framework can be extend to other object classes or sub-classes (*e.g.* cars, vans and trucks).

6 REFERENCES

[1] B. Bose and E. Grimson. Improving object classification in far-field video. In CVPR, 2004.

[2] L. Brown. View independent ehicle/person classification. In VSSN, 2004.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.

[4] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In ECCV, 2000.

[5] H. Grabner and H. Bischof. Online boosting and vision. In CVPR, 2006.

[6] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. In WMVC, 2002.

[7] O. Javed and M. Shah. Tracking and object classification for automated surveillance. In ECCV, 2002.

[8] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006.

[9] L. Li, S. Yuan, and S. Xiang. Real-time object classification in video surveillance based on appearance learning. In CVPR, 2007.

[10] A. Lipton. Local application of optic flow to analyse rigid versus non-rigid motion. In CMU Technical Report, 1999.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60(1):91–110, 2004.

[12] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied for faces and cars. In CVPR, 2000.

[13] A. Senior. Tracking with probabilistic appearance models. In PETS, 2002.

[14] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In ICCV, 2003.

[15] B. Wu and R. Nevatia. Tracking of multiple, partially occluded humans based-on static body part detection. In CVPR,2006