VEHICLE DETECTION AND TRACKING IN VIDEO

A.N. Rajagopalan and R. Chellappa

Center for Automation Research
University of Maryland
College Park, MD 20742-3275

ABSTRACT

In this paper, we present a scheme for vehicle detection and tracking in video. The proposed method effectively combines statistical knowledge about the class of vehicles with motion information. The unknown distribution of the image patterns of vehicles is approximately modeled using higher-order statistical information derived from sample images. Statistical information about the background is learnt 'on the fly'. A motion detector identifies regions of activity. The classifier uses a higher-order statistical closeness measure to determine which of the objects actually correspond to moving vehicles. The tracking module uses position co-ordinates and difference measurement values for correspondence. Results on real video sequences are given.

1. INTRODUCTION

Detection of vehicles in images represents an important step towards achieving automated roadway monitoring capabilities. It can also be used for monitoring activities in parking lots. The challenge lies in being able to reliably and quickly detect multiple small objects of interest against a cluttered background which usually consists of road signs, trees and buildings.

In recent works, the concept of site-model-based image exploitation has been used for the detection of prespecified vehicles in designated areas as well as the detection of global vehicle configurations in aerial imagery [1, 2]. The approach consists of maintaining a geometric functional model of the site of interest. Before an acquired image can be processed, it needs to be registered with respect to the site. In [3], an attentional mechanism based on the characterization and analysis of spectral signatures using context information is described. Moon et al. [4] use a simple geometric edge model in conjunction with contextual information for detecting vehicles from aerial images of parking areas. However, the method is sensitive to low illumination and/or acquisition angles. There is an increasing interest in the vision community to detect and track vehicles from video data. These approaches usually extract foreground objects from the background using frame differencing or background subtraction. The foreground objects are then classified as vehicles or otherwise using some matching criterion such as the Hausdorff measure [5] or trained neural networks [6, 7].

We present a new scheme that combines statistical knowledge of the 'vehicle class' with spatio-temporal information for classification and tracking. The unknown distribution of the image patterns of vehicles is approximately modeled by learning the higher-order statistics (HOS) of the 'vehicle class' from training images. To reduce false alarms, statistical properties of the background scene are learnt from the given test video sequence. A motion detector outputs regions of activity while the classifier uses an HOS-based closeness measure with very good discriminating capability to determine which of the moving objects actually correspond to vehicles in motion. The tracker uses position co-ordinates and HOS-based difference measurement values to establish correspondence across frames. When tested on real video sequences, the performance of the method is found to be very good.

2. DETECTION AND TRACKING

The proposed vehicle detection and tracking system is comprised of the following modules: statistical learning and parameter estimation, moving object segmentation, object discrimination, and tracking. We now discuss each of these modules in brief.

2.1. Statistical Learning and Parameter Estimation

Statistical information about the 'vehicle class' is derived off-line using a training set of vehicle image patterns. Let a random vector $\underline{X}$ of length N represent the class of vehicle image patterns and $\underline{x}$ be a lexicographically ordered sample image pattern from this class. Since the conditional density function for this class is unlikely to be well-modeled by a simple Gaussian fit, the unknown probability density function (p.d.f) is approximated up to its m^{th} order joint moment

This work is partially supported by the ARL Federated Laboratory on Advanced Sensors Consortium under contract DAAL-01-96-2-0001.

0-7803-6297-7/00/$10.00 © 2000 IEEE

using a finite-order HOS-based expansion which is given by

$$f(\underline{x}) = \mathcal{N}(\underline{\mu}, R)\left(1 + \sum_{n=3}^{m} E\left[\underline{H}_n^T(R^{-\frac{1}{2}}(\underline{X} - \underline{\mu}))\right] \underline{H}_n(R^{-\frac{1}{2}}(\underline{x} - \underline{\mu}))\right) . \quad (1)$$

Here E is the expectation operator with respect to $f(\cdot)$ while $\underline{H}_n(\underline{x})$ is a vector whose elements are given by the product $\left(\prod_{i=1}^{N} \frac{H_{k_i}(x_i)}{\sqrt{k_i!}}\right)$ for all permutations of k_i, $i = 1, \ldots N$, such that $\sum_{i=1}^{N} k_i = n$. The term $H_{k_i}(x_i)$ is the Hermite polynomial of order k_i which is given by

$$H_{k_i}(x_i) = \left[\frac{\partial^{k_i}}{\partial t_i^{k_i}} \exp\left(t_i x_i - \frac{1}{2} t_i^2\right)\right]_{t_i=0} .$$

Equation (1) uses higher-order statistics (the terms inside the summation) to get a better approximation to the unknown p.d.f and to capture deviations from Gaussianity. Details of the derivation can be found in [8]. The distribution of $\underline{X}$ is modeled by fitting the data samples of vehicles with six clusters. The idea of using multi-dimensional clusters to model the p.d.f may be traced back to [9, 10].

Because the underlying distribution is expected to be non-Gaussian, an HOS-based k-means clustering algorithm is used on the training set to derive information about the mean, the covariance and the higher-order statistics of the 'vehicle class'. Clustering is done using an HOS-based closeness measure which is given by $-\log f(\underline{x})$ where $f(\underline{x})$ is expressed as in (1). The higher-order information contained in (1) gives an enhanced discriminating capability to the closeness measure as compared to Euclidean and Mahalanobis distances which use only first and second-order statistics [11, 12].

2.2. Moving Object Segmentation

For a stationary camera, foreground objects are segmented from the background by frame differencing followed by thresholding. Simple thresholding can result in partial extraction of moving objects. If the threshold is too low, camera noise and shadows will produce spurious objects; whereas, if the threshold is too high, some portions of the objects in the scene will fail to be separated from the background. Hence, morphological operations are used to reconstruct incomplete targets and to remove extraneous noise. The net result is a binary image with the areas of motion identified. If the camera is in motion, then the image frames are first stabilized using a recently developed image stabilization algorithm [13]. Thresholding and morphological operations are carried out on the stabilized frames to detect the motion regions.

2.3. Object Discrimination

The task of this module is to determine which of the moving objects actually correspond to vehicles in motion. This is achieved by background learning, image search and classification.

2.3.1. Background Learning

Even though the HOS-based closeness measure has good discriminating capability, in the absence of any information about the background scene, there are many naturally occurring background patterns in the real world that could well be confused with the image patterns of vehicles. This can result in an unacceptable number of false detections. For a given image sequence, statistical information of the background scene can be used to enhance vehicle detection capability. Knowledge about the background scene helps to relax the detection threshold which in turn leads to an improvement in the vehicle detection rate while simultaneously keeping down the number of false matches.

Given an image frame and the knowledge-base of the vehicle class as derived in Section 2.1, the statistical parameters of the background (mean, covariance and HOS) are learnt as follows. The test image frame is scanned for square patches that are (most likely) not vehicles. The difference measurement of a test patch $\underline{x}$ with respect to the i^{th} vehicle cluster is computed as

$$d_{\underline{x}}^i = -\log N(\underline{\mu_i}, R_i)\left(1 + \sum_{n=3}^{m} E\left[\underline{H}_n^T(R_i^{-\frac{1}{2}}(\underline{X} - \underline{\mu_i}))\right] \underline{H}_n(R_i^{-\frac{1}{2}}(\underline{x} - \underline{\mu_i}))\right) .$$

If the minimum value of $d_{\underline{x}}^i$, $i = 1, 2, \ldots, 6$, is greater than a suitably chosen threshold $\overline{T}_b$, then the test patch is treated as a non-vehicle pattern. Since the background usually constitutes a major portion of the test image, one can obtain with good confidence, sufficient number of samples that are not vehicles by choosing a reasonably large value for T_b. The non-vehicle patterns are also distributed into six clusters using HOS-based k-means clustering and the statistical parameters corresponding to each of these clusters are estimated. For simplicity, the background statistics are assumed to be constant across the frames.

2.3.2. Image Search

Having learnt the statistical information of the background scene, we search for the presence of vehicles in and around the motion regions detected by the segmentation module. A window is chosen about each of these regions and searched for possible target at all points and across different scales to account for any variations in size. For every test pattern $\underline{x}$, a vector of HOS-based difference measurements is computed with respect to each of the 12 clusters; the first 6 clusters correspond to the class of vehicles while the other 6 are used to model the background. The minimum distance value and the corresponding cluster are obtained as

$$d_{\underline{x}} = \min_i - \log N(\underline{\mu_i}, R_i)\left(1 + \sum_{n=3}^{m} E\left[\underline{H}_n^T(R_i^{-\frac{1}{2}}(\underline{X} - \underline{\mu_i}))\right] \underline{H}_n(R_i^{-\frac{1}{2}}(\underline{x} - \underline{\mu_i}))\right), \quad 1 \le i \le 12 . \quad (2)$$

$$i_{\underline{x}} = \arg\min_i - \log N(\underline{\mu_i}, R_i)\left(1 + \sum_{n=3}^{m} E\left[\underline{H}_n^T(R_i^{-\frac{1}{2}}(\underline{X} - \underline{\mu_i}))\right] \underline{H}_n(R_i^{-\frac{1}{2}}(\underline{x} - \underline{\mu_i}))\right), \quad 1 \le i \le 12 . \quad (3)$$

2.3.3. *Classification*

The procedure for pattern classification is as follows.

- Based on the minimum difference measurements obtained from the search step, a set $\mathcal{A}$ of test patterns is generated such that $\underline{x} \in \mathcal{A}$ if the minimum difference value of $\underline{x}$ is less than an optimally selected threshold T_0 and the cluster corresponding to the minimum value belongs to the set of vehicle clusters. Equivalently, $\underline{x} \in \mathcal{A}$ if

  $$d_{\underline{x}} < T_0 \quad \text{and} \quad 1 \le i_{\underline{x}} \le 6 ,$$

  where $d_{\underline{x}}$ and $i_{\underline{x}}$ are derived from (2) and (3), respectively.

- If the set $\mathcal{A}$ is non-empty, then that test pattern $\underline{x}^*$ which has the smallest difference measurement value among the patterns in $\mathcal{A}$ is declared as the detected vehicle pattern within that motion region. The centroid of the detected vehicle pattern $\underline{x}^*$ along with the average of the HOS-based difference measurement values (which is given by $\frac{1}{12}\sum_{i=1}^{12} d_{\underline{x}^*}^i$) are passed onto the tracking module.

- If the set $\mathcal{A}$ is empty, then it is decided that the moving object is not a vehicle.

- The above steps are repeated for every motion region to check for the presence of a vehicle.

2.4. Tracking

Most tracking systems are based on either the Kalman filter or the correlation technique. In our scheme, the centroidal locations in conjunction with the average of the HOS-based difference measurement values of the detected vehicle pattern are used for tracking.

Step 1. The centroid corresponding to a detected foreground region is compared with the centroids of the objects detected in the earlier frame using the simple Euclidean norm.

Step 2. If the difference in the displacements of the centroids is less than a certain value, then correspondence is established (assuming that the frame rate is high enough).

Step 3. If there are multiple foreground regions that are likely candidates for match with an object in the previous frame, then the average of the HOS-based difference measurement values is used to establish a unique correspondence. This can be looked upon as matching using higher-order statistical correlation.

The tracking method described above works satisfactorily as long as the spatial positioning of the vehicles is not very complex. Velocity estimates could be computed for the motion regions and used together with the locations of the centroids for improved performance.

3. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed HOS-based vehicle detection and tracking system in natural real imagery against a cluttered background. The training set consisted of about 500 grey-scale patterns of vehicles (cars here), each of dimension 16 × 16 pixels. The method was then tested on real image sequences of vehicular activity on roadways captured with a moving camera. The training set was distinct from the test set. As a compromise between accuracy of representation and computational complexity, we chose $m = 3$ in (1) for our experiments.

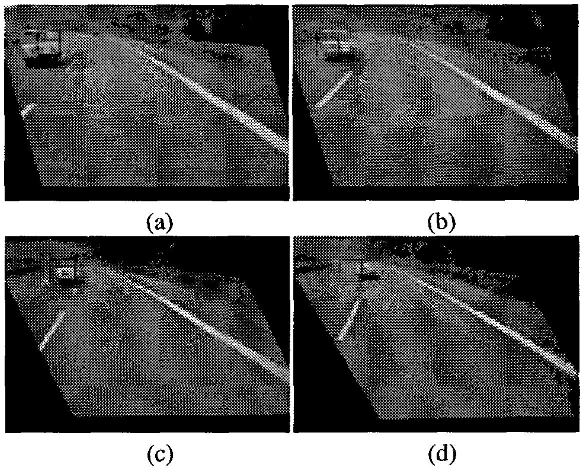

Fig. 1. Stabilized image frames of a moving car sequence captured with a moving camera. Detection and tracking results correspond to frame (a) 20, (b) 25, (c) 35, and (d) 43.

Figures 1 and 2 show the output results corresponding to some of the stabilized frames for two different car image sequences. Note that the background is very different for the two cases. For computational speedup, test patterns were evaluated every fourth pixel along the rows as well as the columns. Hence, the boxes are sometimes not exactly centered about the target. Note that the vehicles are successfully detected and tracked in each of the frames, despite the presence of heavy clutter. Even non-frontal views are detected. The potential of the HOS-based scheme for detection and tracking is quite evident from these results.

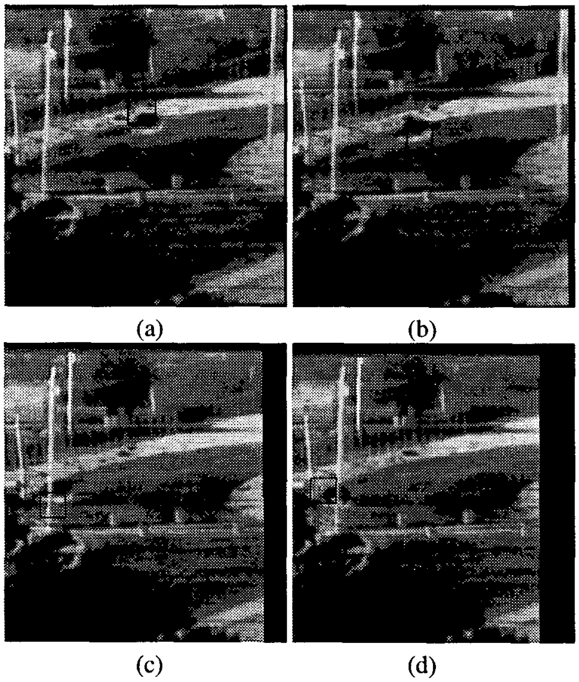

Fig. 2. Stabilized images for the second car sequence. Results corresponding to frames (a) 1, (b) 6, (c) 22, and (d) 34 are shown here.

4. CONCLUSIONS

We have described a new scheme for vehicle detection and tracking in video. The method effectively combines higher-order statistical information about the image patterns of vehicles with motion information for classification and tracking. The system successfully detects and tracks vehicles, even against complex backgrounds. The method is also reasonably robust to orientation, changes in scale, and lighting conditions. We are currently working on best-view selection and automatic detection of change in background to adaptively update statistical information about the background scene.

Acknowledgments

The test car sequence in Fig. 1 was kindly provided by Dr. Margrit Betke, Boston College.

5. REFERENCES

[1] R. Chellappa, Q. Zheng, L. Davis, C. Lin, X. Zhang, C. Rodriguez, A. Rosenfeld and T. Moore, "Site model based monitoring of aerial images", in *Proc. DARPA Image Understanding Workshop*, (Monterey, CA), pp. 295-318, 1994.

[2] P. Burlina, V. Parameswaran and R. Chellappa, "Sensitivity analysis and learning strategies for context-based vehicle detection algorithms", in *Proc. DARPA Image Understanding Workshop*, pp. 577-583, 1997.

[3] P. Burlina, R. Chellappa and C.L. Lin, "A spectral attentional mechanism tuned to object configurations", *IEEE Trans. Image Processing*, vol. 6, pp. 1117-1128, Aug. 1997.

[4] H. Moon, R. Chellappa and A. Rosenfeld, "Performance analysis of a simple vehicle detection algorithm", in *Proc. Fed. Lab. Symposium on Advanced Sensors*, (College Park, MD), 1999, pp. 249-253.

[5] D.H. Huttenlocher and R. Zabih, "Aerial and ground-based video surveillance at Cornell university", in *Proc. DARPA Image Understanding Workshop*, (Monterey, CA), 1998, pp. 77-83.

[6] T. Kanade, R.T. Collins and A.J. Lipton, "Advances in cooperative multi-sensor video surveillance", in *Proc. DARPA Image Understanding Workshop*, (Monterey, CA), 1998, pp. 3-24.

[7] A.J. Lipton, H. Fujiyoshi and R.S. Patil, "Moving target classification and tracking from real-time video", in *Proc. DARPA Image Understanding Workshop*, (Monterey, CA), 1998, pp. 129-136.

[8] A.N. Rajagopalan, Philippe Burlina and R. Chellappa, "Higher order statistical learning for vehicle detection in images", *Proc. IEEE Intl. Conf. on Computer Vision*, (Corfu, Greece), 1999, pp. 1204-1209.

[9] T. Poggio and K. Sung, "Finding human faces with a Gaussian mixture distribution-based face model", in *Proc. Asian Conf. on Computer Vision*, (Singapore), Springer Verlag, Eds. S. Z. Li, D. P. Mittal, E. K. Teoh and H. Wan, 1995, pp. 437-446.

[10] K. Sung and T. Poggio, "Example-based learning for view-based human face detection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, Jan. 98.

[11] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall Inc., Englewood Cliffs, 1988.

[12] R. O Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons Inc., 1973.

[13] S. Srinivasan and R. Chellappa, "Noise-resilient estimation of optical flow by use of overlapped basis functions", *Journal of the Optical Society of America - A*, vol. 16, pp. 493-506, 1999.