Object Classification in Videos—An Overview

Insaf Setitra
Research Center in Scientific and Technical Information, University of Sciences and Technology Houari Boumeriene, Algiers, Algeria
Email: isetitra@cerist.dz

***Abstract*—In this article, we will discuss the classification of moving objects in videos. An overview of classical steps in video classification will be given and a particular attention will be given to classification in video surveillance since classification in this kind of systems is very important and plays a primary role in several functions such as event classification, speed control, classification of intrusions and so on.**

***Index Terms*—classification, moving objects, video sequence, Background subtraction, Feature selection, feature extraction, video surveillance, SVM**

I. Introduction

Object classification in video sequences is subject of several studies worldwide. The objective of this kind of research is to develop intelligent systems of video surveillance capable not only of capturing videos in real-time but also interpret them in terms of objects, classes and even behavior. A final end is to replace the traditional video surveillance systems -often inefficient when the number of cameras exceeds the number of human operators- by more intelligent systems capable of interpreting scenes accurately without any human intervention.

In this work we have explored the classification of objects in videos and focused on classification in video surveillance systems. The study is not exhaustive, but is simple and intuitive and will serve as a starting point for the communities concerned.

The paper is organized as follows: In the first part we address the classification in general, we expose conditions of a good classifier and we discuss some work belonging to the state of the art. Second part describes steps of classification in video sequences. More precisely steps of classification in video surveillance systems are addressed. The third part details a little more each of the steps mentioned above. Finally, a conclusion on what was discussed and guidance on future research will be presented.

II. Object Classification in Video Sequences

In a classical video surveillance system, an object classifier must have the following properties [1]:

A. Perform Under Real-Time Constraints

B. Be Robust to Several Conditions

- Large variation in natural conditions [2] e.g. weather conditions, including snow [3], fog [4], rain [5] and haze [6]
- Various lighting conditions and clouds casting shadows [7] and indirect illumination on objects (illumination from the headlights of a car, for example).
- Variation of the movement such as shakes of the camera due to strong winds.
- Total or partial occlusions of objects.
- Color variation in human clothes and cars [8].
- False moving objects detected such as branches and leaves swaying in the wind [9].
- Posture variations of humans [10].
- Spurious objects landing in the camera (birds, insects, etc.).
- Deterioration of the video due to compression operations or resolution of camera (high and low resolution).

C. Resolve a Multiclass Problem

Moreover, the latter property depends on several parameters:

- The number of classes to be used for classification:
    a. Vehicles, bicycles and persons [11]
    b. Vehicles, bicycles and group of persons [1]
    c. Vehicles and pedestrians [8], humans and vehicles [12]
- The choice of features to be considered;
- The classification algorithm chosen;

The three criteria described above enhances greatly the performance of the classification systems, however, meeting the three together is not always possible and systems proposed in the literature as far rarely do that.

In what follows, we describe, in a comprehensive manner, how an object classification system operates. Functions of the classification system are divided in three mandatory steps and each step is then detailed. Figure 1 depicts the mechanism of a classification system in a video sequence.

Manuscript received September 25, 2012; revised December 23, 2012.

doi: 10.12720/joace.1.2.106-109

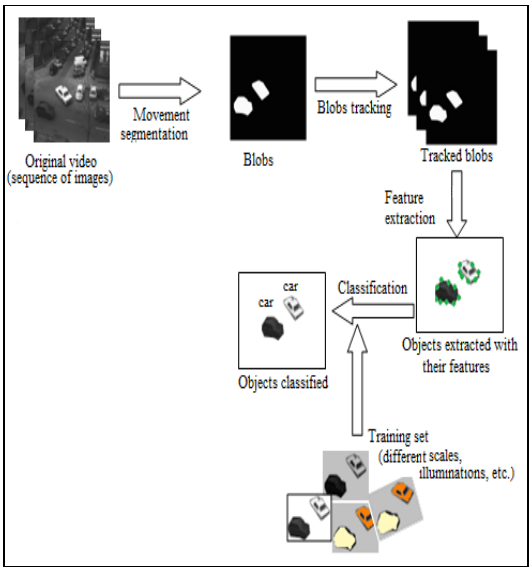

Figure 1. The mechanism of a classification system in a video sequence

III. Steps of Object Classification in Video Sequences

In the classification of one or more objects in a video sequence, several steps have to be respected. These steps are described below.

A. Movement Segmentation

Movement segmentation is the first step of classification since identifying moving objects from a video sequence is a fundamental and critical task in video surveillance. It consists of detecting regions of interest rather than considering the whole image of the sequence. By this way, time of execution is reduced since the classification system will focus only on interesting objects which are generally the moving ones.

This step is widely used in video surveillance systems since moving objects are the most likely to be interesting. Considering for example detection of intrusions, it is always more significant to classify a foreground of humans and interpret their behavior than classify a motionless background such as walls, trees and roads.

To distinguish these moving objects from stationary ones, segmentation methods use generally two types of information: temporal and spatial information and are categorized as follow:

1) Background Subtraction

Background subtraction is a very popular method of movement segmentation. In this method, an image -commonly called a frame- generally the first one of the sequence is taken as a reference and each remaining image of the sequence is compared to this one. Differences of images are considered as moving objects. At this time, moving objects are in the form of blobs, they then are tracked. By this, every object has its own sequence and so, its own properties.

The most common technique of background subtraction is the one of Stauffer and Grimson [13]. In this method, background is updated continuously using a mixture of Gaussians. Much works use this method to treat the classification problem, examples are works described in [14], [15], [9] and [16].

- Advantages: Simple, does not need an a priori knowledge of the object to be segmented and focus only on objects of interest rather than dealing with the whole image content.
- Disadvantages: Does not apply to non static background, is sensitive to lightening conditions, does not distinguish objects of their shadows, can detect false moving objects and is less accurate in case of occlusions.

2) Temporal Differentiation

This method calculates the absolute difference between two or three frames of the sequence and applies a movement threshold to determine which regions have undergone a significant movement. Then, the extracted moving parts are grouped into moving regions according to a connected component analysis. This method was introduced in the work of [17] and has been used particularly in the work of [18].

- Advantages: Adequate in the case of a moving background.
- Disadvantages: Does not always detect all moving pixels which diminish classification accuracy since some useful pixels can be ignored.

3) Optical Flow

This method uses optical flow to separate regions. The optical flow is a visual displacement field which considers variations in motion as a displacement of points in the image and assumes that color of pixels is independent of time.

The most common techniques used in calculating optical flow are: the Lucas-Kanade technique [19] for «sparse» optical flow and the Horn-Schunck technique [20] for «dense» optical flow. More information about optical flow can be found in [21]

- Advantages: Can be used on any camera configuration (static, non-static, calibrated, uncalibrated).
- Disadvantages: Has heavy computation and is sensitive to noise.

After choosing the method of segmentation, comes naturally the notion of feature extraction which will be discussed in the following section.

B. Feature Extraction

Feature extraction, sometimes confounded with feature selection, consists in extracting from each object detected previously in the video sequence a set of descriptors which represents the object. Feature selection then, consists in evaluating features extracted so that only significant descriptors and representing best the object and so the class to which the object belongs are hold. Features are commonly represented in a form of vectors. Thus, feature extraction is done for every object to be classified.

According to [16], features selected for classification must have the following properties:

- Class informative: all objects of the same class must be represented by the same set of descriptors.
- Discriminative: Each class has its proper set of descriptors.
- Minimal: should not contain redundant descriptors.

Thereby, descriptors should be a high interclass variability and low intra-class one.

Bose [8], in another hand, considers that features to be considered depend on objects to be classified. Hence, he selects descriptors manually and then evaluates them automatically by calculating their mutual information.

Besides, Bileschi, and Wolf [22] divide features into two broad categories: histograms of oriented edges and patch based features.

1) Histograms of Oriented Edges

A Histogram of oriented edges is described as a weighted histogram wherein each histogram bin collects votes from gradients near particular locations and particular orientations.

The most popular example of this approach is the SIFT (Scale Invariant Feature Transform). Other examples are the geometric blur, histogram of oriented gradient (HOG) and Hu moments.

- Advantages:This representation is discriminative and tolerates a large number of transformations that the image can undergo (scale, translations, etc.).
- Disadvantages: has sometimes a heavy computation.

2) Patch based Features

A patch based feature vector is an image description which depends on comparing the image with a set of stored image crops, also known as templates or fragments. Different implementations select different balances between invariance and the representation of geometric structure.

- Advantages: The use of multiple patches allows a better distinction between images.
- Disadvantages: Need to have a set of prototypes.

Many other features are present in the literature, such as symmetry, velocity, size and so on.

C. Object Classification

After getting interested in collecting objects and representing them by their features, the next step consists on choosing the right classification algorithm.

Classification problems are usually posed as a supervised learning problem, where a set of examples also called a training set -in our case images of the objects to be classified- and their classes are provided before any new classification. This is done as mentioned above, by extracting and selecting informative features which best represents the class. After preparing this training set, any new observation -in our case a video sequence- is treated to extract its moving objects. Moving objects are extracted by the movement segmentation described above. Their features are then extracted and compared to the set of labeled examples provided in the training set. At that moment, a particular classification algorithm is chosen for this comparison and an output of objects contained in the new observed video with their respective classes is provided.

In the following, only the most used classification algorithms in video sequences will be treated.

1) Support Vector Machine SVM

This algorithm was introduced by [23]. It solves the problem of binary classification (+1, -1) and aims to define a hyperplane of the formula:

$$y(x) = sign\ (w \times x + b) \tag{1}$$

where w and b are the parameters of the decision surface, x is the vector of descriptors and $y(x)$ the binary decision of the classifier. This hyperplane allows better separating the set of classes in the basis of support vectors by maximizing the margin between these vectors and the hyperplane.

Decision bounds separating the data are defined by the kernel function:

$$k(x,y) = exp(-c||x-y||2) \tag{2}$$

Garcia-Pedrajas and Ortiz-Boyer proposed in 2006 an extension of the SVM classifier to resolve a multi classification problem. The approach was taken back in 2010 by Gurwicz and al. [16]. However, it is worth noting that the SVM classifier is the most widespread classification algorithm in the area of pattern classification and more precisely in object classification in images and videos. More information on SVM can be found on [24].

2) Bayesian Network

Proposed by Pearl [25], this classifier aims to predict, based on a training set, to which class belongs any new observation of an object. Bayesian network is a directed cyclic graph with a set of conditional probabilities $P(Xi|Pai)$ where Pa is the parent set of the node X in the graph. The likelihood probability is given by:

$P(X) = P(X1,\dots, Xn)= \Pi\ P(Xi \mid Pai)$ with i from1 to n.

And the object is assigned to the class with best likelihood probability.

3) K Nearest Neighbors

This algorithm is among the easiest algorithms and generally allows a comparative study of classification algorithms.

The goal of this algorithms is to estimate the value of the unknown probability density function at a given point x. According to the k-nearest neighbor estimation technique, the following steps are performed:

- Choose a value for k.
- Find the distance between x and all training points x_i , i = 1, 2, . . . ,N. Any distance measure can be used (e.g., Euclidean, Mahalanobis, etc.). But Euclidean distance is usually used.
- Find the k-nearest points to x.
- Compute the volume $V(x)$ in which the k-nearest neighbors lie.

If the Euclidean distance is employed and the distance between the k-furthest neighbor and x is ρ, the volume $V(x)$ is equal to:

$V(x) = 2\rho$ in the 1-dimensional space

$V(x) = \pi\rho 2$ in the 2-dimensional space

- Compute the estimate by $p(x) \approx k / NV(x)$

The object is then assigned to the class with the best estimate.

Other algorithms exist in the state of the art but for the most part, a combination of several classifiers is used to ensure a best performance. In [16], K nearest neighbors was used as a reference.

IV. Conclusion

The goal of this paper was to give an overview of moving objects classification in videos. It described, briefly, what is object classification in videos and how does it work. Hence, it acted as a way to introduce the concept of object classification as a start point to readers who are unfamiliar with this area and to provide a review for more advanced researchers.

At this time, more works based on a combination of the different classifiers described in this paper can be projected. Choosing the right set of features for the representation of objects is also a wide area of research and need to be deepened. More future works should also include a special consideration to the feature selection since this step can improve classification in terms of accuracy and execution time. Finally, improving the segmentation and the tracking may provide a better classification.

References

[1] L. Chen, R. Feris, Y. Zhai, L. Brown, and A. Hampapur, "An Integrated System for Moving Object Classification in Surveillance Videos," *AVSS 08 IEEE*, pp 52-59, 2008.

[2] S. Munder and D. Gavrila, "An experimental study on pedestrian classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 1863–1868, 2006.

[3] H. Hase, K. Miyake, and M. Yoneda, "Real-time snowfall noise elimination," in *Proc. IEEE ICIP*, 1999, vol. 2, pp. 406–409.

[4] S. G. Narasimhan and S. K. Nayar. "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp 713–724, 2003.

[5] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," in *Proc. IEEE CVPR*, 2004, vol. 1, pp. 528–535.

[6] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proc. IEEE CVPR.*, 2001, vol. 1, pp. 325–332.

[7] S. G. Narasimhan and S. K. Nayar, "Shedding light on the weather," in *Proc. IEEE CVPR*, 2003, vol. 1, pp. 665–672.

[8] B. Bose, "Classifying Tracked Objects in Far-Field Video Surveillance," M.S. thesis, Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, February 2004.

[9] O. Javed and M. Shah, "Tracking and object classification for automated surveillance," *ECCV.* vol. 4, pp. 343–357, 2002.

[10] C. J Taylor, A. Lanitis, T. F. Cootes, G. Edwards, and T. Ahmed, *Computer Vision for Human-Machine Interaction*, UK: Cambridge University Press, 1998, pp. 217- 234.

[11] Z. Zhang, Y. Cai, K. Huang, and T. Tan, "Real-time moving object classification with automatic scene division," *IEEE ICIP*, 2007, vol. 5, pp. v-149-v-152.

[12] R. N. Hota, V. Venkoparao, and A. Rajagopal, "Shape based Object Classification for Automated Video Surveillance with Feature Selection," in *Proc. IEEE 10th International Conference on Information Technology*, 2007, pp. 97-99.

[13] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real time tracking," *IEEE Trans. PAMI 22*, vol. 8, pp. 747–757, 2000.

[14] B. Bose and E. Grimson, "Improving Object Classification in Far-Field Video," in *Proc. IEEE Computer Society Conference onComputer Vision and Pattern Recognition*, 2004, vol. 2, pp. II.

[15] B. Lei and L. Xu, "Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management," *Pattern Recognition Letters*, vol. 27, pp. 1816–1825, 2006.

[16] Y. Gurwicz, R. Yehezkel, and B. Lachover, "Multiclass Object Classification for Real-Time Video Surveillance Systems," *Pattern Recognition Letters*, vol. 32, pp. 805-815. 2011.

[17] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proc. IEEE Workshop Applications of Computer Vision*, 1998, pp. 8–14.

[18] M. Tsuchiya and H. Fujiyoshi, "Evaluating Feature Importance for Object Classification in Visual Surveillance," in *Proc. 18th International Conference on Pattern Recognition*, 2006, vol. 2, pp. 978-981.

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *DARPA Imaging Understanding Workshop*, pp. 121–130, 1981.

[20] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[21] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *Int J. Comput. Vis.*, vol. 12, no. 1, pp. 42–77, 1994.

[22] S. Bileschi and L. Wolf, "Image representations beyond histograms of gradients: The role of Gestalt descriptors," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.

[23] V. N. Vapnik, *The Nature of statistical learning theory*, New York: Springer-Verlag, 1995.

[24] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, 1998.

[25] J. Pearl, "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference," *Morgan-Kaufman*, 1988.

Insaf Setitra received the bachelor degree in Information systems and the master degree in Software engineering from the University of Sciences and technology Houari Boumediene Algeria in 2008 and 2010 respectively. She also received a bachelor degree in management from the university of management and economics of Algiers in 2006. She's currently working as a research associate in the Information and Multimedia systems' department of the Research Center on Scientific and Technical Information in Algeria. Her current research interests concern computer vision in video surveillance applications and embedded systems.